



HAL
open science

Intégration, caractérisation et modélisation des mémoires non-volatiles à nanocristaux de silicium

Stéphanie Jacob

► **To cite this version:**

Stéphanie Jacob. Intégration, caractérisation et modélisation des mémoires non-volatiles à nanocristaux de silicium. Micro et nanotechnologies/Microélectronique. Université de Provence - Aix-Marseille I, 2009. Français. NNT: . tel-00408813

HAL Id: tel-00408813

<https://theses.hal.science/tel-00408813>

Submitted on 3 Aug 2009

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.

UNIVERSITE DE PROVENCE – AIX-MARSEILLE I

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE DE PROVENCE

Ecole doctorale : Sciences Pour l'Ingénieur : Mécanique, Physique, Micro et Nanoélectronique

Présentée et soutenue publiquement le 2 avril 2008

par

Stéphanie JACOB

TITRE :

**INTEGRATION, CARACTERISATION ET
MODELISATION DES MEMOIRES NON VOLATILES A
NANOCRISTAUX DE SILICIUM**

DIRECTEUR DE THESE : Pascal MASSON, Prof. Université de Nice Sophia-Antipolis

CO-ENCADRANTS : Barbara DE SALVO, HDR, Dr. Ing. CEA-LETI Minatec

Gilles FESTES, Dr. Ing. ATMEL Rousset

JURY

Président : M. Rachid BOUCHAKOUR, Prof. Université de Provence, Marseille
Rapporteurs : M. Salvatore LOMBARDO, Dr. IMM-CNR, Catagne, Italie
M. Georges PANANAKAKIS, Prof. INP Grenoble
Examineurs : Mme Barbara DE SALVO, HDR, Dr. Ing. CEA-LETI Minatec, Grenoble
M. Gilles FESTES, Dr. Ing. ATMEL, Rousset
M. Pascal MASSON, Prof. Université de Nice Sophia-Antipolis
Invités : M. Romain COPPARD, Dr. Ing. R&D, Sofileta-CEA, Grenoble
M. Thierry PEDRON, Dr. Dir. Technologie Avancée, ATMEL, Rousset

*« Sachez vous éloigner, car, lorsque vous reviendrez
à votre travail, votre jugement sera plus sûr. »*,

Leonard de Vinci

*« On peut aussi bâtir quelque chose de beau avec les
pierres qui entravent le chemin. »*,

Johann Wolfgang Von Goethe

REMERCIEMENTS

Ce travail a été effectué au sein du Laboratoire des Nano-Dispositifs (LNDE) du CEA-LETI Minatoc et de la société Atmel Rousset, dans le cadre d'une convention CIFRE.

J'exprime d'abord toute ma gratitude à Simon Deleonibus, chef du LNDE et Olivier Demolliens, chef du D2NT (Département NanNoTechnologies) pour m'avoir accueillie au sein du CEA-LETI de Grenoble. Je remercie également Thierry Pedron, Directeur de la Technologie Avancée, pour en avoir fait de même au sein d'Atmel Rousset. Grâce à leur collaboration, j'ai pu bénéficier des meilleures conditions pour réaliser cette thèse.

Je remercie également Pascal Masson, mon directeur de thèse (professeur à l'Institut Matériaux Microélectronique Nanosciences de Provence (IM2NP) d'abord et maintenant à l'Université de Nice Sophia-Antipolis) de m'avoir fait confiance et de m'avoir soutenue pour ces trois années de thèse et depuis le stage de fin d'étude.

J'exprime toute ma reconnaissance et ma gratitude à Barbara De Salvo, mon encadrante au CEA/LETI, pour m'avoir guidée durant toute la durée de cette thèse, en me faisant partager sa grande expérience et ses précieux conseils. Merci de m'avoir fait confiance une nouvelle fois après le stage. Je tiens également à remercier Romain Coppard, pour avoir été mon encadrant chez Atmel pendant 2 ans et demi. Sa motivation et son enthousiasme ont largement contribué à faire avancer le projet. Merci de m'avoir renouvelé sa confiance pour de nouvelles aventures. J'adresse mes plus vifs remerciements à Gilles Festes pour avoir repris la suite de mon encadrement, pour son aide précieuse sur de nombreux points et pour avoir été toujours disponible malgré sa charge de travail. Qu'ils soient tous les trois assurés de toute ma gratitude.

Mes remerciements sincères vont à Rachid Bouchakour, Professeur à l'Université de Provence, pour avoir bien voulu présider le jury. J'exprime toute ma gratitude à Salvatore Lombardo, directeur de recherche à l'IMM-CNR et à Georges Pananakakis, Professeur à l'INPG, pour avoir bien voulu accepter la tâche de rapporteur.

Je souhaite également remercier Damien Deleruyelle, maître de conférence à l'Université de Provence, pour sa disponibilité au cours de ces trois ans et en particulier pour avoir assuré le relais de Pascal et pour sa présence régulière à Grenoble au début la thèse.

Je remercie ensuite très chaleureusement les membres de l'équipe mémoire : Marc Bocquet, Julien Buckley, Guillaume Gay, Marc Gély, Eric Jalaguier, Carine Jahan, Gabriel

Molas, Etienne Nowak, Luca Perniola et Tiziana Pro. Ils ont toujours été disponibles pour répondre à toutes mes questions. Leur bonne humeur a sans aucun doute contribué au bon déroulement de cette thèse. Un grand merci à Luca qui m'a énormément aidée au cours de ces trois ans. Merci pour son support sur la caractérisation électrique et pour son aide indispensable sur la modélisation. Merci à Gabriel pour son expertise. Merci à Eric pour son expérience, sa disponibilité pour nos discussions sérieuses (ou non) ainsi que son expertise en carnets de lots de nanocristaux « 3D » partagée à celle de Marc G. à l'accent chantant. Je souhaite bon courage aux thésards Tiziana, Marc B. (initiateur du Atchou'show qui m'a ruinée lors de son passage au 416), Guillaume et Etienne pour la suite.

J'adresse tous mes remerciements aux équipes du LSCDP (Laboratoire de Simulation et Caractérisation des Dispositifs et Procédés) dirigé par Fabien Boulanger pour leur support indispensable sur la caractérisation électrique (Alain Toffoli, Denis Blachier, Patrick Grosgeorges, Vincent Vidal, Fabienne Allain et Jacques Cluzel) et sur la simulation TCAD (Pascal Scheiblin, Pierrette Rivallin, Sylvain Barraud, Gilles Lecarval). Merci en particulier à Pascal pour ses précieux conseils et son aide sur les simulations porteurs chauds.

Ce travail n'aurait pu être effectué sans le support des équipes d'Atmel. Je remercie la fab pour avoir sorti les lots. Merci aux personnes de diffusion : Sylvie Bodnar, Stéphanie Angle et Gael Borvon. Merci en particulier à Sylvie pour les nanocristaux, pour s'être toujours intéressée à ces travaux et enfin pour sa gentillesse. Merci à Arnaud Talagrand pour le développement des étapes de gravure nécessaires à la fabrication du démonstrateur. Merci à Thibaut Pate-Cazal pour les caractérisations électriques sur PCM. Merci à Jean-François Thiery pour les caractérisations des matrices mémoires et pour avoir pris le temps de me former sur ces mesures ainsi que sur les méthodes de test des produits Flash, malgré sa charge de travail. Merci à Laurence Morancho avec qui j'ai posé mes premières pointes.

Je tiens également à remercier toutes les personnes de la salle blanche et des autres laboratoires du LETI (LFE, SDOT, SSIT, SCPIO) qui ont apporté leur contribution (de jour comme de nuit !) : Pierre Mur, Lilian Masarotto, Jean-Philippe Colonna...et beaucoup d'autres.

Comment ne pas citer les thésards du 416 qui ont partagé mon bureau durant ces années et qui ont apporté beaucoup de bonne humeur (par ordre d'apparition) :

Romain (qui m'a donné un avant-goût des joies et surtout des peines de la rédaction), Atsushi (pas folle la guêpe qui maîtrise le subjonctif), Fred (toujours en avance, au grand désespoir de ses voisins), Vince (merci pour ta culture musicale et footballistique), JP (collectionneur obsessionnel de boîtes de lot 2 pouces et fournisseur officiel de barres

chocolatées du labo), Guillaume (soit digne de mon bureau et surtout de mon PC...tant attendu) et Louis (la relève en matière de culture musicale est assurée !).

Je salue également les thésards CIFRE avec qui j'ai découvert le langage Atmélien au cours des premières semaines de ma thèse : Michel, Loeizig et Joël qui a également été l'autre « représentant » d'Atmel au LETI (merci pour les blagues raffinées... et aussi pour le travail en diffusion sur les nanocristaux et les oxydes).

Je n'oublie pas les autres thésards et post-doc du LNDE que je remercie pour avoir participé à l'ambiance chaleureuse qui règne à l'étage : Marco, Juliano, Cécilia, Emilie, Perrine, Michael, Jyotshna, Estelle, Stéphane, Sophie et Alexandre. Je leur souhaite à tous bonne continuation.

Je n'oublie pas de remercier toutes les autres personnes que j'ai côtoyées durant ces trois années et qui m'ont apporté leur aide ou qui ont contribué à créer une ambiance agréable, que ce soit au LETI : Corine, Julie, François, Olivier W., Cyrille, Christel, Bernard G., Georges, Thomas, Maud, Thierry, Olivier F, Sophie, Virginie, Laurent, Arnaud, Claude, Bernard P., Nathalie, Marie-Pierre, Xavier, Jérôme, Florent, Marie-Anne, Jean-Charles, Olga, Emmanuelle, Sébastien... ou que ce soit chez Atmel : Eva, Pascal, Alexis, Eleonore, Willem, Magali, Nadia, Didier, Gilles L., Stéphane, Patrick, Jean-Yves, Gilles M., Serguei, Florence, Mathieu, Jérôme, Jean-Paul, Jean-Marc B., François J., Elsa, Cécile, Catherine, Bruno, Bertrand, Lionel, Eric Y., Tamzin, Virginie, Bernard B., Yves T., Michel M., Luc J...

Je ne terminerai pas sans adresser un grand merci à mes parents pour tout ce qu'ils ont fait et pour le soutien qu'ils m'ont apporté durant toutes mes études. Je souhaite une bonne continuation à mon « petit » frère (bien que plus grand en taille).

Enfin, un immense merci à Ludo pour avoir été à mes côtés depuis toutes ces années. Merci pour sa patience et pour avoir supporté les sautes d'humeur d'une thésarde ainsi que les soirées, les nuits, les week-ends... et les mois de rédaction passés cloîtrés à la maison. Encore merci de m'avoir soutenue et motivée pendant ces trois années.

SOMMAIRE

Liste des abréviations	5
Liste des symboles	7
Introduction générale	9
Chapitre I : Introduction : Les mémoires non-volatiles à base de silicium.....	13
I.1 Introduction.....	17
I.2 Contexte	17
I.3 Historique des mémoires non-volatiles.....	20
I.4 Fonctionnement et architectures des mémoires Flash.....	22
<i>I.4.1 Présentation et fonctionnement de la cellule mémoire.....</i>	<i>22</i>
<i>I.4.2 Architectures des mémoires Flash.....</i>	<i>26</i>
I.5 Limites des mémoires Flash.....	28
I.6 Les solutions	31
<i>I.6.1 Les mémoires Flash à sites de piégeage discrets.....</i>	<i>32</i>
I.6.1.1 Les mémoires à base de nitrure	33
I.6.1.2 Les mémoires à nanocristaux de silicium	36
I.6.1.2.A Résultats publiés par Freescale	37
I.6.1.2.B Résultats publiés par STMicroelectronics.....	43
I.7 Conclusion	46
ANNEXE : Les solutions alternatives à la miniaturisation des mémoires Flash.....	47
Bibliographie.....	54

Chapitre II : Fabrication des mémoires Flash à nanocristaux de silicium.....	61
II.1 Introduction.....	65
II.2 Méthodes de fabrication des nanocristaux de silicium.....	65
<i>II.2.1 Précipitation de silicium en excès.....</i>	<i>65</i>
<i>II.2.2 Synthèse sous forme d'aérosol.....</i>	<i>65</i>
<i>II.2.3 Technique de croissance par CVD.....</i>	<i>66</i>
II.2.3.1 Procédé à une étape.....	66
II.2.3.2 Procédé à deux étapes.....	67
II.2.3.3 Influence de la préparation de surface avant dépôt des nanocristaux de silicium.....	68
II.2.3.4 Nitruration des nanocristaux de silicium.....	69
II.3 Intégration des nanocristaux de silicium dans un produit Flash NOR 130 nm.....	70
<i>II.3.1 Organisation d'une mémoire Flash.....</i>	<i>70</i>
<i>II.3.2 Fabrication d'une mémoire Flash standard.....</i>	<i>71</i>
<i>II.3.3 Les différentes possibilités d'intégration des nanocristaux de silicium</i>	<i>80</i>
<i>II.3.4 Niveaux de masques.....</i>	<i>93</i>
II.4 Conclusion.....	95
ANNEXE : Autres procédés de fabrication possibles intégrant les nanocristaux de silicium.....	96
Bibliographie.....	106
Chapitre III : Caractérisation électrique des mémoires Flash à nanocristaux de silicium.....	109
III.1 Introduction.....	113
III.2 Caractérisation électrique des cellules mémoires simples à nanocristaux de silicium.....	113
<i>III.2.1 Influence des méthodes de programmation.....</i>	<i>113</i>
III.2.1.1 Ecriture par porteurs chauds.....	113
III.2.1.2 Effacement par Fowler-Nordheim.....	114

III.2.2	<i>Influence des paramètres technologiques</i>	115
III.2.2.1	Influence de l'implant canal	117
III.2.2.2	Influence de l'épaisseur de l'oxyde tunnel	117
III.2.2.3	Influence de la taille et de la densité des nanocristaux	120
III.2.2.3.A	Influence du procédé de fabrication	120
III.2.2.3.B	Influence de la taille des nanocristaux	120
III.2.2.4	Influence de la passivation des nanocristaux.....	124
III.2.2.4.A	Passivation avec NH ₃ et NO	124
III.2.2.4.B	Dépôt d'une couche de Si ₃ N ₄	126
III.2.2.5	Influence du diélectrique de contrôle	127
III.2.2.5.A	Influence de la nature du diélectrique : HTO ou ONO.....	127
III.2.2.5.B	Influence de l'épaisseur, densification du HTO	129
III.3	Caractérisation électrique des matrices mémoires à nanocristaux de silicium	135
III.3.1	<i>Introduction</i>	135
III.3.2	<i>Influence du procédé de fabrication des nanocristaux de silicium</i>	136
III.3.2.1	Influence de la taille des nanocristaux	136
III.3.2.2	Comparaison des procédés une étape/deux étapes	137
III.3.3	<i>Influence des conditions de programmation</i>	137
III.3.3.1	Amélioration de la distribution des tensions écrites	138
III.3.3.1.A	Influence du temps d'écriture	138
III.3.3.1.B	Influence de la polarisation du substrat.....	139
III.3.3.2	Amélioration de la distribution des tensions effacées.....	140
III.4	Fiabilité	141
III.4.1	<i>Endurance</i>	141
III.4.2	<i>Rétention</i>	143
III.4.3	<i>Rétention après endurance</i>	144
III.4.4	<i>« Gate disturb »</i>	146
III.5	Conclusion	148
	Bibliographie	150
Chapitre IV : Modélisation des mémoires Flash à nanocristaux de silicium		
		151
IV.1	Introduction	155

IV.2 Le modèle de la quasi-grille flottante.....	155
<i>IV.2.1 Résultats des simulations avec le modèle de la quasi-grille flottante</i>	158
IV.2.1.1 Effacement Fowler-Nordheim	158
IV.2.1.1.A Influence de l'épaisseur de l'oxyde tunnel.....	158
IV.2.1.1.B Influence de l'épaisseur du HTO de contrôle.....	161
IV.2.1.1.C Influence du taux de couverture des nanocristaux de silicium.....	163
IV.2.1.1.D Comparaison avec la Flash standard	164
IV.2.1.2 « Gate disturb »	165
IV.3 Ecriture par porteurs chauds	169
<i>IV.3.1 Etude de la localisation de la charge dans une cellule mémoire à nanocristaux de silicium</i>	169
IV.3.1.1 Introduction	169
IV.3.1.2 Dispositifs d'étude	170
IV.3.1.3 Simulations électrostatiques	171
IV.3.1.4 Simulations dynamiques de l'écriture par électrons chauds et interprétation des données expérimentales	172
IV.3.1.4.A Simulations TCAD.....	172
IV.3.1.4.B Modélisation analytique.....	177
IV.4 Conclusion	182
ANNEXE 1 : Mécanismes de conduction tunnel Fowler-Nordheim et tunnel direct.....	184
ANNEXE 2 : Quelques éléments de la théorie des porteurs chauds.....	189
Bibliographie.....	192
Conclusions et perspectives.....	197
Bibliographie de l'auteur	203

Liste des abréviations

CHE	Channel Hot Electron
CI	Circuits Intégrés
DMA	Direct Access Memory
EEPROM	Electrically Erasable Programmable Read Only Memory
EFTEM	Energy-Filtered Transmission Electron Microscopy
EOT	Equivalent Oxide Thickness
EPROM	Electrically Programmable Read Only Memory
FAMOS	Floating-gate Avalanche-injection MOS
FLOTOX	FLOating gate Thin Oxide
HDP	High Density Plasma
HHI	Hot Hole Injection
HTO	High Temperature Oxyde
ITRS	International Roadmap for Semiconductors
LDD	Low-Doped Drain
MEB	Microscope Electronique à Balayage
MIMIS	Metal-Insulator-Metal-Insulator-Semiconductor
MNOS	Métal-Nitruire-Oxyde-Semi-conducteur
MOS	Métal Oxyde Semi-conducteur
MTJ	Magnetic Tunnel Junction
ONO	Oxyde/Nitruire/Oxyde
PCM	Phase Change Memory
RAM	Random Access Memory
ROM	Read Only Memory
SASTI	Self Aligned Shallow Trench Isolation
SILC	Stress Induced Leakage Current
Si-NCs	Nanocristaux de Silicium
SNOS	Silicium-Nitruire-Oxyde-Silicium
SONOS	Silicium-Oxyde-Nitruire-Oxyde-Silicium
STI	Shallow Trench Isolation
TEM	Transmission Electron Microscopy
UV	Ultra Violet

Table des symboles

Nom	Description	Valeur	Unité
ϵ_0	Permittivité du vide	$8,854.10^{-12}$	F/m
E_C	Energie du bas de la bande de conduction du Si		J
E_F	Energie du niveau de Fermi		J
ϵ_{ox}	Permittivité du diélectrique considéré		F/m
E_{ox}	Champ électrique dans le diélectrique considéré		V/m
ϵ_{Si}	Permittivité du Si	$11,9.\epsilon_0$	F/m
h	Constante de Planck	$6,62620.10^{-34}$	J.s
\hbar	Constante de Planck / 2π	$1,0546.10^{-34}$	J.s
J_{HTO}	Densité de courant à travers le HTO de contrôle		A/m ²
J_{tun}	Densité de courant à travers l'oxyde tunnel		A/m ²
k	Constante de Boltzmann	$1,38062.10^{-23}$	J.K ⁻¹
m_{ox}	Masse de l'électron dans le diélectrique considéré		kg
m_{Si}	Masse de l'électron dans le Si		kg
Φ_{dot}	Diamètre du nanocrystal		m
Φ_{SiO_2}	Barrière de potentiel Si/SiO ₂	3,15	eV
q	Charge élémentaire	$1,60219.10^{-19}$	C
Q_{FG}	Charge dans la grille flottante		C
R_{dot}	Portion de surface couverte par les nanocristaux		
T	Température		K
t_{HTO}	Epaisseur du HTO de contrôle		m
t_{ONO}	Epaisseur de l'ONO de contrôle		m
t_{ox}	Epaisseur du diélectrique considéré		m
t_{tun}	Epaisseur de l'oxyde tunnel		m
V _b	Tension appliquée sur le substrat d'un transistor		V
V _{CG}	Potentiel de la grille contrôle		V
V _d	Tension appliquée sur le drain d'un transistor		V
V _{FG}	Potentiel de la grille flottante		V
V _g	Tension appliquée sur la grille d'un transistor		V

V_{OX}	Chute de potentiel dans le diélectrique considéré	V
V_s	Tension appliquée sur la source d'un transistor	V
V_{th}	Tension de seuil	V
V_{thF}	Tension de seuil en lecture directe	V
V_{thR}	Tension de seuil en lecture inverse	V

INTRODUCTION GENERALE

Contexte

Depuis une vingtaine d'années, l'industrie de la microélectronique connaît une évolution considérable, en termes d'augmentation de la capacité d'intégration et de diminution du prix de revient. Ceci a permis au grand public d'accéder aux produits électroniques tels que les téléphones et ordinateurs portables, baladeurs MP3, clés USB ou encore appareils photos numériques, qui connaissent actuellement un énorme succès.

Cet essor considérable de l'industrie du semiconducteur a été possible notamment grâce à l'augmentation perpétuelle des performances du transistor MOS qui est la brique élémentaire des circuits intégrés. La diminution des dimensions du transistor MOS suit une loi exponentielle, connue sous le nom de la loi de Moore, établissant que la densité de transistors par microprocesseur double tous les deux ans.

Un autre marché, celui des mémoires non-volatiles et en particulier des mémoires Flash a également fortement contribué à la croissance de l'industrie des semiconducteurs. Ces dispositifs permettant le stockage de l'information sont en effet présents dans tous les nouveaux produits à succès cités précédemment.

Cependant, la miniaturisation des mémoires Flash standard risque de rencontrer des limitations au-delà du nœud technologique 45 nm, prévu vers 2010, principalement en ce qui concerne la réduction des diélectriques de la mémoire. C'est pourquoi les industriels et les laboratoires recherchent actuellement de nouvelles voies qui permettraient de prolonger la durée de vie de ces dispositifs. L'idée d'une cellule mémoire qui utilise des sites de piégeage discrets pour stocker la charge fait aujourd'hui l'objet de beaucoup d'attention, car c'est un candidat potentiel pour des futurs dispositifs mémoires à haute densité d'intégration et faible consommation en puissance. Plusieurs types de mémoires à pièges discrets ont été reportés dans la littérature. Dans ces dispositifs, le matériau de stockage peut être soit une couche continue avec des défauts naturels (pièges électriquement actifs) ou bien des nanocristaux réalisés avec différentes technologies.

C'est pourquoi dans cette thèse nous nous sommes attachés à étudier ces dispositifs et en particulier les mémoires Flash à nanocristaux de silicium.

Le travail de thèse

Cette thèse porte sur l'intégration, la caractérisation et la modélisation des mémoires non-volatiles à nanocristaux de silicium. L'objectif premier de cette thèse est l'étude expérimentale et théorique de ces mémoires.

Les points majeurs de notre étude sont présentés selon quatre chapitres.

Le **chapitre I** présente le contexte et le marché des mémoires Flash, ainsi que leur fonctionnement. Ensuite, les limitations de la réduction des dimensions de ces dispositifs sont exposées, ainsi que les solutions qui semblent les plus prometteuses pour repousser ces limites. Les études réalisées sur les cellules à nanocristaux de silicium et en particulier les résultats sur des matrices de plusieurs Méga bits (Mb) obtenus par des industriels sont présentés.

Le **chapitre II** traite de la fabrication des mémoires à nanocristaux de silicium. Dans la première partie, les différentes méthodes de fabrication des nanocristaux de silicium sont présentées. La deuxième partie du chapitre porte sur l'intégration des nanocristaux de silicium dans un produit ATMEL Flash NOR 32 Mb, basé sur une technologie 130 nm. L'organisation d'un produit mémoire Flash est d'abord présentée. Ensuite, nous détaillons le procédé de fabrication de la mémoire Flash standard à grille flottante continue et nous expliquons les différentes façons d'intégrer les nanocristaux de silicium à partir de ce procédé. Enfin, la dernière partie traite du nombre de masques pouvant être économisés avec un procédé à nanocristaux de silicium par rapport à un procédé standard.

Le **chapitre III** porte sur la caractérisation électrique des mémoires à nanocristaux de silicium. Les résultats électriques obtenus sur des cellules mémoires simples sont d'abord présentés à travers une étude exhaustive de l'influence des méthodes de programmation ainsi que des paramètres technologiques sur les caractéristiques d'écriture par électrons chauds et d'effacement par Fowler-Nordheim. Dans une deuxième partie, nous présentons les caractéristiques électriques de matrices mémoires et en particulier celles d'un démonstrateur ATMEL Flash NOR 32 Mb. L'influence du procédé d'élaboration des nanocristaux de silicium ainsi que des conditions de programmation est étudiée. Enfin, nous proposons une étude de fiabilité sur les matrices mémoires.

Dans le **chapitre IV**, nous nous intéressons à la modélisation des mémoires Flash à nanocristaux de silicium. Dans une première partie, l'effacement Fowler-Nordheim et le « gate disturb » sont simulés grâce au modèle dit de la quasi-grille flottante. L'influence des épaisseurs de l'oxyde tunnel et de contrôle ainsi que celle du taux de couverture des

nanocristaux de silicium sur l'effacement sont étudiées. Concernant le gate disturb, l'influence de la tension de grille de lecture et de l'épaisseur du HTO sont mises en évidence.

La deuxième partie du chapitre porte sur l'écriture par électrons chauds et en particulier sur la localisation de la charge dans les nanocristaux. Nous présentons une étude de l'influence des conditions d'écriture sur la localisation de la charge à l'aide de simulations TCAD et d'un modèle analytique couplé à des mesures expérimentales.

Enfin, le manuscrit se termine par une conclusion générale qui résume les principaux résultats obtenus dans le cadre de cette étude et des perspectives de ce travail sont proposées.

Chapitre I : Introduction : Les mémoires non-volatiles à base de silicium

Chapitre I : Introduction : Les mémoires non-volatiles à base de silicium.....	13
I.1 Introduction.....	17
I.2 Contexte	17
I.3 Historique des mémoires non-volatiles.....	20
I.4 Fonctionnement et architectures des mémoires Flash.....	22
<i>I.4.1 Présentation et fonctionnement de la cellule mémoire.....</i>	22
<i>I.4.2 Architectures des mémoires Flash.....</i>	26
I.5 Limites des mémoires Flash.....	28
I.6 Les solutions	31
<i>I.6.1 Les mémoires Flash à sites de piégeage discrets.....</i>	32
I.6.1.1 Les mémoires à base de nitrure	33
I.6.1.2 Les mémoires à nanocristaux de silicium.....	36
I.6.1.2.A Résultats publiés par Freescale	37
I.6.1.2.B Résultats publiés par STMicroelectronics.....	43
I.7 Conclusion.....	46
ANNEXE : Les solutions alternatives à la miniaturisation des mémoires Flash.....	47
Bibliographie.....	54

I.1 Introduction

L'objectif de ce premier chapitre est de présenter les principes de base des mémoires non-volatiles. Nous verrons dans quel contexte s'inscrivent les mémoires Flash et quel est leur fonctionnement. Nous présenterons également les limites des dispositifs Flash actuels et quelles sont les solutions envisagées pour les dépasser. Nous verrons enfin qu'une des solutions est l'utilisation des mémoires à sites de piégeage discrets et en particulier à nanocristaux de silicium, qui sont le sujet principal de cette thèse.

I.2 Contexte

Depuis le milieu des années 60, le marché des circuits intégrés (CI) a connu un essor exceptionnel et avec lui, celui des mémoires. Les revenus du marché des CI pour l'année 2006 sont d'environ 150000 millions de dollars (**Fig. I-1**), ce qui montre l'importance de celui-ci dans l'économie mondiale. Le marché des mémoires, dont les revenus représentent presque 30 % du marché des CI, est un moteur essentiel de l'industrie des semi-conducteurs.

En effet, tous les produits électroniques qui connaissent actuellement un grand succès auprès du grand public (téléphones portables, ordinateurs, clés USB, lecteurs MP3, assistants personnels) contiennent des mémoires.

La solution idéale serait une mémoire qui retiendrait l'information sans alimentation électrique extérieure, avec un accès en lecture et une programmation rapide et tout cela avec une haute densité d'intégration et une basse consommation en énergie. Cependant, la mémoire idéale, regroupant tous ces avantages à la fois n'existe pas, même si les mémoires Flash en présentent plusieurs d'entre eux (**Fig. I-2**). Il existe donc plusieurs catégories de mémoires selon les applications visées.

Les mémoires à semi-conducteur se divisent en deux catégories différentes : volatiles et non-volatiles (**Fig. I-3**). Les mémoires volatiles perdent leur information dès qu'elles ne sont plus alimentées. Ces mémoires ont un accès en lecture et programmation très rapide. Les mémoires non-volatiles, au contraire, retiennent l'information stockée indépendamment de l'alimentation extérieure. La programmation de ces dispositifs est plus lente que celle des mémoires volatiles.

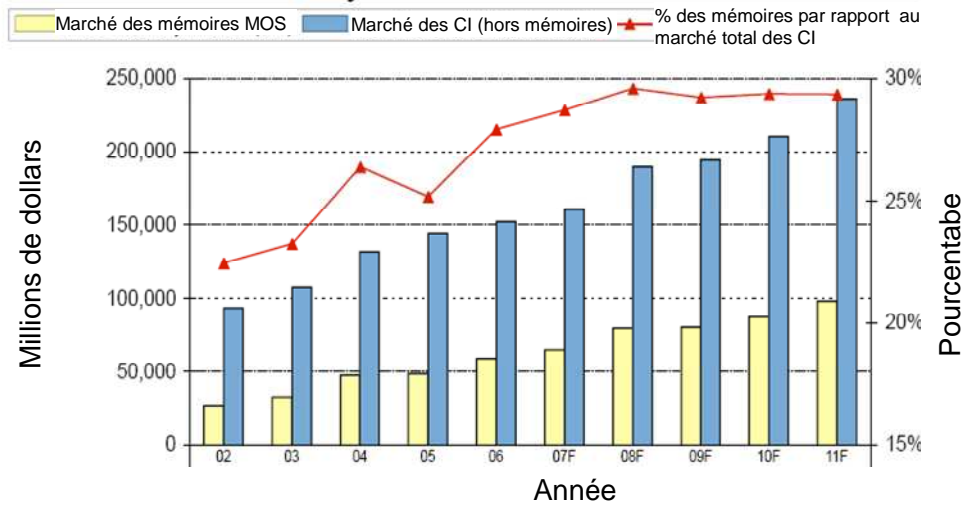


Fig. I-1: Evolution des revenus du marché des CI et du marché des mémoires à technologie MOS (F : prévisions). Source : WSTS, IC Insights.

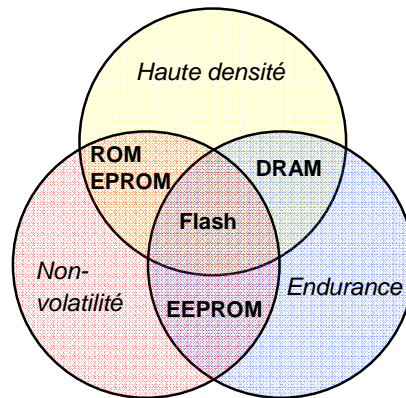


Fig. I-2 : Classification des mémoires à semi-conducteurs selon les critères de performance.

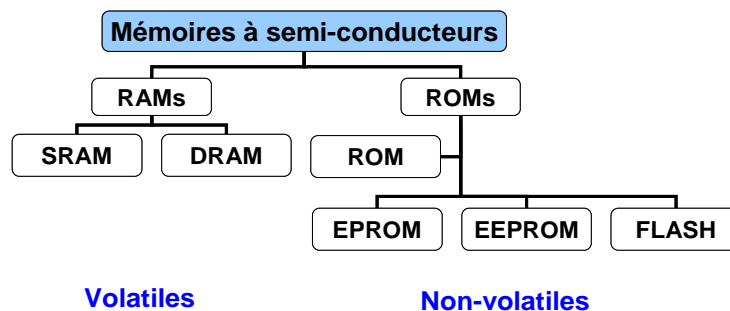


Fig. I-3 : Tableau récapitulatif des différentes classes de mémoires à semi-conducteurs.

Les mémoires à accès aléatoire RAM (Random Access Memory) constituent l'essentiel des mémoires volatiles. Ce nom se réfère au fait que l'on peut accéder à n'importe quel endroit de la mémoire très rapidement et dans n'importe quel ordre. Les principales catégories de RAM sont les SRAM (Static RAM) et les DRAM (Dynamic RAM).

La cellule SRAM utilise six transistors et retient l'information aussi longtemps qu'elle est alimentée. La taille de chaque cellule est donc assez grande et cela limite l'utilisation de la SRAM à des mémoires à basse densité d'intégration. De plus, son coût est plus élevé que celui de la DRAM. Par contre, l'accès aux données est plus rapide et la consommation plus faible.

Le point mémoire DRAM est constitué d'un transistor et d'une capacité servant à stocker la charge. Les capacités ne retenant pas l'information indéfiniment, il est nécessaire de rafraîchir la cellule mémoire régulièrement, d'où le nom « dynamique ». Les DRAM, de part leur petite taille, peuvent être utilisées dans les applications avec une haute densité d'intégration. De plus leur coût de fabrication est faible.

C'est pourquoi, le marché des DRAM est largement supérieur à celui des SRAM, comme le montre la **Fig. I-4**.

La première mémoire non-volatile a été la ROM (Read Only Memory) dans laquelle les données sont écrites de façon définitive au cours de la fabrication. Ce dispositif peut être lu mais jamais reprogrammé. Par la suite, la capacité à programmer la mémoire électriquement a successivement été ajoutée.

L'EPROM (Electrically Programmable ROM) peut être écrite électriquement mais elle doit être effacée par un passage sous rayons UV. Le point mémoire est constitué d'un seul transistor.

L'EEPROM (Electrically Erasable and Programmable ROM) peut s'écrire et s'effacer électriquement, mais au prix d'une complexité de cellule accrue. Le point mémoire utilise une surface équivalente à deux transistors. Le premier est un transistor de sélection et le second est l'élément de stockage. L'inconvénient de cette mémoire est donc la surface occupée qui est très importante et le coût qui est plus élevé.

La mémoire Flash EEPROM (généralement appelée mémoire Flash) peut tout comme l'EEPROM être écrite et effacée électriquement. Son nom provient du fait qu'un secteur ou une page entière peuvent être effacés en même temps. De plus, la surface occupée est faible car le point mémoire est constitué d'un seul transistor. Pour toutes ces raisons, les mémoires Flash sont aujourd'hui le type de mémoires non-volatiles le plus utilisé. La **Fig. I-4** montre la croissance remarquable du marché des mémoires Flash, qui bien qu'inférieur à celui des DRAM, tend à le rejoindre d'ici les prochaines années.

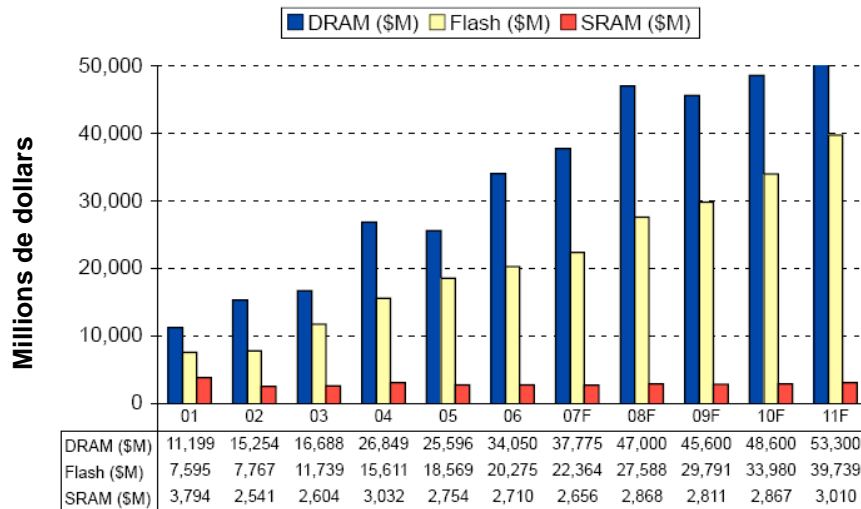


Fig. I-4 : Evolution des revenus du marché des mémoires à technologie MOS : les DRAM dominant mais les mémoires Flash gagnent du terrain (F : prévisions) : WSTS, IC Insights.

I.3 Historique des mémoires non-volatiles

Depuis le milieu des années 60, plusieurs solutions ont été proposées afin de résoudre le problème de la volatilité des mémoires à technologies MOS.

Deux solutions furent proposées en 1967 : le concept de grille flottante par D. Kahng des laboratoires Bell [**Kahng'67**] et la mémoire MNOS (Métal-Nitride-Oxyde-Semi-conducteur) par H.A.R. Wegener [**Wegener'67**]. La première était une structure de type MIMIS (Metal-Insulator-Metal-Insulator-Semiconductor). Le diélectrique du bas devait être assez fin (>5 nm) pour permettre aux électrons de passer par effet tunnel du substrat vers la grille flottante. Mais à cette époque, il était très difficile de déposer une couche aussi fine sans introduire de défauts. C'est pourquoi une autre structure utilisant un autre mécanisme d'injection avec un oxyde plus épais fut développée par Intel en 1971 [**Frohman'71**]. Son nom était la structure FAMOS (Floating-gate Avalanche-injection MOS). Le canal était de type P et la cellule n'avait pas de grille de contrôle. L'écriture était réalisée en appliquant une forte tension négative (-30 V) sur le drain afin de mettre la jonction drain-substrat en avalanche et créer des électrons fortement énergétiques sous la grille flottante. L'inconvénient de ce dispositif, outre les fortes tensions appliquées, était le fait que l'effacement devait se faire par ultra-violets, l'absence de grille de contrôle ne permettant pas de le faire électriquement.

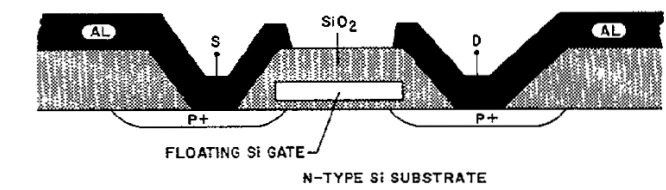


Fig. I-5 : Schéma du dispositif à grille flottante FAMOS (Floating-gate Avalanche-injection MOS) proposé par D. Frohman-Bentchkowsky en 1971 [**Frohman'71**].

La mémoire MNOS, proposée également en 1967 par Wegener, comme son nom l'indique, était une structure de type MOS avec en plus une couche de nitrure entre le métal et l'oxyde (**Fig. I-6**). La couche de nitrure agit comme une couche de piégeage des électrons et des trous. Cette structure était effaçable électriquement, par l'injection de trous par courant tunnel à travers l'oxyde et programmable de manière symétrique par injection d'électrons. Par la suite, afin d'améliorer la rétention de la charge, de nouvelles structures ont été développées : d'abord la SNOS (Silicium-Nitrure-Oxyde-Semiconducteur) puis la SONOS (Silicium-Oxyde-Nitrure-Oxyde-Semiconducteur) permettant de limiter l'injection de trous depuis la grille de contrôle.

En 1980, W. S. Johnson d'Intel présente une nouvelle structure électriquement effaçable qui sera la base de l'EEPROM telle qu'on la connaît aujourd'hui [**Johnson'80**]. Cette mémoire appelée FLOTOX (FLOating gate Thin OXide) est inspirée à la fois de la structure MNOS et de celle à grille flottante. A la différence des autres structures, celle-ci utilise le mécanisme tunnel Fowler-Nordheim pour l'écriture et l'effacement et elle est constituée de deux transistors (un transistor de sélection et un transistor de stockage), comme le montre la **Fig. I-7**.

La mémoire Flash EEPROM, quant à elle a été présentée pour la première fois dans un produit en 1984 par un ingénieur de Toshiba, F. Masuoka [**Masuoka'84**]. Cette cellule se base sur le concept de grille flottante. L'objectif était de combiner la haute densité d'intégration des EPROM avec l'effacement électrique des EEPROM (**Fig. I-8**). Le terme Flash provient du fait qu'elle permet un effacement très rapide de plusieurs blocs en simultanément. La commercialisation d'une mémoire Flash de type NOR (voir I.4.2) a été réalisée en 1988 par Intel [**Kynett'88**], suivi de près par celle d'une mémoire type NAND (voir I.4.2) par Toshiba en 1989.

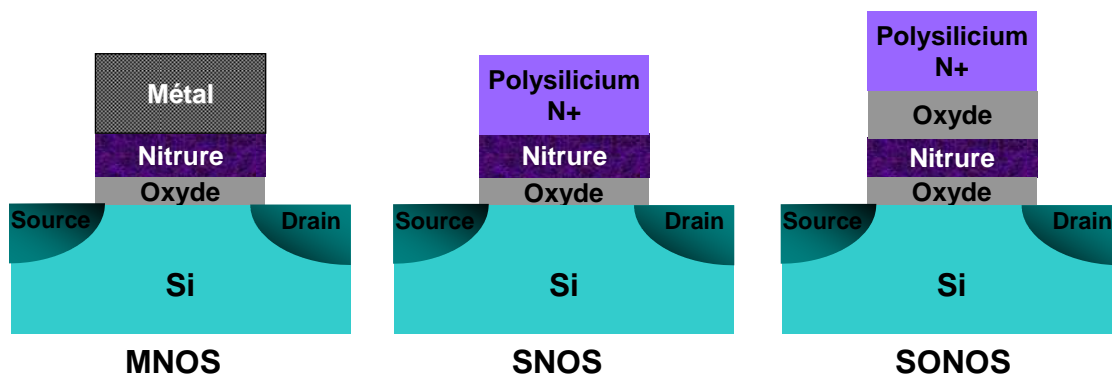


Fig. I-6 : Schémas des structures mémoires à base de nitrure comme couche de stockage.

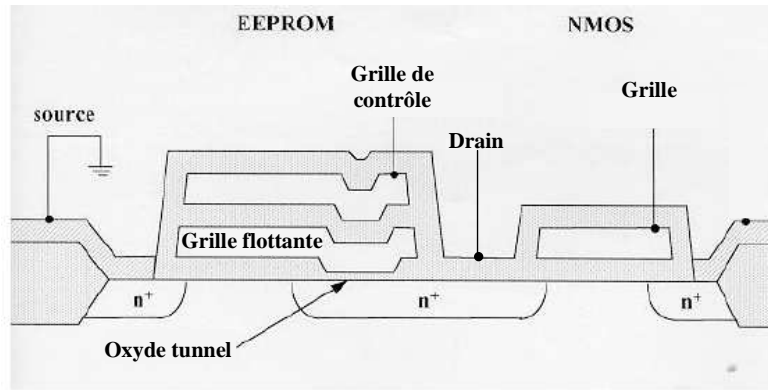


Fig. I-7 : Coupe schématique d'une cellule mémoire FLOTOX avec le transistor de sélection.

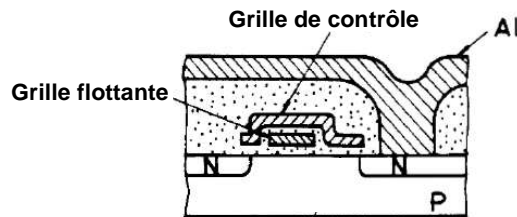


Fig. I-8 : Coupe schématique d'une cellule mémoire Flash selon [Masuoka'84].

I.4 Fonctionnement et architectures des mémoires Flash

I.4.1 Présentation et fonctionnement de la cellule mémoire

La cellule mémoire Flash à grille flottante peut être décrite selon la **Fig. I-9**. Cette structure est basée sur celle d'un transistor MOS avec en plus une grille flottante en polysilicium qui sert de couche de stockage des électrons. La charge dans la grille flottante induit un décalage de la tension de seuil du transistor MOS ΔV_{th} qui peut être exprimé de la façon suivante :

$$\Delta V_{th} = -\frac{Q_{FG}}{C_{CG}} \quad \text{Équation I-1}$$

Avec Q_{FG} : charge stockée dans la grille flottante et C_{CG} : capacité grille de contrôle/grille flottante.

Si la grille flottante est déchargée, on dit que la cellule est effacée. Si la grille flottante est chargée en électrons et donc avec une tension de seuil plus élevée, la cellule est dite écrite (**Fig. I-10(a)**). La lecture de la cellule, c'est-à-dire la détermination de l'état dans lequel se trouve la cellule s'effectue en appliquant une tension de grille $V_{G-lecture}$, comprise entre les deux tensions de seuil correspondant à l'état effacé et écrit, que nous nommerons respectivement V_{th1} et V_{th2} . Si la cellule mémoire est effacée (i.e. $V_{th} = V_{th1} < V_{G-lecture}$) alors le transistor constitué par la cellule est passant. Si la cellule mémoire est écrite (i.e. $V_{th} = V_{th2} > V_{G-lecture}$) alors le transistor constitué par la cellule est bloqué.

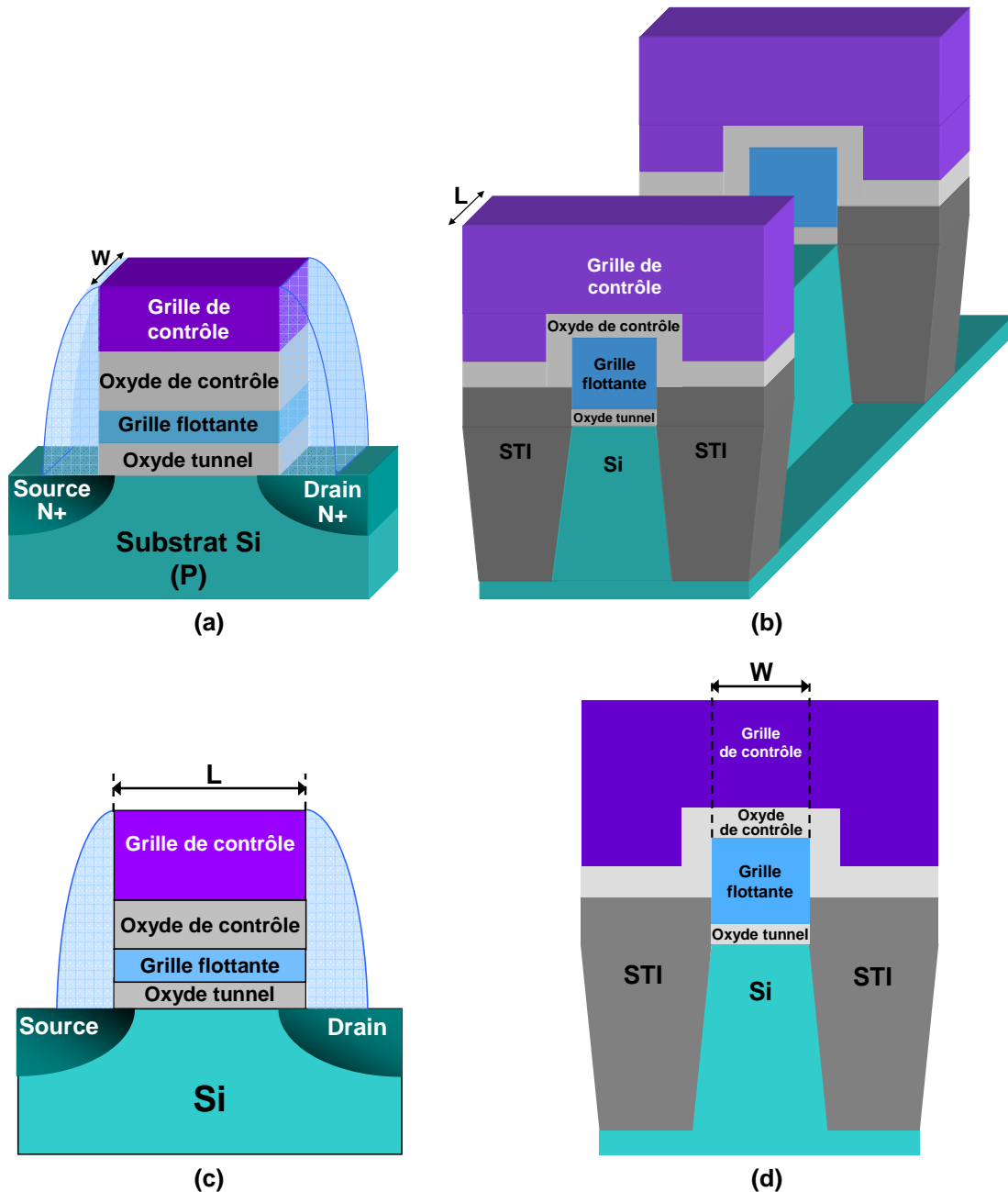


Fig. I-9 : Schémas d'une cellule mémoire Flash : (a) vue générale de la cellule mémoire, (b) vue schématique de deux cellules mémoires Flash dans un plan mémoire, (c) coupe longitudinale de la cellule, (d) coupe latérale de la cellule.

La modélisation du point mémoire habituellement utilisée est basée sur la structure capacitive de la **Fig. I-10(b)**. Il s'agit d'un modèle simple permettant de calculer la répartition du champ électrique dans chacun des diélectriques d'isolation [San'92].

Ce modèle permet d'accéder au potentiel de la grille flottante V_{FG} :

$$V_{FG} = \frac{Q_{FG}}{C_T} + \alpha_G V_{CG} + \alpha_S V_S + \alpha_D V_D + \alpha_B V_B \quad \text{Équation I-2}$$

$$\text{Avec } \alpha_G = \frac{C_{CG}}{C_T}, \alpha_D = \frac{C_D}{C_T}, \alpha_S = \frac{C_S}{C_T}, \alpha_B = \frac{C_{TUN}}{C_T} \quad C_T = C_{CG} + C_{TUN} + C_S + C_D$$

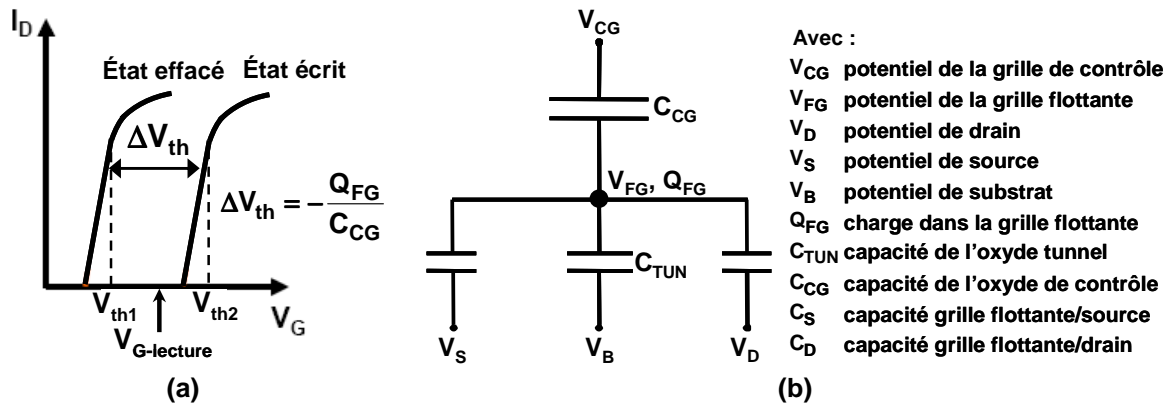


Fig. I-10 : (a) : Caractéristiques I_D - V_G (I_D courant de drain) d'une cellule mémoire Flash dans l'état effacé et écrit. (b) : Modélisation électrique du point mémoire Flash.

De plus, si la source et le substrat sont à la masse et si la tension de drain est faible, le potentiel de la grille flottante vaut approximativement :

$$V_{FG} = \frac{Q_{FG}}{C_T} + \alpha_G V_{CG} \quad \text{Équation I-3}$$

Cette dernière équation montre l'importance du coefficient α_G . Plus il est élevé, plus le potentiel de la grille flottante est proche de celui de la grille de contrôle. Le couplage de grille indique en fait la manière dont le champ électrique va se répartir entre l'isolant interpoly et l'oxyde tunnel. Ce coefficient doit donc être le plus grand possible pour exploiter au mieux les tensions utilisées. Typiquement, ce coefficient est de 0,7/0,8 dans une structure standard.

Les modes d'écriture de la mémoire Flash sont détaillés dans la **Fig. I-11**. On distingue principalement deux modes d'écriture : le mode Fowler-Nordheim et le mode porteurs chauds (CHE : Channel Hot Electrons). La **Fig. I-12** spécifie les modes d'effacement, qui sont plus nombreux. Nous avons ajouté le cas de la NROM (Nitride Read Only Memory) dont la couche de stockage est une couche de nitrure et dont le mode d'effacement est l'injection de trous chauds (HHI : Hot Hole Injection).

Modes d'écriture

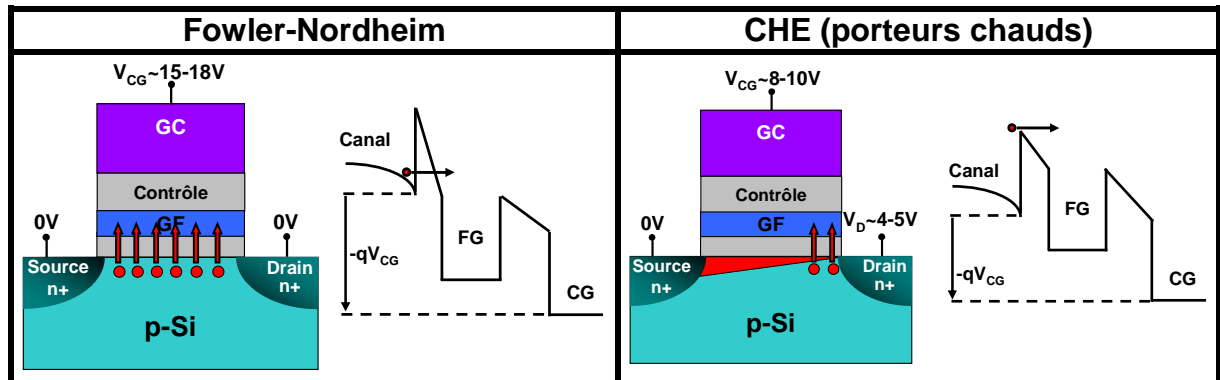


Fig. I-11 : Synthèse des différents mécanismes d'écriture pour les mémoires Flash. A droite de chaque cellule mémoire se trouvent des schémas de bande en énergie dans l'empilement de grille de la cellule. (CHE : Channel Hot Electron).

Modes d'effacement

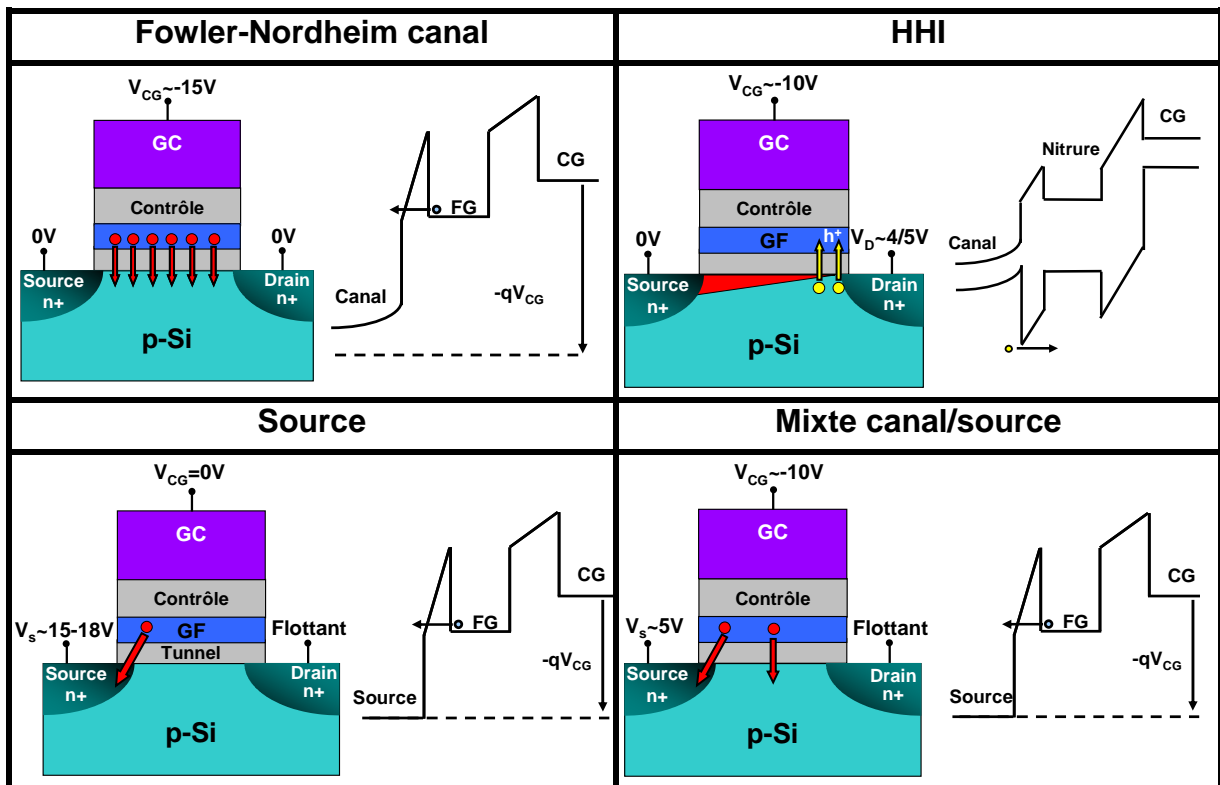


Fig. I-12 : Synthèse des différents mécanismes d'effacement pour les mémoires Flash. A droite de chaque cellule mémoire se trouvent des schémas de bande en énergie dans l'empilement de grille de la cellule. Dans le cas de l'effacement HHI (Hot Hole Injection), le diagramme de bande représente également la bande de valence associée aux trous. Ce mode d'effacement est uniquement utilisé par les mémoires de type NROM [Larcher'04].

1.4.2 Architectures des mémoires Flash

Les mémoires Flash sont organisées en matrices. Les architectures les plus utilisées sont la NOR et la NAND [Cappelletti'99].

Dans le cas de l'architecture NOR, les cellules mémoires sont connectées en parallèle (Fig. I-13(a)). La grille de contrôle est commune à une ligne de cellules appelée ligne de mots (« word line ») et le drain est commun à une ligne de bits (« bit line »). Cette architecture permet un accès aléatoire à n'importe quelle position. Le drain de chaque cellule étant accessible, l'écriture peut s'effectuer par porteurs chauds. Les cellules sont effacées par Fowler-Nordheim. Les temps d'écriture et d'effacement d'un bloc de cellules sont longs, mais le temps d'accès est très court. La Flash NOR est essentiellement utilisée pour le stockage de codes d'instruction (téléphones portables, assistants personnels, ...).

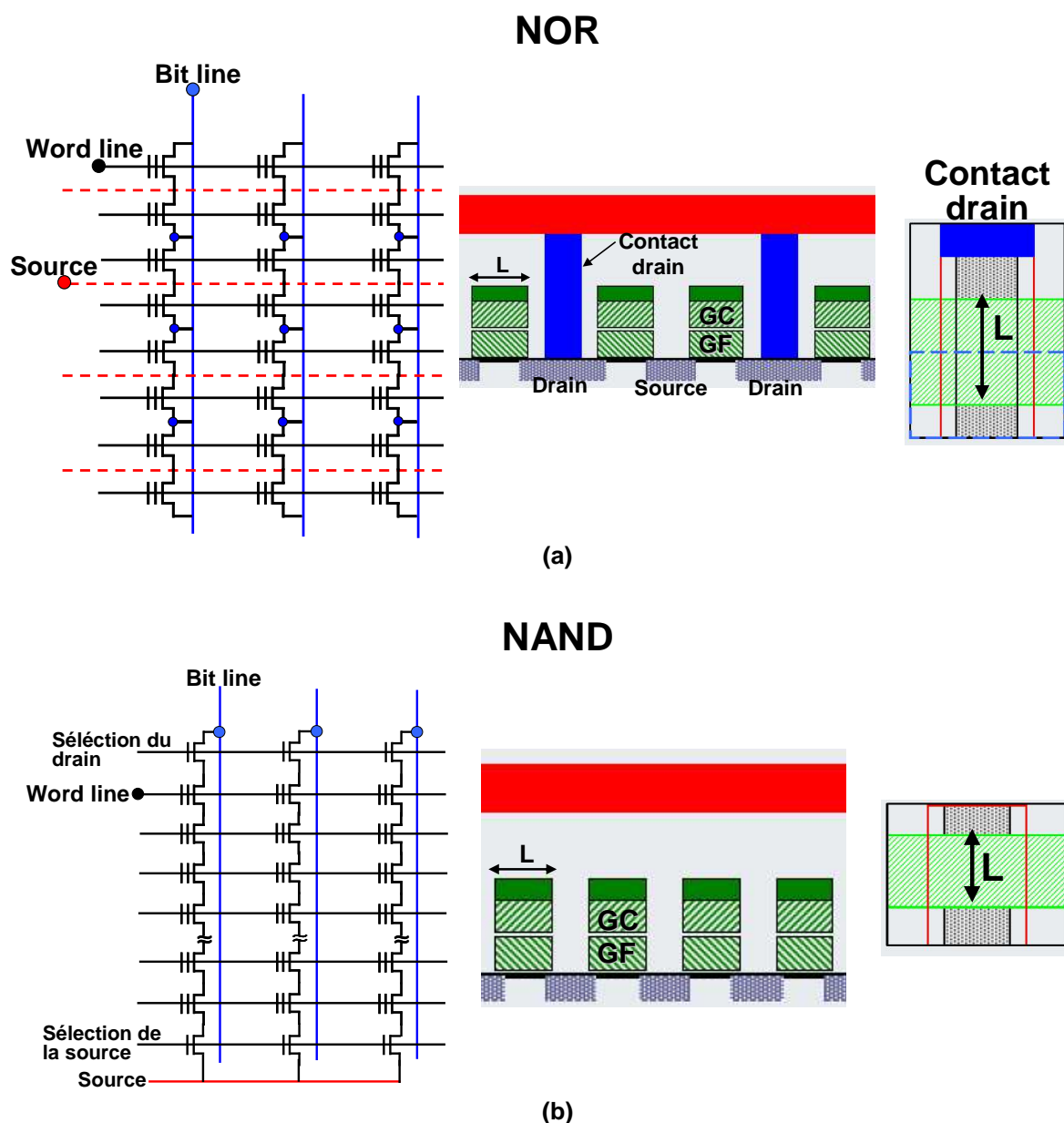


Fig. I-13 : Architecture et coupe longitudinale d'une matrice NOR (a) et NAND (b). Le dessin de masques d'un point mémoire de la matrice est également représenté.

Dans l'architecture NAND, les cellules mémoires sont connectées en série (**Fig. I-13(b)**). Comme la NOR, la grille de contrôle est commune à une « word line », mais le drain des cellules n'est pas adressable. Les cellules ne peuvent donc être écrites et effacées qu'en Fowler-Nordheim. L'accès en lecture à une cellule de la « bit line » s'effectue en polarisant la grille des autres cellules de façon à ce qu'elles soient passantes. C'est pourquoi le temps d'accès est plus long que celui des NOR. Par contre, de part la connexion en série des cellules, la programmation peut être réalisée par blocs, d'où des temps de programmation plus rapides que la NOR. Enfin, comme on peut le voir sur les dessins de masques de la **Fig. I-13**, l'absence de contacts drain permet une diminution de la taille de la Flash NAND (environ 40% par rapport à une Flash NOR). La densité des Flash NAND est donc plus élevée pour un coût par bit plus faible. Elle est donc plutôt utilisée pour le stockage de données (clés USB, cartes mémoires des appareils photos...).

	NOR	NAND
Densité maximale	1 Gb	16 Gb
Accès	Aléatoire	Séquentiel
Temps d'accès	90 ns (accès aléatoire)	7 µs (accès initial) 50 ns (accès série)
Temps d'écriture	8 µs/octet 16 µs/mot	200 µs
Temps d'effacement	1 s	2 ms

Fig. I-14 : Tableau récapitulatif des performances des mémoires Flash de type NOR et NAND.

Le marché des Flash NOR a dominé jusqu'en 2004 et par la suite, le marché des NAND à presque doublé en taille (**Fig. I-15**) grâce notamment, au succès des lecteurs MP3, clés USB et appareils photo numériques. Les flash NAND devraient encore accroître leur domination sur les NOR dans les prochaines années : le marché des NOR devrait croître de 6% d'ici 2011 et celui des NAND de 18% (source : IC Insights).

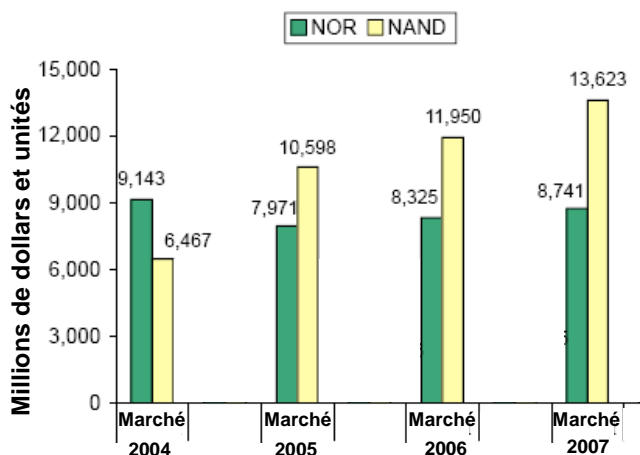


Fig. I-15 : Evolution des revenus du marché des mémoires Flash NOR et NAND : depuis quelques années, les NAND dominent, source : WSTS, IC Insights.

I.5 Limites des mémoires Flash

Bien que connaissant un succès commercial important, les dispositifs mémoires conventionnels à grille flottante ont des difficultés à pousser davantage la miniaturisation.

La réduction des dimensions et l'amélioration des performances sont guidées par l'ITRS (International Roadmap for Semiconductors) [ITRS'06], qui est renouvelée tous les deux ans et prévoit les principales tendances qui seront suivies par l'industrie dans les 15 années suivantes. La dernière édition date de 2005 et une mise à jour a été faite en 2006. On peut voir dans le **Fig. I-16** que toutes les dimensions critiques des cellules vont être réduites, mais certaines d'entre elles sont représentées comme « briques rouges », la solution n'étant toujours pas connue à l'heure actuelle.

Année	2007		2008		2010		2013		2016	
	NOR	NAND	NOR	NAND	NOR	NAND	NOR	NAND	NOR	NAND
Nœud technologique (nm)	65	57	57	51	45	40	32	28	22	20
Taille de la cellule (λ^2)	9-11	4/2	9-12	4/2	9/12	4/1	10-12	4/1	10-13	4/1
Coefficient de couplage	0,6/0,7									
EOT de l'oxyde tunnel (nm)	8-9	6-7	8-9	6-7	8	6-7	8	6-7	7-8	6-7
EOT de l'oxyde de contrôle (nm)	13-15	10-13	13-15	10-13	10-12	10-13	10-12	9-10	8-10	9-10
NOR L _G -empilement (physique- μm)	0,13		0,12		0,11		0,1		0,08	
Tension d'écriture max. (V)	7-9	15-17	7-9	15-17	6-8	15-17	6-8	15-17	6-8	15-17
Endurance (# cycles)	10^5				10^6				10^7	
Rétention (années)	10-20						20			
# bits/cellule max.	2				4					

Des solutions existent
 Solutions connues
 Solutions inconnues

Fig. I-16 : Résumé des prévisions de l'ITRS (mise à jour 2006) dans la catégorie des mémoires Flash. Source [ITRS'06].

Une des limitations concerne la réduction de l'épaisseur de l'oxyde de contrôle. En effet, il est nécessaire de maintenir un couplage suffisant entre la grille de contrôle et la grille flottante, afin que les tensions de programmation ne soient pas trop élevées. On peut voir dans le **Fig. I-16** que le coefficient de couplage doit être maintenu à 0,6/0,7. Ceci est réalisé en diminuant l'épaisseur de l'oxyde de contrôle. Cependant, cette réduction de l'épaisseur est critique car elle entraîne l'augmentation des courants de fuite. On peut d'ailleurs voir qu'aucune solution technologique n'est connue pour réduire l'EOT (Equivalent Oxide Thickness) au-dessous de 10 nm. La solution adoptée dans les mémoires actuelles est d'utiliser comme diélectrique une triple couche oxyde/nitride/oxyde (ONO) [Yim'03] [Park'04]. Mais aujourd'hui, la diminution de l'épaisseur de l'ONO atteint ses limites. C'est pourquoi l'intégration de matériaux diélectriques à haute permittivité ou « high-k » (HfO_2 , Al_2O_3 , HfAlO , HfSiO ...) est envisagée pour remplacer l'interpoly ONO des mémoires Flash standard. Ces matériaux permettent en effet d'obtenir un coefficient de couplage α_G élevé tout en maintenant une bonne rétention des données. Il a été démontré que l'intégration de HfO_2 comme diélectrique de contrôle dans des mémoires standard à grille flottante continue permettait une diminution des tensions de programmation grâce à un meilleur coefficient de

couplage [VanDuuren'06]. De plus, de récents travaux ont montré que l'utilisation d'une grille flottante discrète et d'un interpoly high-k couplée à celle d'une grille de contrôle métallique permettait de réduire le phénomène d'injection des électrons par la grille de contrôle pendant l'effacement [VanDuuren'06] [Lai'06] [Lee'06]. Parmi les différents matériaux étudiés, un intérêt particulier est donné à l'aluminate d'hafnium (HfAlO), qui combine à la fois la constante diélectrique élevée du HfO₂ et la barrière énergétique élevée et la bonne stabilité thermique de l'Al₂O₃ [Molas'07].

Une autre limitation de la miniaturisation (et en parallèle de l'augmentation de la densité) est l'augmentation des interférences entre cellules. Lorsque la densité d'intégration augmente, les cellules se rapprochent et le couplage capacitif entre grilles flottantes augmente (Fig. I-17). Ainsi, la charge stockée dans une grille flottante peut modifier la tension de seuil de la cellule voisine. Pour palier à cela, les industriels ont remplacé le nitrure des espaceurs par du SiO₂ et l'utilisation de matériaux à très basse permittivité dits « low-k » entre les cellules [Kim'07] est envisagée.

La limitation principale de la miniaturisation des mémoires Flash est la réduction de l'épaisseur de l'oxyde tunnel. En effet, il est nécessaire de diminuer cette épaisseur lorsque les dimensions de la cellule diminuent si l'on veut garder un bon contrôle du canal. Mais la diminution de l'épaisseur du diélectrique tunnel affecte la rétention de la charge. La limite intrinsèque de cette épaisseur permettant de respecter le critère de rétention de dix ans est d'environ 6 nm [Lai'98]. En dessous de cette limite, les électrons stockés dans la grille flottante fuient par courant tunnel direct.

En plus de cette limite intrinsèque, il existe une autre limite due au vieillissement prématuré de l'oxyde tunnel à cause des nombreux cycles d'écriture/effacement au cours desquels l'oxyde est soumis à de forts champs électriques (~10 MV/cm). De nombreuses études [Maserjan'82] [Olivo'88] [Naruke'88] ont montré une augmentation du courant de fuite à faible champ à travers l'oxyde après des cycles de programmation successifs. Ce phénomène de dégradation, appelé SILC (Stress Induced Leakage Current), est lié à des défauts dans l'oxyde qui donnent lieu à un courant de fuite important (conduction tunnel assistée par pièges). Etant donné que le SILC augmente lorsque l'épaisseur de l'oxyde tunnel diminue, l'épaisseur de l'oxyde tunnel est limitée à 8 nm (Fig. I-18).

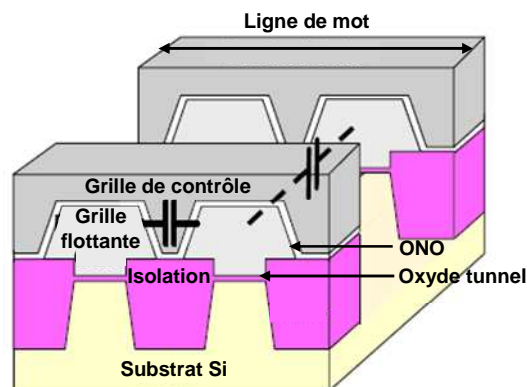


Fig. I-17 : Couplage capacitif entre les grilles flottantes voisines d'une matrice mémoire.

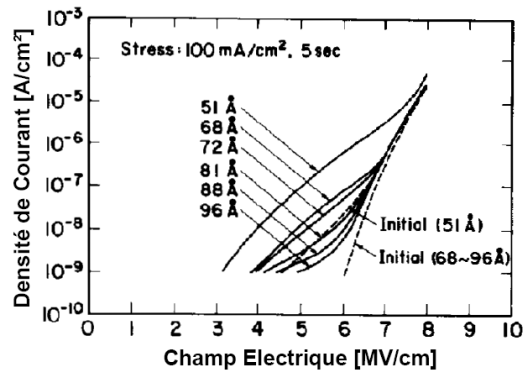


Fig. I-18 : Courant de SILC pour différentes épaisseurs d'oxyde après une contrainte cumulée correspondant à plusieurs milliers d'écritures d'après [Naruke'88].

Outre les limites technologiques, il existe une limite physique intrinsèque. La réduction des dimensions des dispositifs implique que la quantité d'électrons utilisés pour coder un bit (réaliser la séparation entre les deux états logiques de la mémoire) devient de plus en plus faible (**Fig. I-19(a)**). Cette réduction, qui peut aller jusqu'à une dizaine d'électrons pour les générations décananométriques peut induire des problèmes de fiabilité [Molas'04][Molas'06]. Pour les dimensions ultimes, les phénomènes de chargement/déchargement ne peuvent plus être décrits comme continus. Il faut considérer le chargement/déchargement électron par électron. Ces phénomènes stochastiques qui sont aléatoires induisent une dispersion sur le temps de rétention et la fenêtre de programmation (**Fig. I-19(b)**).

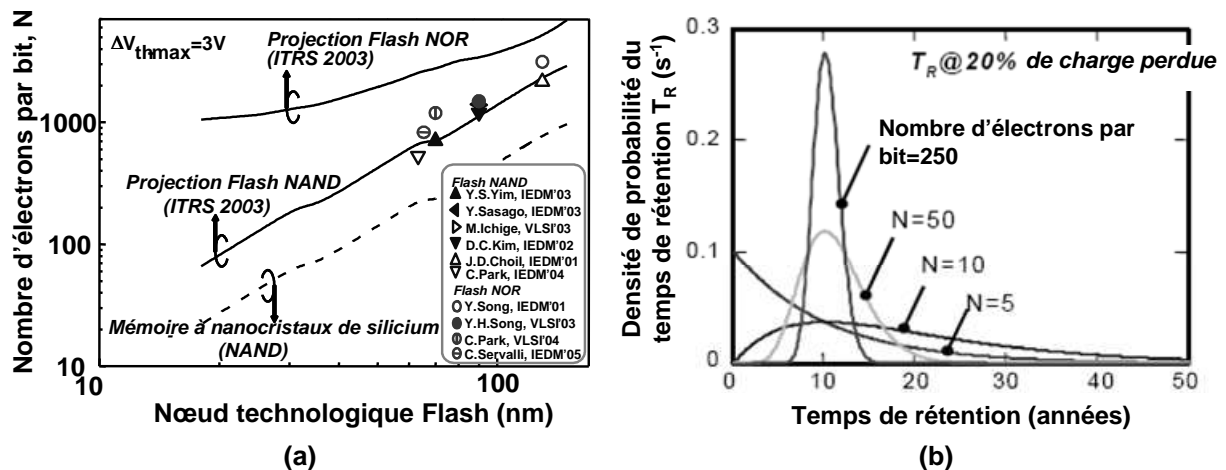


Fig. I-19 : (a) : Représentation du nombre d'électrons par cellule en fonction du nœud technologique Flash. (b) Dispersion du temps de rétention en fonction du nombre d'électrons par bit : plus le nombre d'électrons diminue, plus la dispersion augmente [Molas'04].

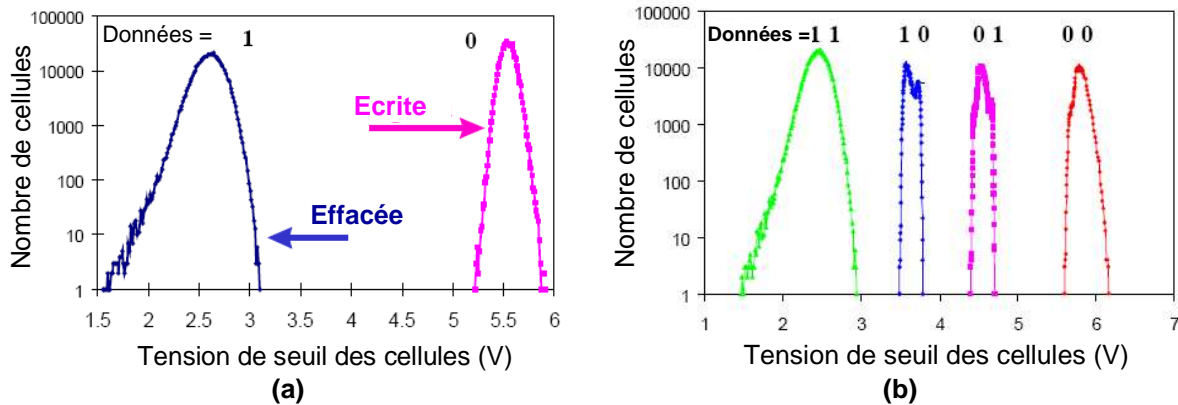


Fig. I-20 : (a) : Distributions des tensions de seuil d'une matrice mémoire à technologie simple niveau : un bit codé par cellule, donc 2 états 0 et 1. (b) : Distributions des tensions de seuil d'une matrice mémoire à technologie multi-niveaux : deux bits codés par cellule, donc 4 états 00, 01, 10 et 11. D'après [Atwood'97].

Il existe une manière artificielle de résoudre le problème de la miniaturisation des mémoires qui est l'utilisation de cellules multi-niveaux. Le principe est de stocker plusieurs bits dans une seule cellule. Ceci est réalisé en stockant plusieurs niveaux de charge dans la grille flottante (2^n niveaux avec n : nombre de bits), ce qui permet d'obtenir autant de niveaux de tensions de seuil (**Fig. I-20**). Ceci permet d'augmenter la capacité de stockage sans diminuer les dimensions de la cellule et de diminuer les coûts de fabrication. Le point critique de cette méthode est qu'il faut être capable de contrôler précisément les niveaux de tensions de seuil et donc la charge injectée au cours de l'écriture ainsi que la dérive au cours de l'endurance, afin que les différents états soient suffisamment différenciables. Cette technologie est en production chez Intel en NOR depuis 1995 sous le nom de StrataFlashTM [Atwood'97] et chez Toshiba et Samsung en NAND.

I.6 Les solutions

Pour les nœuds technologiques sub-45 nm, la miniaturisation des dispositifs mémoires tels que nous les connaissons aujourd'hui semble atteindre ses limites. C'est pourquoi de nombreux efforts se portent sur le développement de nouvelles technologies d'une part et le développement des technologies actuelles grâce à de nouveaux matériaux ou de nouvelles structures d'autre part. Les principales voies de recherche sont listées dans la **Fig. I-21**. Parmi les nouvelles technologies, on peut citer les mémoires FeRAM (Ferroelectric RAM) qui utilisent l'orientation de la polarisation des matériaux ferroélectriques, les MRAM (Magnetoresistive RAM) qui exploitent la polarité magnétique d'un film ferromagnétique, les PCM (Phase Changed Memories) qui sont basées sur le changement de résistivité d'un matériau à changement de phase et à plus long terme les mémoires moléculaires basées sur les mécanismes d'oxydoréduction de molécules électriquement actives. Ces solutions alternatives sont détaillées en annexe.

Dans les sections suivantes, nous présenterons une des alternatives à la miniaturisation qui sont les mémoires à sites de piégeage discrets et en particulier les mémoires à base de nitrure et les mémoires à nanocristaux de silicium.

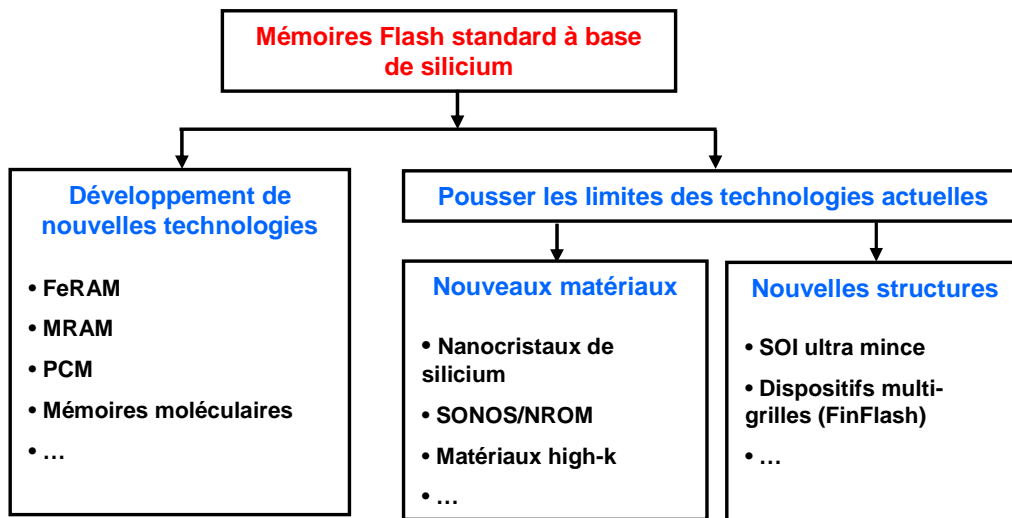


Fig. I-21 : Organisation des activités de recherche sur les mémoires Flash avancées.

1.6.1 Les mémoires Flash à sites de piégeage discrets

L'idée de base est de remplacer la grille flottante continue standard en polysilicium par des sites de stockage discrets, qui peuvent être soit les pièges naturels d'un diélectrique approprié (comme par exemple une couche de nitrure), soit des nanocristaux (en silicium par exemple) [DeSalvo'04]. Ces mémoires offrent plusieurs avantages permettant de repousser les limites liées à la miniaturisation dont le problème du SILC dans l'oxyde tunnel. En effet, le SILC étant dû à une conduction assistée par pièges, dans le cas d'une grille flottante continue, la présence de défauts dans l'oxyde tunnel peut provoquer la fuite de la totalité de la charge stockée dans la grille flottante. Au contraire, avec des sites de stockage discrets isolés les uns des autres, le même défaut n'entraîne que la fuite des électrons piégés dans le site situé au-dessus du défaut (Fig. I-22). Ce concept permet la réduction de l'épaisseur de l'oxyde tunnel sans compromettre la rétention et l'endurance de la cellule mémoire.

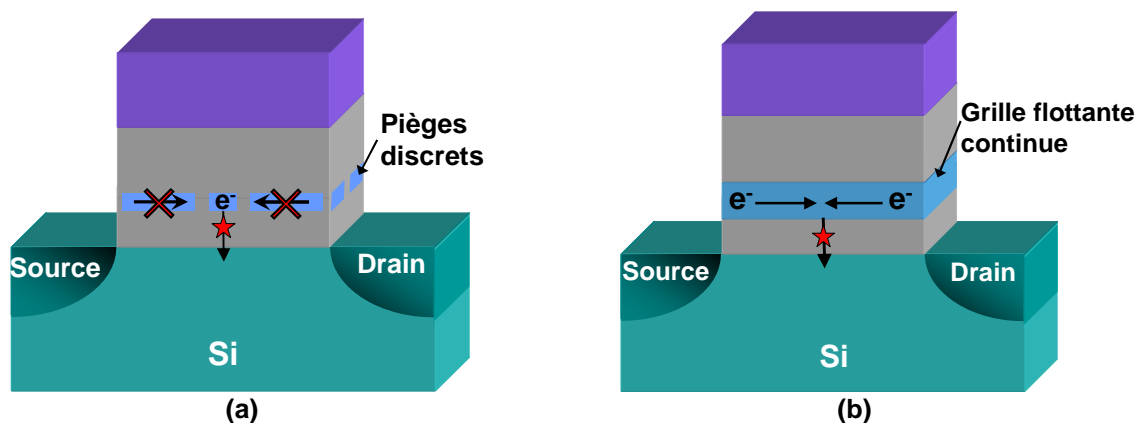


Fig. I-22 : Amélioration de l'immunité au SILC des mémoires à sites de piégeage discrets (a) par rapport aux mémoires à grille flottante continue (b).

Une autre particularité des mémoires à sites de stockage discrets programmées par porteurs chauds est la localisation de la charge. Ceci permet le stockage de deux bits dans la même cellule. En effet, si la cellule est écrite par électrons chauds avec la source à la masse, les électrons sont injectés au niveau de la jonction canal/drain, là où le champ électrique latéral est le plus élevé. L'absence de conduction latérale entre les sites de stockage induit une localisation des électrons injectés au-dessus de la jonction drain. Les deux principales catégories de mémoires à sites de piégeage discrets sont les mémoires à couche de piégeage en nitrure (SONOS et NROM) et les mémoires à nanocristaux de silicium. Elles sont détaillées dans les paragraphes suivants.

1.6.1.1 Les mémoires à base de nitrure

Les deux principaux types de mémoires à couche de piégeage en nitrure sont la SONOS et la NROM. La couche de piégeage est du nitrure de silicium Si_3N_4 .

La SONOS est à l'étude depuis de nombreuses années [Libsch'87][Wann'95] et son intérêt s'est accru ces dernières années [Swift'02]. Ses avantages par rapport aux mémoires à grille flottante en polysilicium sont son immunité face aux défauts dans l'oxyde tunnel et donc une bonne tolérance face aux radiations et des tensions de programmation réduites. De plus, la densité de pièges est assez élevée par rapport à celle des nanocristaux de silicium ($\sim 10^{13} \text{ cm}^{-2}$) et enfin, le procédé de fabrication est simple et compatible avec le procédé CMOS.

On peut voir une structure SONOS typique dans la **Fig. I-23(a)** : l'oxyde tunnel est très fin ($\sim 2 \text{ nm}$), l'oxyde de contrôle également ($\sim 4 \text{ nm}$) et la couche de Si_3N_4 a une épaisseur d'environ 4 nm . L'oxyde tunnel très fin permet l'écriture et l'effacement en mode tunnel direct. Des caractéristiques typiques de programmation sont présentées dans la **Fig. I-23(b)**.

Cependant, cette structure a plusieurs inconvénients. Un des problèmes majeurs concerne l'effacement au cours duquel des électrons sont injectés dans le nitrure depuis la grille de contrôle. Ceci entraîne la saturation des caractéristiques d'effacement en fonction de la tension appliquée sur la grille (**Fig. I-23(c)**). Le second problème est la rétention en température. En effet, au-delà de 85°C , les 10 années de rétention ne sont plus garanties (**Fig. I-23(d)**).

Le second type de mémoire à base de nitrure est la NROM qui a été introduite par Saifun en 1999 [Eitan'99]. Comme la SONOS, le stockage des charges s'effectue dans une couche de nitrure située entre deux couches d'oxyde (**Fig. I-24(a)-(b)**). L'écriture est réalisée par injection d'électrons chauds (CHE) et l'effacement par l'injection de trous chauds (HHI), ce qui permet l'utilisation d'un oxyde tunnel plus épais ($\sim 5 \text{ nm}$) et donc une meilleure rétention. En intervertissant le rôle de source et drain, il est possible de stocker physiquement 2 bits par cellule, grâce à la localisation de la charge près des jonctions (**Fig. I-24(a)**). Ceci permet d'augmenter la densité de stockage tout comme le font les cellules multi-niveaux. Comme la SONOS, la NROM se caractérise par son immunité face aux défauts dans l'oxyde

tunnel. De plus, le problème de saturation des caractéristiques d'effacement est supprimé puisque l'effacement est réalisé par trous chauds.

La nouvelle génération de NROM présentée en 2005 [Eitan'05] propose le stockage de 4 bits par cellule en ajoutant une programmation multi-niveaux sur chacun des 2 bits physiques : on a donc deux bits stockés de chaque côté du canal (Fig. I-24(c)). L'obtention de 4 bits suffisamment séparés est obtenue grâce à un algorithme de programmation approprié (Fig. I-24(d)).

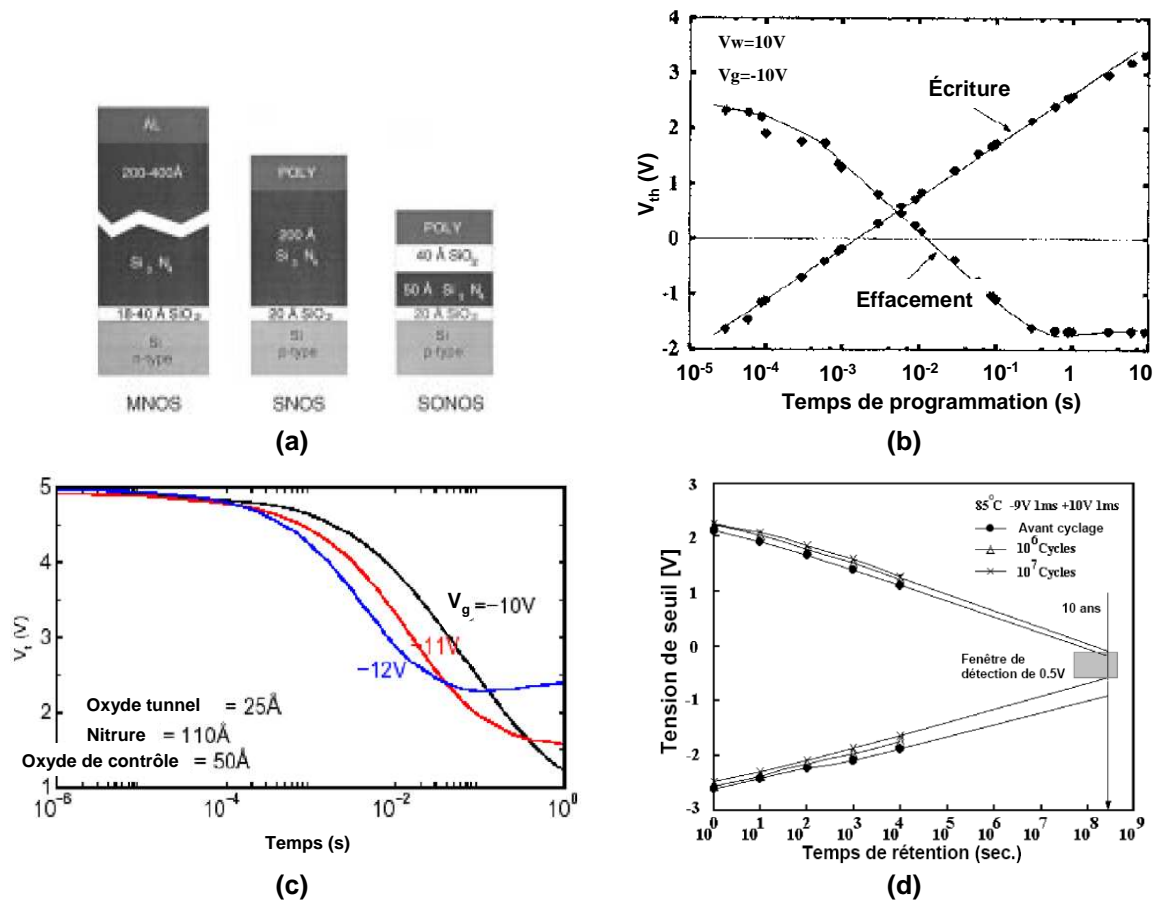


Fig. I-23 : (a) : Evolution des mémoires non-volatiles à base de nitride. (b) : Caractéristiques d'écriture et d'effacement d'une mémoire SONOS avec un oxyde tunnel de 2 nm, une couche de nitrure de 5 nm et un oxyde de contrôle de 3,5 nm d'après [White'97]. (c) : Saturation des caractéristiques d'effacement en fonction de la tension appliquée sur la grille d'après [Steimle'07]. (d) : Rétention d'une mémoire SONOS à 85°C avant et après endurance : après 10⁷ cycles d'écriture/effacement et 10 ans, la fenêtre de programmation n'est plus que de 0,5V d'après [Bu'02].

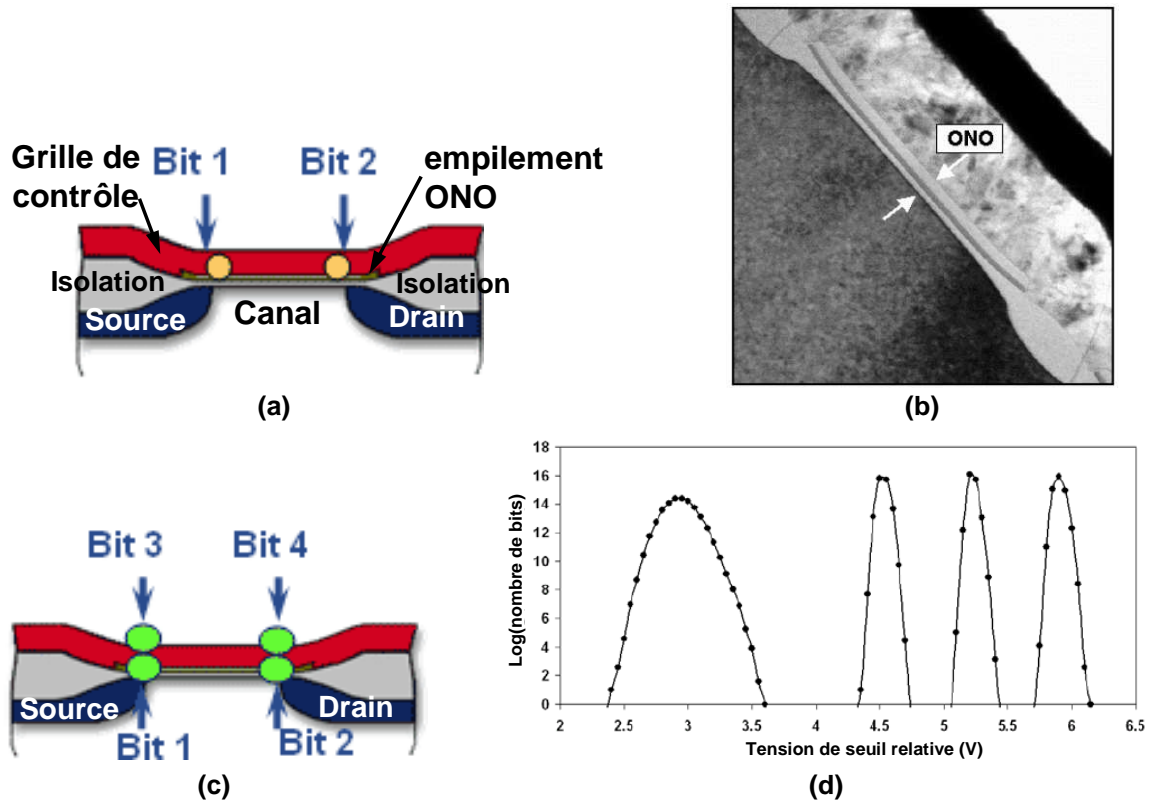


Fig. I-24 : Schéma (a) et coupe TEM (b) d'une cellule NROM le long du canal. Représentation schématique du stockage de 4 bits par cellule (c) et distributions de la tension de seuil mesurées sur un produit 4 bits/cellule (nœud technologique 130 nm) [Eitan'05].

Nous avons vu les deux principales structures de base utilisant le nitrure comme couche de stockage. Toutefois, une autre structure mémoire à base de nitrure développée récemment pour des applications NAND haute densité mérite également d'être citée : c'est la cellule TANOS (TaN/AIO/Nitrure/Oxyde/Silicium) (**Fig. I-25(a)**) [Kim'07]. Cette structure est une amélioration de la mémoire SONOS, avec une grille métallique en TaN et un diélectrique de contrôle « high-k » en AIO. La bonne stabilité thermique de ces deux matériaux et le travail de sortie de la grille relativement élevé (~4,8 V) permettent d'améliorer le temps d'effacement par rapport à une mémoire de type SONOS en réduisant le phénomène d'injection d'électrons depuis la grille de contrôle. Samsung a présenté en 2006 une mémoire NAND TANOS (SiO₂ 4 nm, SiN LPCVD 7 nm, Al₂O₃ ALCVD 15 nm, TaN 17 nm) de 32 Gb multi-niveaux avec une technologie 40 nm [Park'06]. Une fenêtre de programmation de 3 V est obtenue avec une écriture et un effacement (en mode Fowler-Nordheim) de respectivement 100 μs à 17 V et 10 ms à -19 V (**Fig. I-25(b)**). Les mémoires de type TANOS pourraient remplacer les mémoires à grille flottante à partir du nœud technologique 30 nm car leurs dimensions sont plus faciles à réduire. Cependant, des recherches sont nécessaires afin d'améliorer les problèmes de rétention et la difficulté à effacer qui subsistent encore.

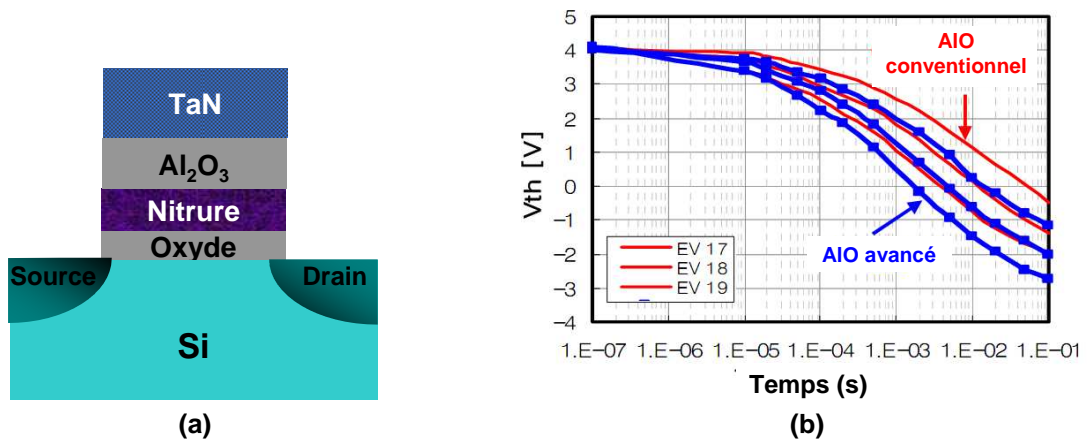


Fig. I-25 : (a) Schémas de la structure mémoire TANOS. (b) : Dynamiques d'effacement de cellules TANOS en fonction de l'optimisation de la couche d'AIO [Park'06].

1.6.1.2 Les mémoires à nanocristaux de silicium

Le second type de mémoires à sites de stockage discrets sont les mémoires à nanocristaux de silicium (Si-NCs). Le principe est similaire à celui des mémoires à grille flottante, à la différence que ce sont des nanocristaux de silicium qui constituent la couche de piégeage. Si les nanocristaux sont suffisamment espacés, il n'y a pas de conduction latérale entre eux. Il existe plusieurs méthodes d'élaboration qui sont détaillées au chapitre II. L'une d'entre elles, le dépôt LPCVD (Low Pressure Vapor Deposition) en deux étapes permet un bon contrôle de la taille et de la densité des Si-NCs (Fig. I-32).

Les cellules sont programmées soit par porteurs chauds, soit par Fowler-Nordheim, selon les applications visées.

Les mémoires à Si-NCs ont été présentées pour la première fois en 1995 par [Tiwari'95] avec comme objectif de répondre aux applications de types DRAM volatiles et l'avantage de n'avoir qu'un seul transistor par cellule. Par la suite, d'autres groupes ont mis en évidence l'intérêt d'utiliser les nanocristaux pour les mémoires non-volatiles. Des matrices de 4 Mb basées sur cette technologie ont été démontrées [Muralidhar'03]. La société Freescale a annoncé fin novembre 2005, qu'elle avait réussi à intégrer cette solution sous la forme d'une matrice 24 Mb.

Ces dispositifs apportent de nombreux avantages.

Elles sont tout d'abord plus robustes aux défauts du diélectrique tunnel, ce qui permet de réduire l'oxyde tunnel jusqu'à environ 5 nm d'épaisseur (par rapport aux 7-8 nm dans le cas de la grille flottante continue) tout en garantissant 10 années de rétention, ce qui rend possible une réduction des tensions de fonctionnement.

Ensuite, ces dispositifs présentent une bonne rétention à haute température, étant donné que la perte de charges stockées dans la bande de conduction des Si-NCs n'est pas activée thermiquement.

Le fait que les nanocristaux soient physiquement isolés les uns des autres, permet de stocker des charges dans des sites séparés (côté drain et côté source), ce qui donne lieu au

stockage de deux bits par cellule. De plus, comme l'isolant entre nanocristaux est du SiO₂, la migration de charges entre sites de piégeage est fortement réduite par rapport au cas d'une couche en nitrure.

Le caractère discret de la couche de stockage et le diamètre des nanocristaux (<10 nm) étant bien inférieur à l'épaisseur d'une couche de polysilicium (~200 nm), le couplage capacitif entre cellules voisines est fortement réduit.

Le procédé de fabrication est totalement compatible avec le procédé CMOS. De plus, Freescale a annoncé que plusieurs niveaux de masques pouvaient être supprimés grâce aux faibles tensions de programmation qui permettent de se passer des transistors haute tension [Swift'02].

Une fenêtre de programmation de 3 V peut être obtenue avec des temps de programmation relativement courts et des tensions faibles. Cependant, il faut noter que les mémoires à Si-NCs présentent aussi quelques inconvénients. Une des principales limitations réside dans le fait que le couplage entre les Si-NCs et la grille de contrôle est faible ($\alpha_G \sim 0,3$) à cause du taux de recouvrement de la grille de contrôle sur la surface des Si-NCs qui est plus faible que dans le cas d'une grille flottante continue. Ceci pose surtout un problème lors de l'écriture par Fowler-Nordheim et donc pour les applications NAND. Une solution à ce problème peut être l'intégration de matériaux high-k dans l'oxyde de contrôle, ce qui permet de réduire les fuites et d'améliorer le couplage. Une autre limitation des mémoires à nanocristaux de silicium est la faiblesse de la fenêtre de programmation (~3 V), qui peut poser problème pour une programmation multi-niveaux. Pour ces applications, différentes technologies sont actuellement à l'étude comme les nanocristaux auto-organisés [Guarani'03] ou métalliques [Liu'02-a][Liu'02-b].

Nous allons maintenant nous intéresser plus particulièrement aux études réalisées en ce qui concerne les matrices et produits mémoires à nanocristaux de silicium. Les deux principaux industriels, outre Atmel, ayant publié des résultats sur des matrices de plusieurs Mbits sont Freescale et STMicroelectronics. Les publications de Freescale seront détaillées dans le premier paragraphe, suivi par les résultats obtenus par STMicroelectronics dans le deuxième paragraphe.

1.6.1.2.A Résultats publiés par Freescale

L'intégration des nanocristaux de silicium dans des matrices mémoires de grandes capacités a été présentée pour la première fois par Freescale en 2003 [Muralidhar'03]. Une mémoire Flash embarquée de 4 Mb en technologie 90 nm a été réalisée. Les cellules mémoires ont une longueur de grille de 0,12 μm , la densité de Si-NCs est comprise entre $5.10^{11} \text{ cm}^{-2}$ et $1,1.10^{12} \text{ cm}^{-2}$ et leur diamètre varie entre 3 et 7 nm. L'oxyde tunnel fait environ 5 nm et l'oxyde de contrôle est un HTO (High Temperature Oxide) d'environ 10 nm. Les dynamiques d'écriture/effacement de la cellule sont représentées sur la **Fig. I-26(a)**.

L'écriture est réalisée en porteurs chauds et l'effacement en Fowler-Nordheim. Une fenêtre de programmation de 3 V peut être obtenue avec des tensions relativement faibles. Cependant, l'endurance montre une augmentation de la tension de seuil d'environ 1 V après 100K cycles pour les états écrit et effacé, ce qui traduit un piégeage parasite dans l'oxyde de contrôle et une dégradation de la pente sous le seuil **Fig. I-26(b)**. L'optimisation des diélectriques est donc très importante.

En ce qui concerne l'intégration des Si-NCs avec un procédé CMOS, il est indiqué qu'il est possible de réduire le nombre de masques additionnels par rapport à un procédé grille flottante standard (**Fig. I-27(a)**). Les distributions écrites et effacées de la matrice 4 Mb sont présentées dans la **Fig. I-27(b)**. La fenêtre de programmation est de 3 V, et on peut voir que les distributions sont assez étroites (entre 1,2 et 1,8 V). Toutefois, les conditions de programmation et l'utilisation ou non d'algorithmes d'écriture/effacement ne sont pas précisées.

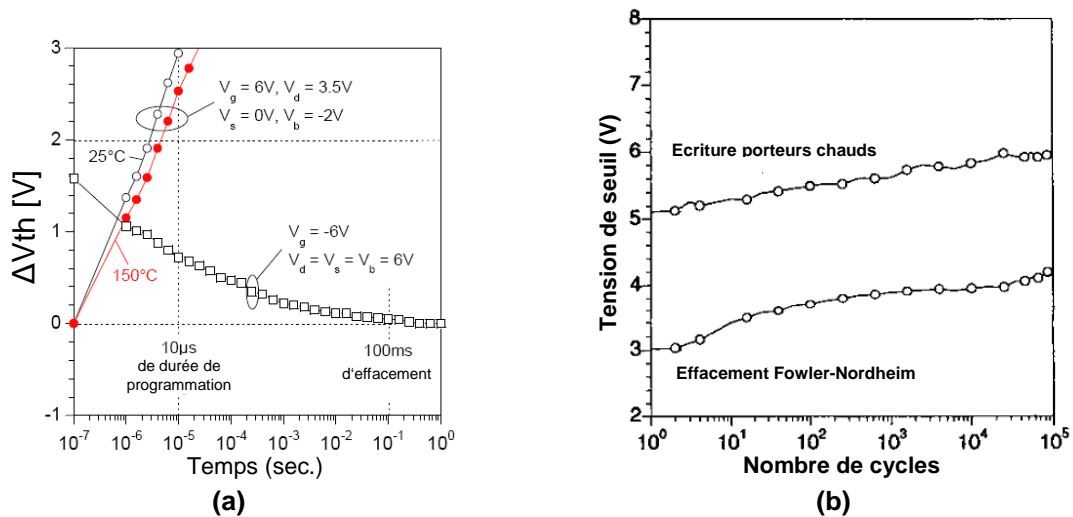
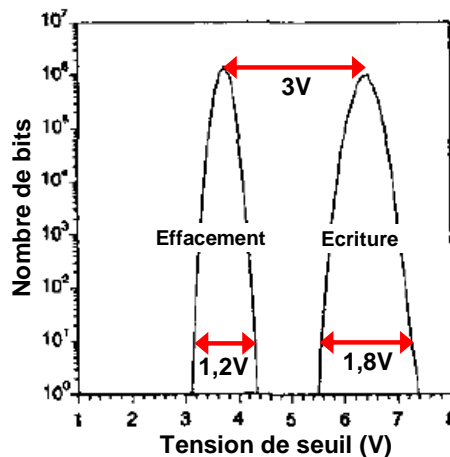


Fig. I-26: (a) : Dynamiques d'écriture en porteurs chauds et d'effacement en Fowler-Nordheim de la cellule à Si-NCs. (b) : Courbes d'endurance d'une cellule : les V_{th} écrits et effacés augmentent sans fermeture de la fenêtre [Muralidhar'03].

Etape du procédé de fabrication	Logique	Grille flottante	Nanocristaux
Isolation	•	•	•
Implants haute tension		2 masques	
Implant mémoire		1 masque	1 masque
Oxyde tunnel		•	
Gravure grille flottante		1 masque	
Gravure ONO/nanocristaux		1 masque	1 masque
Implants basse tension	•	•	•
Implants DGO	•	•	•
Oxydation/gravure haute tension		1 masque	1 masque
Oxydation/gravure DGO	•	•	•
Oxyde basse tension	•	•	•
Dépôt grille de contrôle	•	•	•
Gravure de l'empilement de grille mémoire		1 masque	
Implant halo source mémoire		1 masque	
Implant drain mémoire		1 masque	1 masque
Gravure de l'empilement de grille périphérie	•	•	•
Implants LDD haute tension		2 masques	
Implants LDD DGO	•	•	•
S/D et backend	•	•	•
Ajout de masque		+11	+4

(a)



(b)

Fig. I-27 : (a) : Masques supplémentaires d'un procédé de fabrication d'une mémoire non volatile standard et à Si-NCs par rapport à un procédé CMOS. (b) : Distribution écrite et effacée de la matrice 4 Mb. [Muralidhar'03].

Comme nous l'avons expliqué précédemment, la qualité du diélectrique de contrôle est très importante pour la fiabilité de la cellule. C'est pourquoi Freescale a réalisé un comparatif entre deux diélectriques de contrôle : ONO et HTO [Steimle'04]. Les mémoires Flash standard actuelles utilisent l'ONO, car le nitrure constitue une barrière à l'oxydation très efficace et les fuites à travers cet empilement sont moins importantes qu'avec un HTO. L'utilisation d'un ONO simplifie donc l'intégration des Si-NCs, car elle empêche l'oxydation des Si-NCs au cours de la croissance des oxydes de grille de la périphérie CMOS. Ainsi, la partie mémoire peut être réalisée indifféremment avant ou après la périphérie. Par contre,

l'inconvénient est qu'un piégeage parasite et irréversible d'électrons peut avoir lieu dans le nitrure lors de l'écriture. C'est ce qui est montré dans la **Fig. I-29**. Les caractéristiques d'endurance (**Fig. I-29(a)-(b)**) mettent en évidence la différence entre HTO et ONO : dans le cas du HTO, les tensions de seuil restent constantes jusqu'à 1000 cycles, puis elles augmentent ensuite à cause du piégeage parasite dans les diélectriques. Dans le cas de l'ONO, les courbes montent continuellement, traduisant un piégeage des électrons dans le nitrure pendant l'écriture qui ne se dépiègent pas lors de l'effacement. Les **Fig. I-29 (c) et (d)** montrent les caractéristiques de rétention à différentes températures pour les deux diélectriques. Alors que le HTO présente une très bonne rétention, l'ONO montre des caractéristiques typiques des mémoires à nitrure, c'est à dire une perte de charge qui s'aggrave avec la température, ce qui démontre encore une fois qu'une partie de la charge est injectée dans la couche de nitrure. Il a également été établi que ce piégeage irréversible induisait des distributions écrites et effacées de l'ONO plus larges que celles du HTO (**Fig. I-29(a)-(b)**).

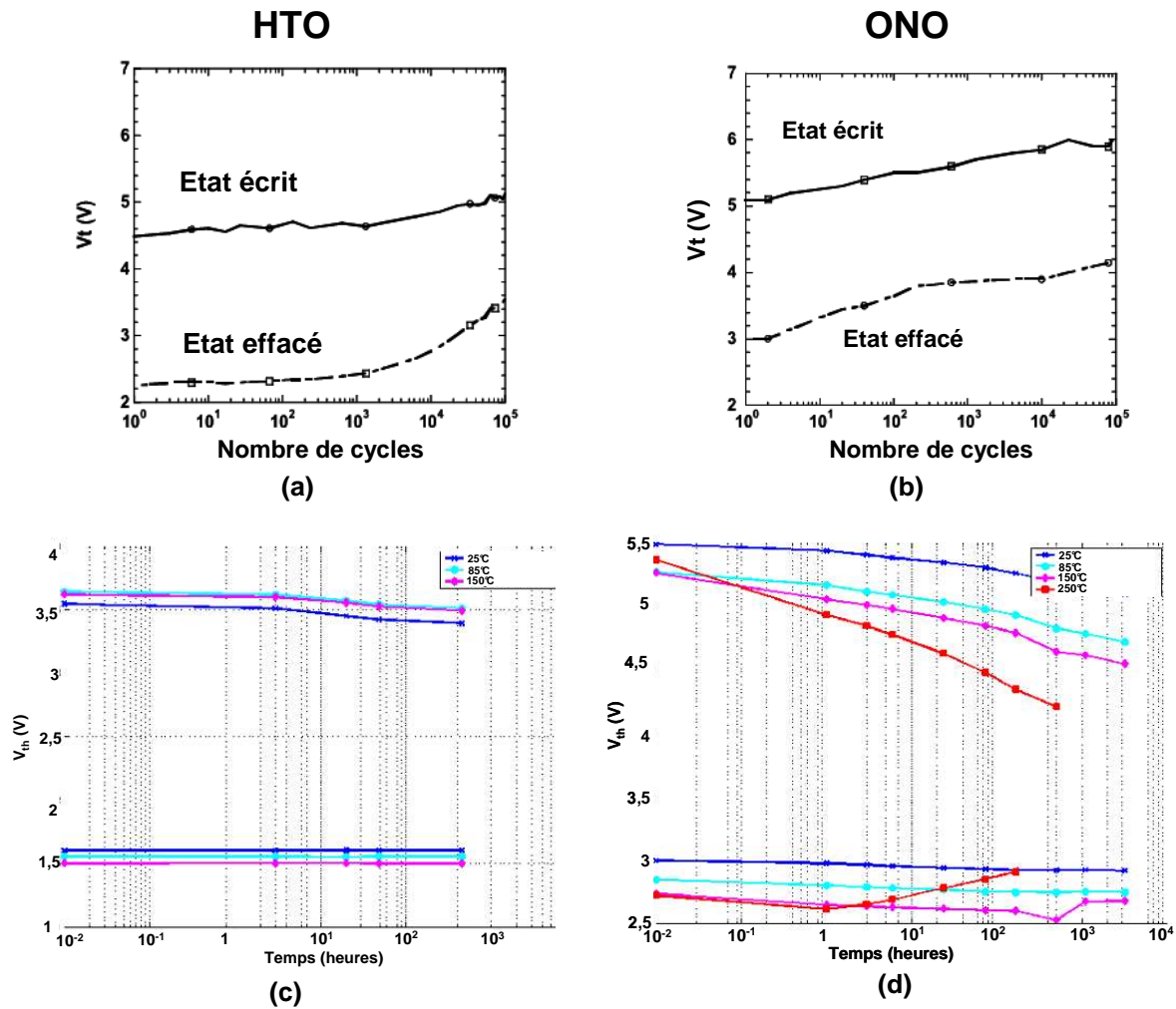


Fig. I-28 : Courbes d'endurance d'une cellule avec HTO (a) et ONO (b). Caractéristiques de rétention à différentes températures 25°C, 85°C, 150°C et 250°C pour une cellule avec HTO (c) et ONO (d). [Steimle'04].

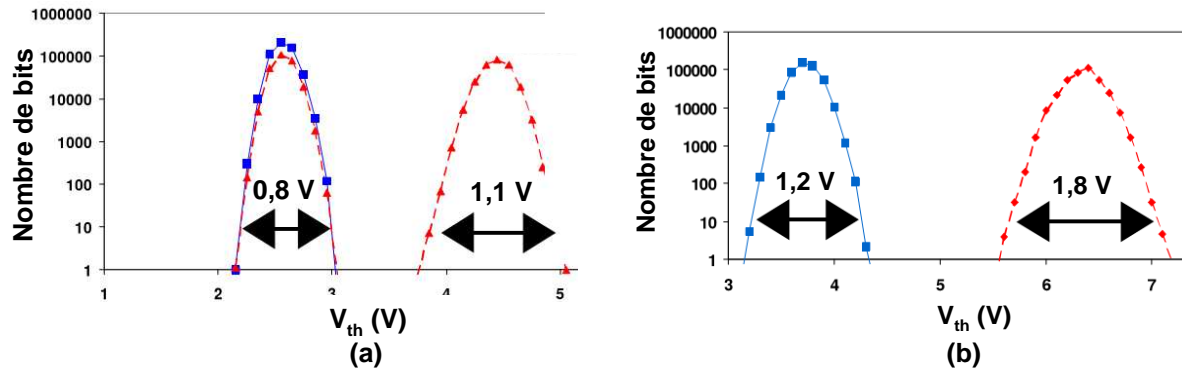


Fig. I-29 : Distributions écrites et effacées d'un secteur de 512 Kb avec HTO (a) et ONO (b). [Steimle'04].

Des études sur l'influence de la densité et de la taille des Si-NCs sur les caractéristiques des cellules et des matrices ont également été présentées [Rao'05] [Gasquet'06]. On peut voir des images MEB des nanocristaux de différentes tailles et densités sur la Fig. I-30 (a). Sur les trois échantillons, deux ont environ la même densité et un diamètre différent ($1.10^{12}/\text{cm}^2$ 36 Å et $8.10^{11}/\text{cm}^2$ 62 Å), c'est-à-dire deux taux de couverture différents. Deux autres ont le même taux de couverture et des diamètres et des densités différentes ($8.10^{11}/\text{cm}^2$ 62 Å et $2.10^{11}/\text{cm}^2$ 120 Å). Les Fig. I-30 (b)-(c) montrent que le taux de couverture des Si-NCs a une influence sur la vitesse d'écriture et d'effacement et sur le niveau de saturation de l'écriture. Plus le taux de couverture est faible, plus le niveau de tension en saturation en écriture est bas et plus l'écriture est lente. Par contre, avec un taux de couverture faible, l'effacement est plus rapide.

Des mesures de rétention à 200°C ont été réalisées sur une matrice 4 Mb pour différentes tailles de Si-NCs (en fait, un taux de couverture des Si-NCs croissant) avec un oxyde tunnel de 5 nm et un HTO de 10 nm. (Fig. I-30 (d)). A noter que les nanocristaux ont été passivés sous NO. La Fig. I-30 (e) ne montre pas d'influence de la taille des Si-NCs sur la rétention des données. Quant à la largeur des distributions écrites et effacées, on observe une augmentation de la distribution écrite lorsque le diamètre des Si-NCs augmente, surtout à partir de 7 nm. La distribution effacée s'élargit également, mais de façon moins importante (Fig. I-30 (f)-(g)). Enfin, concernant l'endurance, une légère amélioration est obtenue avec les Si-NCs les plus gros (Fig. I-30 (h)). On note tout de même une augmentation importante du V_{th} (entre 1,5 et 2 V).

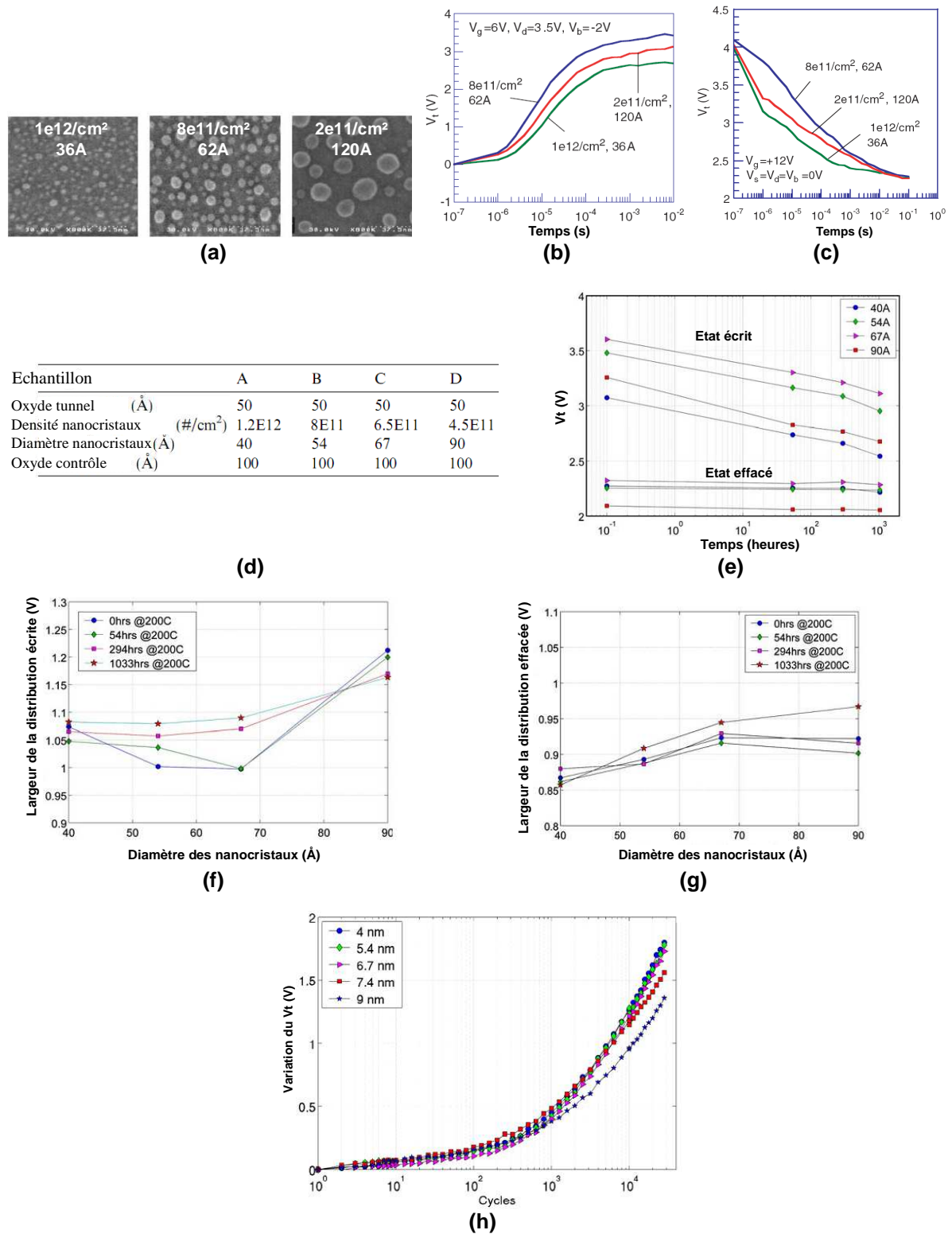


Fig. I-30 : (a) : Images MEB de Si-NCs déposés sur SiO₂ avec trois conditions différentes : 1E12 cm⁻² et 3,6 nm ; 8E11 cm⁻² et 6,2 nm ; 2E11 cm⁻² et 12 nm. (b) et (c) : Dynamiques d'écriture sur cellule en porteurs chauds et d'effacement en Fowler-Nordheim (par l'oxyde de contrôle) pour les trois conditions de taille et densité des Si-NCs de (a). (d) : Tableau descriptif des quatre échantillons utilisés pour les mesures (e), (f), (g). (e) : Rétention à 200°C sur l'état écrit et effacé sur une matrice 4 Mb (8 secteurs de 512 Kb). Le graphique représente la moyenne sur les 8 secteurs des valeurs des bits les moins écrits et les moins effacés. (f) et (g) : Largeur des distributions écrite et effacée pour les quatre échantillons après rétention à 200°C. (h) : Augmentation de la tension de seuil des bits les moins effacés de la distribution effacée au cours de l'endurance. [Rao'05] [Gasquet'06].

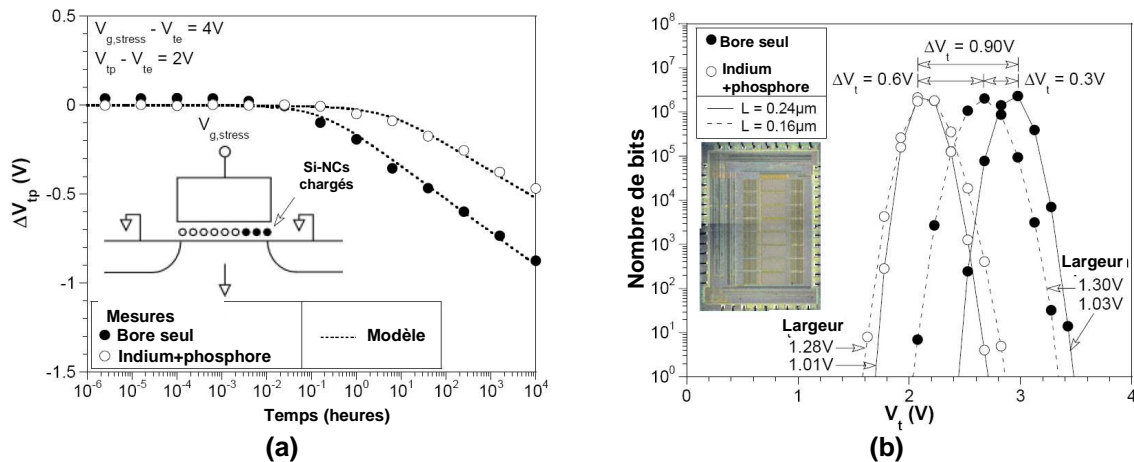


Fig. I-31 : (a) : Gate disturb accéléré sur une cellule écrite : la tension de seuil diminue lorsque l'on applique une tension positive sur la grille. Le fait de diminuer la tension de seuil initiale de la cellule en contre-dopant le canal permet de diminuer la perte de charge au cours de la lecture. (b) : Distributions effacées de la matrice 4 Mb en fonction des implants canal et de la longueur de grille. Le contre-dopage en surface abaisse la tension de seuil effacée de 900 mV. [Gasquet'06].

Enfin, une attention particulière a été portée au « gate disturb », c'est-à-dire aux perturbations subies par les cellules mémoires écrites pendant la lecture [Steimle'04] [Gasquet'06]. En effet, l'épaisseur réduite du HTO de contrôle peut entraîner une fuite des électrons stockés dans les Si-NCs à travers celui-ci, en particulier au cours de la lecture pendant laquelle une tension positive est appliquée sur la grille. Une solution pour diminuer le gate disturb est de diminuer la tension de seuil de l'état écrit, ce qui permet d'abaisser la tension appliquée sur la grille lors de la lecture et donc la perte de la charge stockée. Cependant, si on veut garder une fenêtre de programmation identique, il est nécessaire d'abaisser également la tension de seuil initiale de la cellule mémoire. C'est ce qu'a réalisé Freescale en contre-dopant la surface du canal (Fig. I-31) [Gasquet'06].

I.6.1.2.B Résultats publiés par STMicroelectronics

Les premiers résultats ont été publiés en 2003 avec la collaboration du CEA-Leti [DeSalvo'03], sur une structure CAST (Cell Array Stress Test) 1 Mb (Fig. I-32(a)). Les caractéristiques d'écriture/effacement en Fowler-Nordheim de la matrice 1 Mb sont représentées sur la Fig. I-32(b). Dans cette étude, les Si-NCs ont été déposés par LPCVD (Low Pressure Vapor Deposition) en deux étapes, ce qui permet un bon contrôle de la taille et de la densité des Si-NCs (Fig. I-32(c)-(d)).

Ensuite, STMicroelectronics a présenté des résultats sur une mémoire Flash NOR 16 Mb à nanocristaux de silicium en 2006 et 2007 [Gerardi'06] [Gerardi'07-a] [Gerardi'07-b]. La matrice 16 Mb est divisée en 32 secteurs de 512 Kb. Les cellules mémoires correspondent au nœud technologique 150 nm. L'oxyde tunnel fait 5 nm et le diélectrique de contrôle est un ONO avec une EOT de 12 nm. Les Si-NCs sont déposés par CVD avec une densité d'environ $5.10^{11}/cm^2$ et deux diamètres différents : 3 et 6 nm. La Fig. I-33(a) montre les distributions des tensions de seuil des états écrits et effacés de la matrice 16 Mb. Une fenêtre de programmation de 3 V est obtenue et les distributions sont assez étroites. La distance entre les

bits les moins effacés et les moins écrits est d'environ 800 mV. La **Fig. I-33(b)** montre un élargissement de la distribution effacée avec le diamètre des Si-NCs. De plus, l'utilisation de l'ONO comme diélectrique de contrôle pose les mêmes problèmes qu'avait rencontrés Freescale, c'est-à-dire le piégeage d'électrons dans le nitrure lors de l'écriture par porteurs chauds. La difficulté à revenir au niveau initial pendant l'effacement (**Fig. I-33(c)**) et le décalage des distributions écrites et effacées vers les tensions de seuil plus élevées au cours de l'endurance (**Fig. I-33(d)**) démontrent ce piégeage parasite.

Un autre problème apparaît sur la **Fig. I-33(e)** : la distribution effacée s'élargit de plus en plus au cours des cycles d'endurance et on peut voir l'apparition d'une queue de distribution au niveau des bits les moins effacés. Gerardi [**Gerardi'06**] affirme que ce problème de fiabilité est dû à l'amincissement de l'oxyde tunnel au niveau des coins de l'active, généré par la présence de Si-NCs de grande taille à cet endroit (Insert **Fig. I-33(e)**).

Enfin, la rétention à 250°C après 1000 cycles (**Fig. I-33(f)**) montre une forte chute de la tension de seuil après les premières heures puis une diminution beaucoup plus faible par la suite. La forte chute peut s'expliquer par le dépiégeage des électrons du nitrure de l'ONO qui se produit dès le début de la rétention. Après le déchargement, on retrouve une pente moins élevée correspondant à la perte de la charge des Si-NCs [**Gerardi'07-a**]. On peut noter qu'après 10 ans à 250 °C, les queues des distributions écrites et effacées (bits les moins écrits et bits les moins effacés) sont encore dissociés.

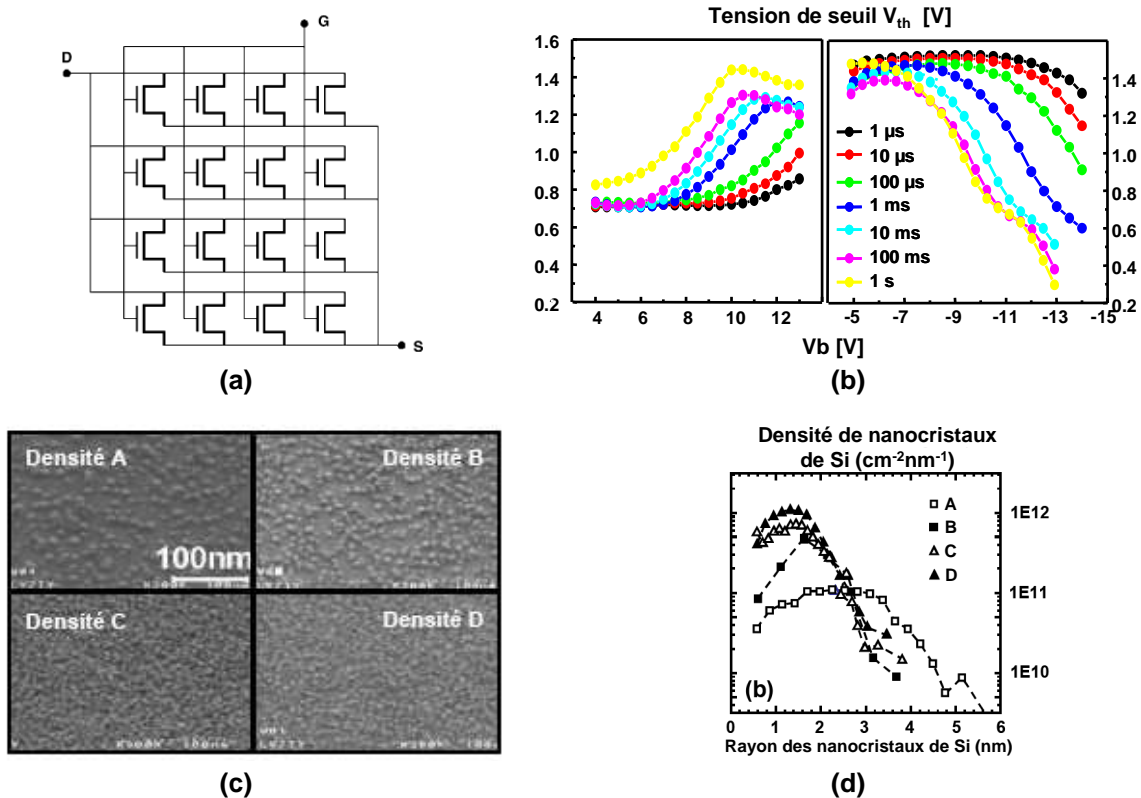


Fig. I-32 : (a) : Dessin schématique de la structure CAST (Cell Array Stress Test) : les cellules mémoires sont connectées en parallèle et les grilles, drains et sources sont reliés entre eux. (b) : Caractéristiques de rétention à 250°C de la structure CAST 1 Mb. (c) : Images MEB de Si-NCs avec des densités croissantes. (d) : Distributions en taille des nanocristaux mesurées par EFTEM (Energy Filtered TEM). [**DeSalvo'03**].

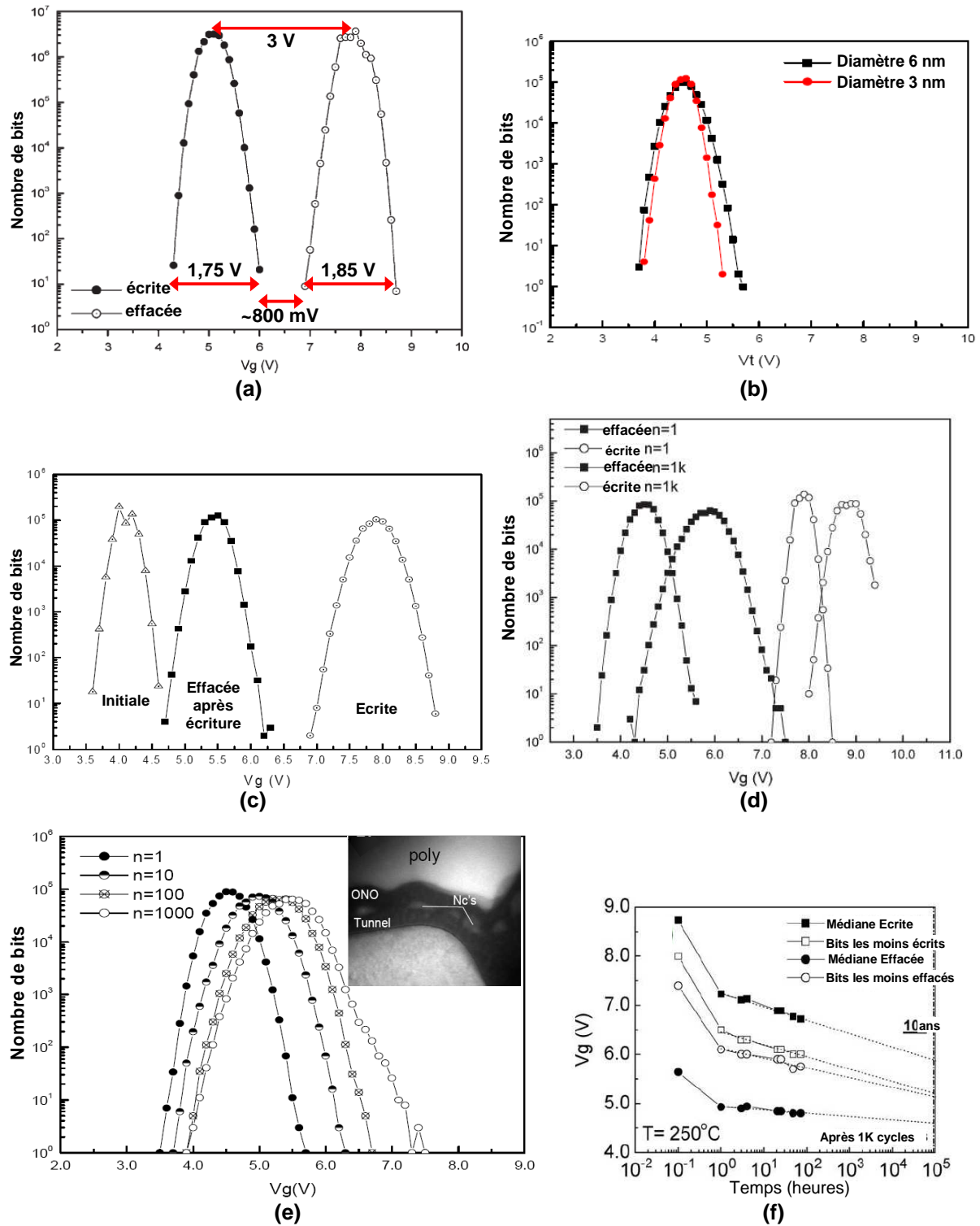


Fig. I-33 : (a) : Distributions écrites et effacées de la matrice NOR 16 Mb. Une procédure de programmation à faibles tensions (« soft programming ») a été effectuée après l'effacement afin de resserrer la partie gauche de la distribution ($V_g \leq 2,5$ V, $V_b = -1,2$ V, $t \leq 100$ μ s). (b) : Comparaison des distributions effacées d'un secteur 512 Kb avec des Si-NCs de 3 et 6 nm. (c) : Comparaison entre la distribution initiale et la distribution effacée d'un secteur 512 Kb. (d) : Evolution des distributions écrites et effacées d'un secteur 512 Kb après 1000 cycles. (e) : Elargissement de la distribution effacée d'un secteur au cours de l'endurance. Insert : Coupe TEM d'une cellule mémoire montrant la présence de Si-NCs de plus grande taille (10 nm) sur les coins de l'active. (f) : Rétention à 250°C sur un secteur 512 Kb après 1K cycles. [Gerardi'06][Gerardi'07-a][Gerardi'07-b].

I.7 Conclusion

Dans ce chapitre d'introduction, nous avons d'abord présenté les enjeux du marché des mémoires. Nous avons montré l'essor considérable que le marché des mémoires non-volatiles et en particulier celui des mémoires Flash a connu ces dernières années, tiré par les applications portables grand public. Ensuite, nous avons présenté un bref historique des mémoires non-volatiles, ainsi que le fonctionnement des mémoires Flash. Cependant, les dispositifs mémoires conventionnels à grille flottante ont des difficultés à pousser davantage la miniaturisation. Le diélectrique « interpoly » doit permettre d'obtenir un coefficient de couplage entre la grille flottante et la grille de contrôle élevé tout en maintenant une bonne rétention des données. L'épaisseur du diélectrique tunnel a atteint aujourd'hui sa taille minimale, fixée par la nécessité d'assurer 10 années de rétention malgré le phénomène du « SILC ». La diminution de l'espace entre cellules voisines entraîne une augmentation des couplages capacitifs entre grilles flottantes voisines. La réduction des dimensions conduit également à un nombre d'électrons plus faible (et donc moins fiable) pour coder l'information. Lorsque l'on arrive à la dizaine d'électrons, il apparaît des phénomènes stochastiques aléatoires, qui induisent une dispersion des caractéristiques électriques de ces dispositifs.

Afin de prolonger la viabilité de ces dispositifs, des solutions comme l'utilisation de nouveaux matériaux ou l'emploi de nouvelles architectures sont envisagées. En particulier, l'une d'entre elles est le remplacement de la grille flottante continue par des sites de piégeage discrets, tels qu'une couche de nitrure (SONOS, TANOS, NROM) ou bien des nanocristaux de silicium. En effet, ces dispositifs présentent une bonne immunité face aux défauts dans l'oxyde tunnel. Cependant, les mémoires de type SONOS montrent une saturation des caractéristiques d'effacement en fonction de la tension appliquée sur la grille. Un autre problème des mémoires avec une couche de nitrure est la rétention en température. C'est pourquoi nous nous sommes plutôt intéressés aux mémoires à nanocristaux de silicium, qui ont fait l'objet de ce manuscrit.

Les dispositifs à nanocristaux de silicium présentent en effet plusieurs avantages. Ils sont robustes aux défauts du diélectrique tunnel et présentent une bonne rétention à haute température. De plus, il est possible de stocker physiquement deux bits par cellule. Le couplage capacitif entre cellules voisines est fortement réduit. Enfin, le procédé de fabrication est totalement compatible avec le procédé CMOS.

Jusqu'à aujourd'hui, deux industriels se sont essentiellement intéressés au sujet : Freescale et STMicroelectronics. Des démonstrateurs mémoires de 4 Mb et 16 Mb, respectivement, ont été réalisés. La fenêtre de programmation obtenue est d'environ 3 V. Leurs études ont montré que l'optimisation du diélectrique tunnel et du diélectrique de contrôle, ainsi que celle des nanocristaux (taille, densité) étaient primordiales pour obtenir des mémoires avec de bonnes performances.

ANNEXE : Les solutions alternatives à la miniaturisation des mémoires Flash

A. Les mémoires FeRAM

Les FeRAM ont une structure similaire aux DRAM, c'est-à-dire que chaque cellule est composée d'un transistor MOS et d'une capacité ferroélectrique. Cette capacité est constituée d'un matériau ferroélectrique cristallin, généralement du PZT (Plomb-Zirconium-Titane). La **Fig. I-34(a)** présente la structure d'un cristal PZT. L'atome Zr/Ti au centre a deux états d'énergie stables, qui correspondent à deux polarisations de signes opposés. Ces états déterminent la position de l'atome. Lorsqu'un champ électrique est appliqué, l'atome se déplace dans la direction de celui-ci et l'atome passe d'un état stable à un autre. L'effet mémoire est obtenu grâce à ce changement de polarisation (**Fig. I-34(b)**).

La mémoire est lue en appliquant une tension positive et le transistor permet de détecter si un courant peut traverser la structure. Si la cellule était initialement à « 0 », le champ électrique appliqué au matériau ferroélectrique est identique à l'orientation de ses domaines. Le courant entre source et drain est alors faible. Au contraire, si la cellule était initialement à « 1 », le champ électrique appliqué au matériau ferroélectrique est opposé à l'orientation de ses domaines. Il y a alors réorientation de ceux-ci et apparition d'un pulse de courant entre source et drain. Il faut noter que la lecture d'un « 1 » est donc destructive, car l'information est perdue et il est nécessaire de reprogrammer la cellule.

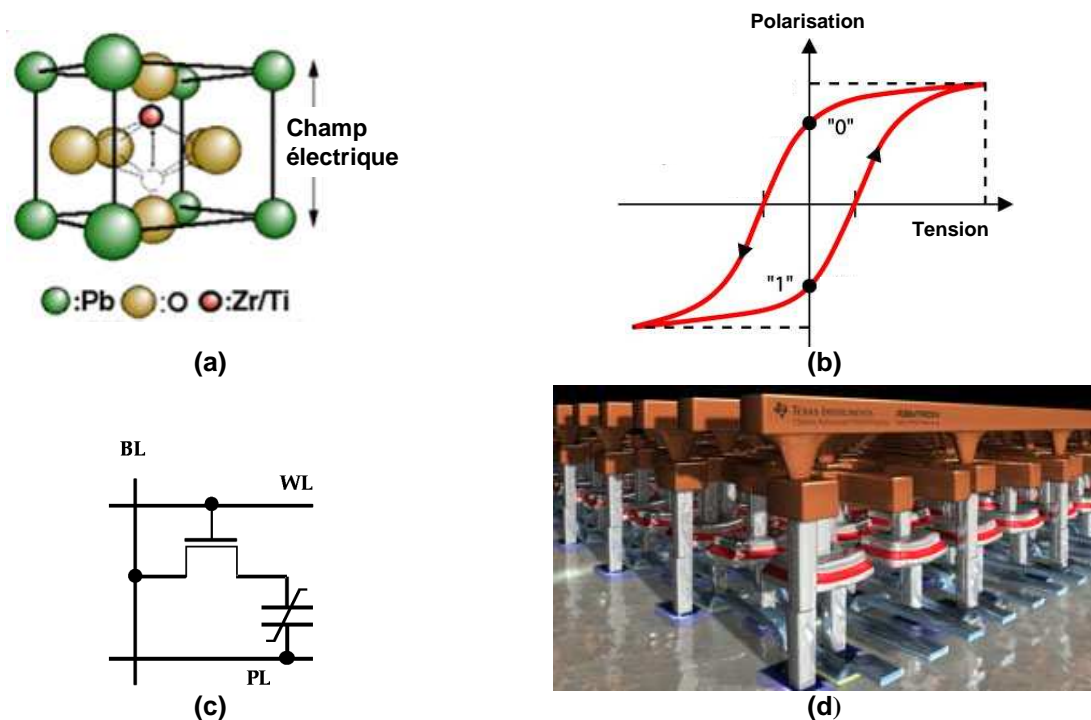


Fig. I-34 : (a) : Schéma de la structure PZT. (b) : Caractéristique typique d'une capacité ferroélectrique mettant en évidence l'hystérésis de la polarisation en fonction de la tension appliquée. (c) : Schéma électrique équivalent d'une FeRAM de type 1 transistor-1 capacité. (d) : Représentation de la FeRAM 4 Mb commercialisée par Ramtron en collaboration avec Texas Instrument [Ramtron'07].

En mars 2007, une mémoire FeRAM de 4 Mb a été mise sur le marché par Ramtron [**Ramtron'07**] avec une technologie 130 nm dessinée par Texas Instrument (**Fig. I-34(d)**).

Les principaux avantages sont les basses tensions de fonctionnement et une écriture et lecture rapides. Les inconvénients sont pour l'instant leur faible densité d'intégration par rapport à la Flash et le fait que leur lecture soit destructive.

B. Les mémoires MRAM

Les mémoires MRAM ou magnétorésistives exploitent le ferromagnétisme pour stocker l'information. Elles sont constituées d'une jonction magnétique tunnel (MTJ) intégrée avec des dispositifs CMOS. Dans la jonction MTJ, deux couches de matériau ferromagnétique sont séparées par une fine couche d'isolant formant une jonction tunnel. L'orientation magnétique de la couche du bas est fixée, tandis que celle du haut peut être réorientée à souhait, par application d'un champ magnétique. Le codage de « 0 » ou « 1 », est donné par la variation de résistivité de la jonction. Lorsque les orientations des domaines magnétiques sont identiques, la jonction magnétique est conductrice, alors que dans le cas contraire elle isole.

Développée depuis les années 90, il existe actuellement trois approches majeures de la technologie MRAM, différenciées par la façon de changer les orientations des domaines magnétiques au cours de l'écriture : changement assisté par champ magnétique, thermiquement assisté et le changement par couple de spin [**Hosomi'05**].

Une MRAM de 4 Mb a été mise en production pour la première fois par Freescale en juin 2006 [**Tehari'06**]. La méthode utilisée est le changement assisté par champ magnétique. La cellule est représentée dans la **Fig. I-35(b)**. L'opération d'écriture consiste à appliquer un courant aux deux lignes proches de la jonction MTJ. Les champs magnétiques créés par les deux courants alignent les domaines magnétiques dans la direction voulue.

La deuxième méthode, développée entre autres par Spintec, utilise en plus un échauffement par effet joule [**Prejbeanu'04**][**Nozières'07**]. Dans ce cas, pendant l'écriture, le transistor de sélection est ouvert, ce qui permet de laisser passer un courant à travers la jonction MTJ qui se comporte comme une résistance et s'échauffe (**Fig. I-35(c)**). L'avantage de cette méthode est la réduction du courant nécessaire à la programmation d'un bit.

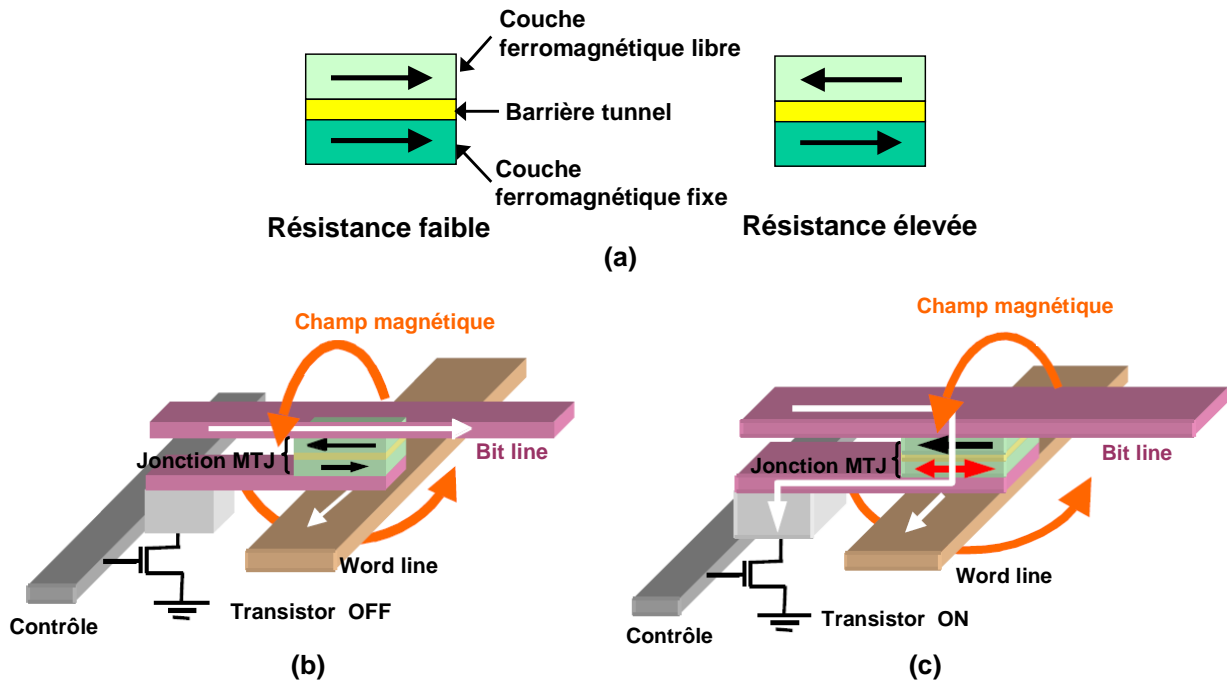


Fig. I-35 : (a) : Jonction tunnel magnétique MTJ. (b) : Cellule MRAM pendant l'écriture : le changement de l'orientation magnétique est obtenu par la création d'un champ magnétique. (c) : Cellule MRAM pendant l'écriture : le changement de l'orientation magnétique est assisté thermiquement : le passage de courant à travers la jonction permet un échauffement par effet Joule. [Nozières'07].

C. Les mémoires PCM ou à changement de phase

Les mémoires PCM (Phase Change Memory) s'appuient sur des matériaux dits chalcogénures (en général un alliage $\text{Ge}_2\text{Sb}_2\text{Te}_5$ ou GST), qui peuvent exister de manière stable dans deux états structuraux différents : un état cristallin, caractérisé par une résistance électrique basse, ou un état amorphe, caractérisé par une résistance électrique élevée. Le passage d'un état à l'autre s'effectue en chauffant localement le matériau chalcogénure. La chaleur est apportée par le passage de courant à travers une électrode résistante, ce qui induit un chauffage par effet joule (**Fig. I-36(a)**). Pour une température au-dessus de son point de fusion, le chalcogénure devient amorphe, alors que pour un chauffage juste en-dessous de ce point, il devient cristallin.

Les avantages de cette technologie sont une lecture et une écriture rapide, des faibles tensions de programmation, une endurance élevée et un faible coût. L'inconvénient majeur est le fort courant nécessaire pour mettre la cellule dans l'état « reset ». La technologie du changement de phase a été découverte dans les années 60 par Stanford Ovshinsky et est à l'origine de nombreux brevets d'invention sur le sujet. La société Ovonyx, formée en 1999, détient les licences de la technologie des mémoires à changement de phase, qu'elle délivre au fil des années à des sociétés telles que Intel, IBM, Samsung ou STMicroelectronics. Samsung a présenté en 2006 [Oh'06] une PCM de 512 Mb en technologie CMOS 90 nm et espère commercialiser des produits PCM début 2008, de même qu'Intel.

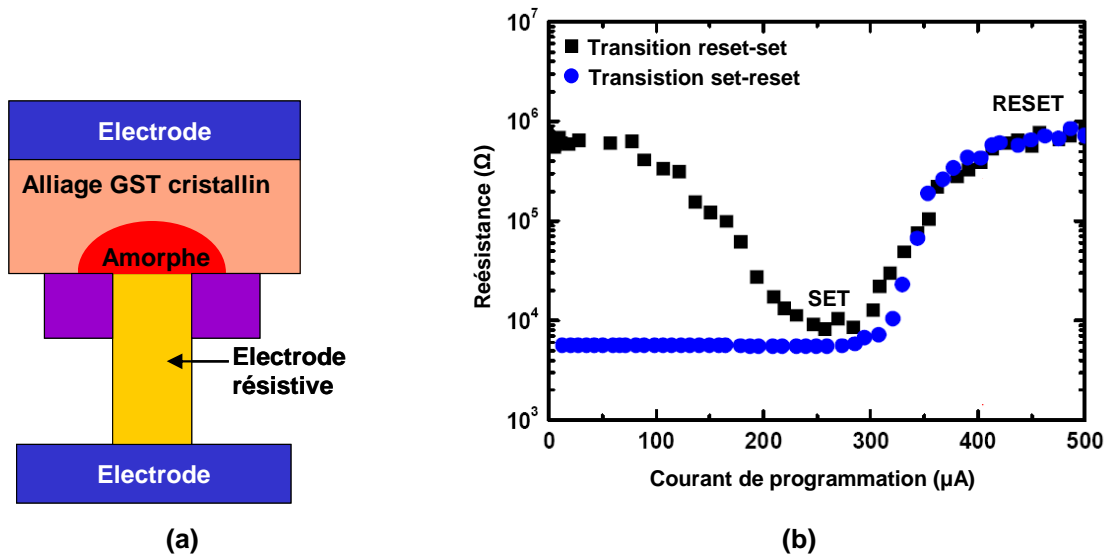


Fig. I-36 : (a) : Schéma de principe d'une cellule mémoire PCM. (b) : Caractéristiques de programmation d'une PCM : la résistance du chalcogénure est modifiée en appliquant un pulse de courant qui chauffe le matériau par effet joule [Pellizier'06].

D. La mémoire millipède

Une des limitations à la miniaturisation est le procédé de lithographie définit les dimensions. C'est pourquoi d'autres solutions sans procédé lithographique sont étudiées afin d'augmenter la densité d'intégration. L'une d'entre elles est la mémoire « millipède » développée par IBM et présentée pour la première fois en 2000 [Lutwyche'00].

Le millipède est une mémoire thermomécanique où la lecture et l'écriture peuvent avoir lieu en parallèle grâce à plusieurs milliers de pointes AFM (**Fig. I-37(a)**). Le dispositif se déplace selon X et Y et un autre dispositif (celui d'adressage) permet d'assurer le contact entre les cantilevers et le support en polymère (**Fig. I-37(b)**). L'information est codée par des trous gravés sur une couche de polymère. L'écriture, c'est-à-dire la gravure d'un trou est réalisée de la façon suivante : la pointe est chauffée à $\sim 400^\circ C$ par un pulse de courant qui traverse un matériau résistif du cantilever. La pointe est alors mise en contact avec le polymère qui sous la pression combinée à la chaleur laisse un trou qui code un « 1 » (**Fig. I-37(c)**). La lecture est réalisée en scannant la surface. Une des deux parties résistives du cantilever est chauffée à $200^\circ C$. Si la pointe rencontre un trou, la distance entre le cantilever et le polymère diminue et la chaleur du cantilever est dissipée. Ainsi, la température diminue, ce qui augmente la résistance (**Fig. I-37(d)**). Cette variation de conductance peut être facilement détectée par la mesure du courant traversant le matériau résistif. Les densités obtenues sont supérieures au Tb/in².

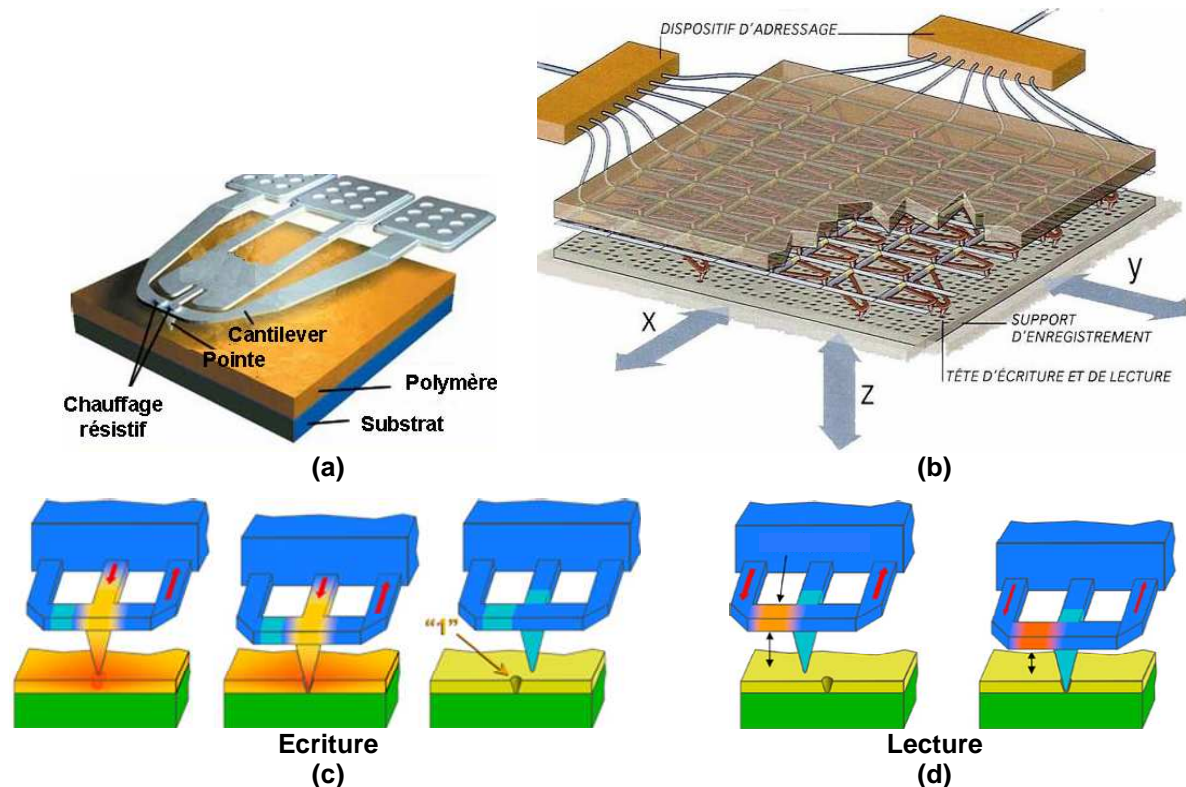


Fig. I-37 : (a) : Agrandissement sur l'une des pointes qui crée un creux dans la couche de polymère. (b) : Schéma de principe de la mémoire millipède : une matrice fixe de points permet de lire les informations stockées en détectant les creux gravés dans la couche de polymère. Le support sur lequel est posé la couche de polymère est quant à lui mobile (c) : Schéma de l'écriture de la mémoire millipède. (d) : Schéma de lecture.

E. Les mémoires moléculaires

L'électronique moléculaire est une autre voie qui permettrait de contourner les limites de la miniaturisation.

Les mémoires moléculaires utilisent le fait que des molécules électriquement actives peuvent piéger quelques charges électroniques suivant leur état d'oxydoréduction. L'avantage principal de ce type de dispositif est l'indépendance de la charge piégée par rapport à la tension qui est appliquée pour le programmer. Les principaux groupes étudiant cette voie sont la NASA et ZettaCore, en collaboration avec l'Université de Caroline du Nord. Elles ont toutes adopté l'approche « hybride » silicium/molécule : une couche de molécules est réalisée sur le substrat silicium ou SiO_2 de transistors ou capacités.

La **Fig. I-38(a)** montre le travail de la NASA. Une couche de molécules est intégrée sur un transistor à nanofil. Cette couche agit comme la couche de stockage et les états « 1 » et « 0 » sont déterminés par la résistance de la cellule qui varie selon les charges piégées par les molécules. Il est possible de faire une programmation multi-niveaux [Li'04]. L'Université de Caroline du Nord, quant à elle, a réalisé des capacités EMOS (Electrolyte-Molécule-Oxide-Silicium) (**Fig. I-38(b)**). Enfin, ZettaCore cherche plutôt à fabriquer des mémoires de type DRAM. Elle propose la ZettaRAMTM basée sur l'architecture DRAM, exceptée que la capacité classique est remplacée par une capacité moléculaire du même type que la **Fig. I-38(b)** (**Fig. I-38 (c)**).

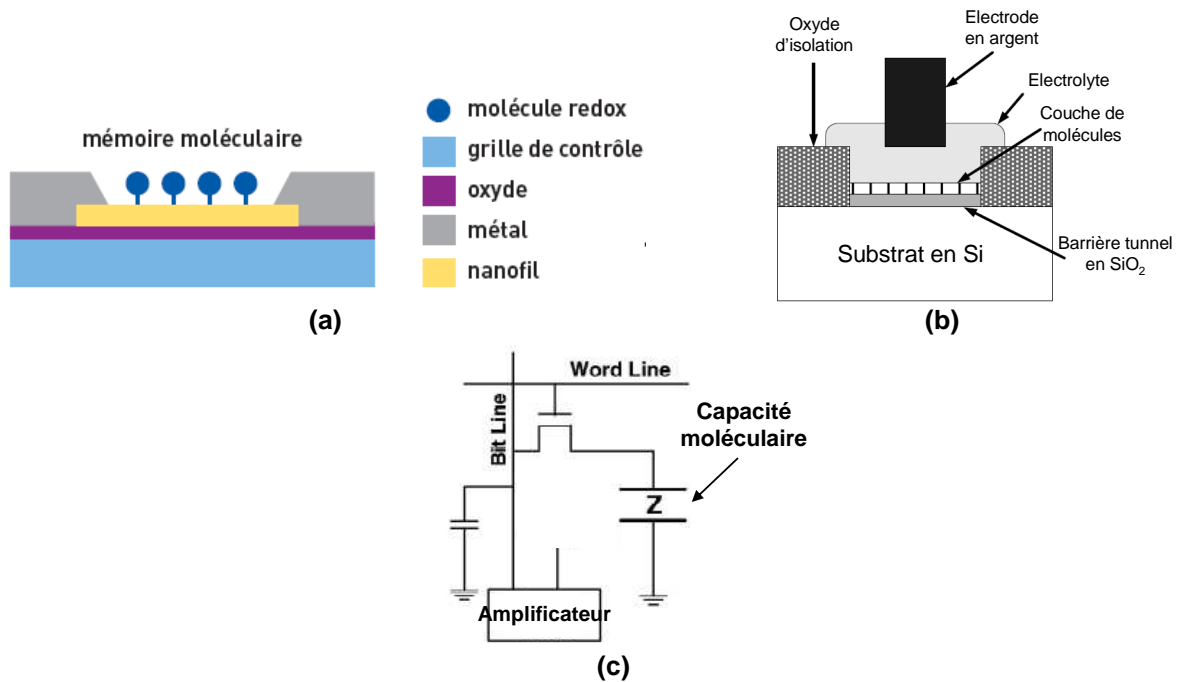


Fig. I-38 : (a) : Schéma d'une mémoire moléculaire pour le stockage multi-niveaux des données. Les différents niveaux d'un élément de mémoire moléculaire sont obtenus en modifiant le rapport de la population des molécules à l'état oxydé et réduit, en appliquant des impulsions de potentiel de grille de diverses amplitudes. (b) : Schéma d'une capacité hybride EMOS (Electrolyte-Molécule-Oxide-Silicium) avec une monocouche de molécules à la surface du SiO₂ ; l'électrolyte de grille est contactée par une électrode en Argent. [Mathur'05]. (c) : Schéma électrique du circuit ZettaRAMTM d'après [VanDuuren'06].

F. Les mémoires organiques

Les matériaux organiques peuvent être utilisés dans des applications mémoires haute densité. L'avantage de ces matériaux est la possibilité d'ajuster leurs propriétés par une modification sélective de leur structure moléculaire. D'autre part, les procédés existants pour les déposer sont de complexité réduite. Leur principale limitation est leur faible stabilité en température, les obligeant à être utilisés en fin de procédé de fabrication.

La cellule élémentaire fonctionne sur le principe de bi-stabilité électrique. Le matériau organique possède la propriété de pouvoir être mis dans un état haute ou faible impédance, suivant le sens du champ électrique qui lui est appliqué (**Fig. I-39(a)**). Plusieurs mécanismes sont évoqués pour expliquer ce phénomène. D'après [Gao'00], l'hypothèse la plus probable serait une réorientation moléculaire en fonction du champ électrique appliqué, qui induirait un désordre local dans le matériau, c'est-à-dire une amorphisation du matériau.

La mémoire hybride consiste en un empilement électrode-matériau organique-électrode de manière croisée (on parle d'architecture « cross bar ») (**Fig. I-39(b)**). La configuration d'un plan mémoire s'en trouve donc très simplifiée. Un des avantages de ce type de mémoire serait la possibilité d'empiler ces éléments afin de réaliser une construction 3D représentée sur la **Fig. I-39(c)** [Sezi'03].

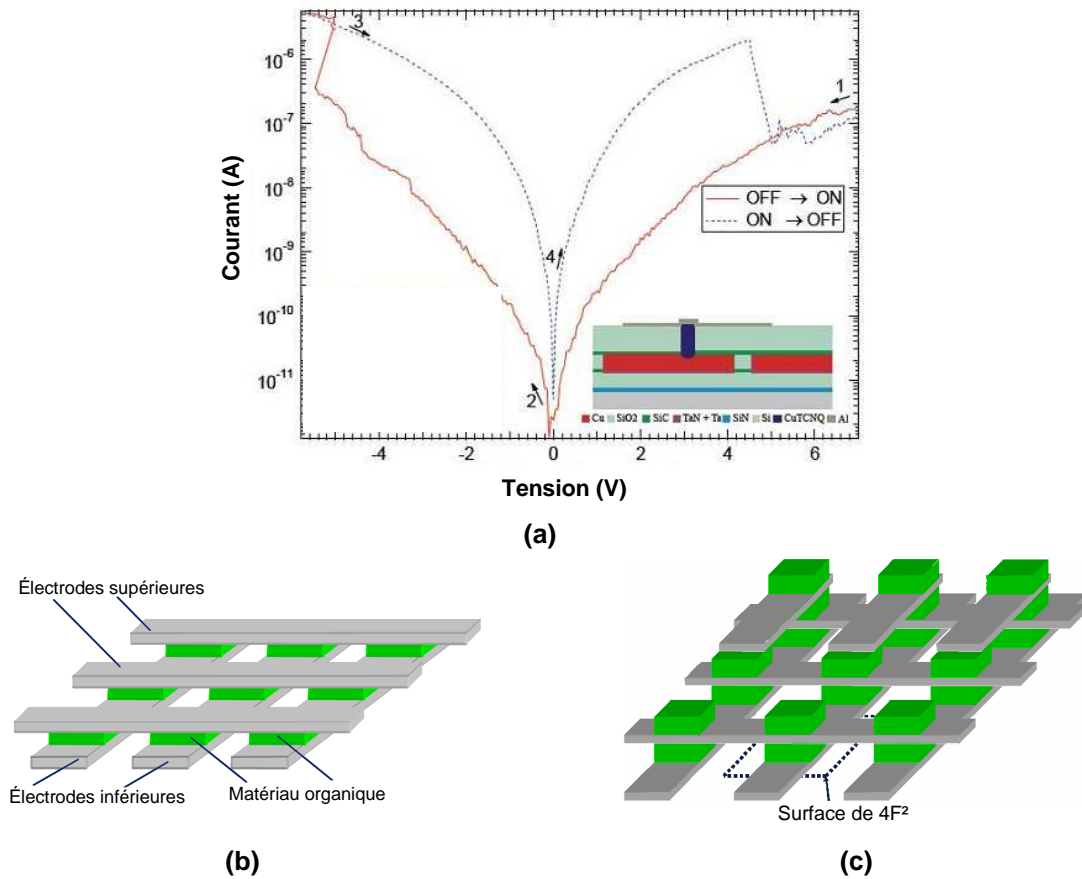


Fig. I-39 : (a) Caractéristiques I(V) typiques d'une mémoire à commutation de résistance [Muller'07]. (b) : Architecture d'une mémoire organique : des îlots de matériau organique sont adressés en mode X-Y. (c) : Du fait de sa simplicité de fabrication, ce type de mémoire pourrait être facilement empilée. [Sezi'03].

Bibliographie

- [Atwood'97] G. Atwood, A. Fazio, D. Mills, B. Reaves, "Intel StrataFlash™ Memory Technology Overview", Intel Technology Journal Q4'97.
- [Bez'07] Roberto Bez et al., "An Overview of Flash Memories", - ST Microelectronics, International Workshop on "Emerging Non Volatile Memories", Satellite Workshop of ESSDERC 2007, Munich, 14 September 2007.
- [Brown'97] W.D. Brown, J.E. Brewer (éditeurs), "Nonvolatile Semiconductor Memory Technology", IEEE PRESS, 1997.
- [Bu'02] J. Bu, M.H. White, "Retention Reliability Enhanced SONOS NVSM with Scaled Programming Voltage", Aerospace Conf. Proc., vol. 5, p.5-2383, 2002.
- [Cappelletti'99] P. Cappelletti, C. Golla, P. Olivo, E. Zanoni, "Flash Memories", Kluwer Academic Publishers, 1999.
- [DeSalvo'01] B. De Salvo, G. Ghibaud, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, B. Guillaumot "Experimental and theoretical investigation of nanocrystals and nitride-trap memory devices", IEEE Trans. on El. Dev., Vol. 48, No. 8, p. 1789, 2001.
- [DeSalvo'03] B. De Salvo, C. Gérardi, S. Lombardo, T. Baron, L. Perniola, D. Mariolle, P. Mur, A. Toffoli, M. Gély, M.N. Séméria, S. Deleonibus, G. Ammendola, V. Ancarani, M. Melanotte, R. Bes, L. Baldi, D. Corso, I. Crupi, R.A. Puglisi, G. Nicotra, E. Rimini, F. Mazon, G. Ghibaud, G. Pananakakis, C.M. Compagnoni, D. Ielmini, A. Lacaita, A. Spinelli, Y.M. Wan, K. Van der Jeugd, "How far will Silicon nanocrystals push the scaling limits NVMs technologies ?", IEEE IEDM Tech. Dig., pp. 26.1.1-26.1.4, 2003.
- [DeSalvo'04] B. De Salvo et al., "Performance and Reliability Features of Advanced Nonvolatile Memories Based on Discrete Traps (Silicon Nanocrystals, SONOS)", IEEE Trans. on Dev. and Mat. Reliab., vol. 4, no. 3, p. 377, 2004.
- [DiMaria'95] D.J. DiMaria, "Stress induced leakage currents in thin oxides", Microelectron. Eng., vol. 28; p. 63, 1995.
- [Eitan'05] B. Eitan, G. Cohen, A. Shappir, E. Lusky, A. Givant, M. Janai, I. Bloom, Y. Polansky, O. Dadashev, A. Lavan, R. Sahar, E. Maayan, "4-bit per Cell NROM Reliability", IEEE IEDM Tech. Dig., 2005.
- [Eitan'99] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer D. Finzi, "Can NROM, a 2-bit, trapping storage NVM cell, give a real challenge to floating gate cells ?", présenté à SSDM international conference, 1999.
- [Engel'05] B. N. Engel, J. Åkerman, B. Butcher, R. W. Dave, M. DeHerrera, M. Durlam, G. Grynkewich, J. Janesky, S. V. Pietambaram, N. D. Rizzo, J. M. Slaughter, K. Smith, J. J. Sun, and S. Tehrani, "A 4-Mbit Toggle MRAM Based on a Novel Bit and Switching Method ," IEEE Trans. on Magnetics, vol. 41, pp. 132-136, 2005.

- [Frohman'71]** D. Frohman-Bentchkowsky, "A fully decoded 2048-bit electrically programmable MOS-ROM", IEEE ISSCC Tech. Dig., p.80, 1971.
- [Gao'00]** H.J. Gao, K. Sohlberb, Z.Q. Xue, H.Y. Chen, S.M. Hou, L.P. Ma, X.W. Fang, S.J. Pang, S.J. Pennycook, "Reversible, Nanometer-Scale Conductance transitions in an Organic Complex", *Phy. Review Lett.*, vol. 84, no. 8, pp. 1780-1783, 2000.
- [Gasquet'06]** H.P. Gasquet, R.F. Steimle, R.A. Rao, R. Muralidhar, "Impact of Nanocrystal Size on NVM Array Reliability Performance", NVSMW Proc., 2006
- [Gerardi'06]** C. Gerardi, S.Lombardo, G.Costa, G.Ammendola, V. Ancarani, O. Brafa, D. Mello, "Study of Nanocrystal Memory Integration in a 16 Mb Flash-like NOR Array", WoDiM Proc., 2006.
- [Gerardi'07-a]** C. Gerardi, V. Ancarani, R. Portoghese, S. Giuffrida, M. Bileci, G. Bimbo, O. Brafa, D. Mello, G. Ammendola, E. Tripiciano, R. Puglisi, S. A. Lombardo, "Nanocrystal Memory Cell Integration in a Stand-Alone 16-Mb NOR Flash Device", *IEEE Trans. On Elec. Dev.*, Vol. 54, No. 6, p. 1376, 2007.
- [Gerardi'07-b]** C. Gerardi, S. Lombardo, G. Ammendola, G. Costa, V. Ancarani, D. Mello, S. Giuffrida, M.C. Plantamura, "Study of nanocrystal memory integration in a Flash-like NOR device", *Microelectronics Reliability*, vol. 47, pp. 593–597, 2007.
- [Guarani'03]** K. W. Guarini et al., "Low voltage, scalable nanocrystal FLASH memory fabricated by templated self assembly", *IEEE IEDM Tech. Dig.*, p.541, 2003.
- [Hosomi'05]** M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K., Yamane, H. Yamada, M. Shoji, H. Hachino, C. Fukumoto, H. Nagao, H. Kano, "A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram", *IEEE IEDM Tech. Dig.*, pp. 459- 462, 2005.
- [ITRS'06]** International Roadmap for Semiconductors,
<http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm>
- [Johnson'80]** W.S. Johnson, G. Perlegos, A. Renninger, G. Kuhn, T.R. Ranganath, "A 16KB Electrically Erasable Nonvolatile Memory", *ISSCC Tech. Dig.*, p.152, 1980.
- [Kahng'67]** D. Kahng and S.M. Sze, "A floating gate and its application to memory devices", *Bell Systems Technology Journal*, vol. 46, p. 1288, 1967.
- [Kim'07]** S. Kim, W. Cho, J. Kim, B. Lee, S. Park, "Air-Gap Application and Simulation Results for Low Capacitance in 60nm NAND Flash Memory", NVSMW Proc., p. 54, 2007.
- [Kynett'88]** V.N. Kynett, A. Baker, M. Fandrich, G. Hoekstra, O. Jungroth, J. Kreifels, S. Wells, "An in-system reprogrammable 256 KCMOS Flash memory", *ISSCC Tech. Dig.*, p.132.
- [Lai'06]** C. H. Lai et al., "Very Low Voltage SiO₂/HfON/HfAlO/TaN Memory with Fast Speed and Good Retention", *Tech. Dig of VLSI*, 2006.
- [Lai'98]** S. Lai, "Tunnel oxide and ETOXtm Flash scaling limitation", *Int. Non Volatile Memory Conf.*, pp. 6-7, 1998.

- [Larcher'04]** L. Larcher, P. Pavan et B. Eitan, "On the Physical Mechanism of the NROM Memory Erase", IEEE Transactions on Electron Devices, Vol. 51, N°10, pp. 1593-1599, 2004.
- [Lee'06]** C.H. Lee et al., "Charge Trapping Memory Cell of TANOS (Si-Oxide-SiN-Al₂O₃-TaN) Structure Compatible to Conventional NAND Flash Memory", Proc. NVSMW 2006.
- [Li'04]** C. Li, W. Fan, B. Lei, D. Zhang, S. Han, T. Tang, X. Liu, Z. Liu, S. Asano, M. Meyyappan, J. Han and C. Zhou, "Multilevel memory based on molecular devices", Appl. Phys. Lett., 84, pp. 1949-1951, 2004.
- [Libsch'87]** F. Libsch, A. Roy, M.H. White, "A true 5V EEPROM cell for high-density NVSM", IEEE Trans. on Electron Dev., vol. 34, p. 2371, 1987.
- [Liu'02-a]** Z. Liu et al., "Metal Nanocrystal Memories—Part I: Device Design and Fabrication", IEEE Trans. on El. Dev, vol. 49, no. 9, p. 1606, 2002.
- [Liu'02-b]** Z. Liu et al., "Metal Nanocrystal Memories—Part II: Electrical Characteristics", IEEE Trans. on El. Dev., vol. 49, no. 9, p. 1614, 2002.
- [Lutwyche'00]** M.I. Lutwyche, G. Cross, M. Despont, U. Drechsler, U. Durig, W. Haberle, H. Rothuizen, R. Stutz, R. Widmer, G.K. Binnig, P. Vettiger, "Millipede—a highly-parallel dense scanning-probe-based data-storage System", IEEE International Solid-State Circuits Conference Tech. Dig., pp. 126-7, 450, 2000.
- [Maserjan'82]** J. Maserjan, N. Zamani, "Observation of positively charged state generation near the Si/SiO₂ interface during Fowler-Nordheim tunnelling", J. Vac. Sci. Technol., vol. 20, no. 3, p. 743, 1982.
- [Masuoka'84]** F. Masuoka, M. Asano, H. Iwashi, T. Komuro, « A new Flash EEPROM cell using triple polysilicon technology », IEEE IEDM Tech. Dig., p.464, 1984.
- [Mathur'05]** G. Mathur, S. Gowda, V. Misra, "Threshold Voltage-Assisted Reduction of Molecules in Hybrid Silicon/Molecular Memory Devices" IEEE Conference on Nanotechnology Proc., pp. 442-445, 2005.
- [Minami'91]** S. Minami, Y. Kamigaki, "New scaling guidelines for MNOS non volatile memory devices", IEEE Trans. Elect. Dev., vol. 38, no. 11, p. 2519, 1991.
- [Molas'04]** G. Molas, D. Deleruyelle, B. De Salvo, G. Ghibaud, M. Gely, S. Jacob, D. Lafond, S. Deleonibus, "Impact of few electron phenomena on floating-gate memory reliability", IEEE IEDM Tech. Dig., pp. 877-880, 2004.
- [Molas'06]** G. Molas, D. Deleruyelle, B. De Salvo, G. Ghibaud, M. Gély, L. Perniola, D. Lafond, S. Deleonibus, "Degradation of Floating-Gate Memory Reliability by Few Electron Phenomena", IEEE Trans. on Electron Dev., vol. 53, no. 10, 2006.
- [Molas'07]** G.Molas et al., "Thorough investigation of Si-nanocrystal memories with high-k interpoly dielectrics for sub-45nm node Flash NAND applications", IEEE IEDM Tech. Dig., 2007.

- [Muller'07]** R. Müller, R. Naulaerts, J. Billen, J. Genoe, P. Heremansb, "CuTCNQ resistive nonvolatile memories with a noble metal bottom electrode", *App. Phys. Lett.* 90, p. 063503, 2007.
- [Muralidhar'03]** R. Muralidhar, R.F. Steimle, M. Sadd, R. Rao, C.T. Swift, E.J. Prinz, J. Yater, L. Grieve, K. Harber, B. Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S.G.H. Anderson, M. Rossow, T. Merchant, M. Paransky, T. Huynh, D. Hadad, K.M. Chang, B.E. White, "A 6V Embedded 90nm Silicon Nanocrystal Nonvolatile Memory", *IEEE IEDM Tech. Dig.*, pp. 26.2.1-26.2.4, 2003.
- [Muralidhar'03]** R. Muralidhar, R.F. Steimle, M. Sadd, R. Rao, C.T. Swift, E.J. Prinz, J. Yater, L. Grieve, K. Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S.G.H. Anderson, M. Rossow, M. Paransky, T. Huynh, D. Hadad, KO-Min Chang, B.E. White Jr., "A 6V Embedded 90nm Silicon Nanocrystal Nonvolatile Memory", *IEEE IEDM Tech. Dig.*, p. 26.2.1, 2003.
- [Naruke'88]** K. Naruke, S. Taguchi et M. Wada, "Stress Induced Leakage Current Limiting to Scale Down EEPROM Tunnel Oxide Thickness", *IEEE IEDM Tech. Dig.*, pp. 424-427, 1988.
- [Nozières'07]** J. P. Nozières, "Magnetic Random Access Memories (M-RAM) : A truly universal memory?", <http://www.spintec.fr/Magnetic-Random-Access-Memories-M.html>
- [Oh'06]** J.H. Oh, J.H. Park, Y.S. Lim†, H.S. Lim†, Y.T. Oh, J.S. Kim, J.M. Shin, J.H. Park, Y.J. Song, K.C. Ryoo, D.W. Lim, S.S. Park, J.I. Kim, J.H. Kim, J. Yu, F. Yeung, C.W. Jeong, J.H. Kong, D.H. Kang, G.H. Koh, G.T. Jeong, H.S. Jeong, Kinam Kim, "Full Integration of Highly Manufacturable 512Mb PRAM based on 90nm Technology", *IEEE IEDM Tech. Dig.*, 2006.
- [Olivo'88]** P. Olivo, T.N. Nguyen, B. Ricco, "High-field-induced degradation in ultra-thin SiO₂ films", *IEEE Trans. Electron Dev.*, vol. 35, p. 2259, 1988.
- [Park'04]** C. Park et al., "A 70nm NOR Flash technology with 0.049 μm² cell size », *VLSI proceedings*, pp.238-239, 2004.
- [Park'06]** Y. Park, J. Choi, C. Kang, C. Lee, Y. Shin, B. Choi, J. Kim, S. Jeon, J. Sel, J. Park, K. Choi, T. Yoo, J. Sim, K. Kim, "Highly manufacturable 32Gb Multi-level NAND Flash memory with 0.0098 μm² cell size using TANOS cell technology", *IEEE IEDM Tech. Dig.*, pp. 29-32, 2006.
- [Pellizier'06]** F. Pellizzer, A. Benvenuti, B. Gleixner, Y. Kim, B. Johnson, M. Magistretti, T. Marangon, A. Pirovano, R. Bez, G. Atwood, "A 90nm Phase Change Memory Technology for Stand-Alone Non-Volatile Memory Applications", *Symposium on VLSI Tech. Dig.*, 2006.
- [Prejbeanu'04]** I.L. Prejbeanu, W. Kula, K. Ounadjela, R.C. Sousa, O. Redon, B. Dieny, J.-P. Nozières, "Thermally Assisted Switching in Exchange-Biased Storage Layer Magnetic Tunnel Junctions", *IEEE Trans. on Magnetics*, vol. 40, no.4, 2004.
- [Pro'07]** T. Pro et al., "Electrical Investigation of Hybrid Molecular / Silicon Memories with Redox-Active Ferrocene Molecules acting as Storage Media", *Proc. of Silicon Nanoelectronics Workshop 2007*.

- [Ramtron'07]** <http://www.ramtron.com>
- [Rao'05]** R.A. Rao , H.P. Gasquet, R.F. Steimle, G. Rinkenberger, S. Straub, R. Muralidhar, S.G.H. Anderson, J.A. Yater, J.C. Ledezma, J. Hamilton, B. Acred, C.T. Swift, B. Hradsky, J. Peschke, M. Sadd, E.J. Prinz, K.M. Chang, B.E. White Jr., "Influence of silicon nanocrystal size and density on the performance of non-volatile memory arrays", *Solid-State Electronics*, vol. 49, pp. 1722–1727, 2005.
- [Saad'03]** M. Saad et al., *IEEE NVSM Workshop*, p.71, 2003.
- [San'92]** K.T. San et al., "A new technique for determining the capacitive coupling coefficients in flash EEPROM's", *IEEE Electron Dev. Lett.*, vol.13, p.328, 1992.
- [Sezi'03]** R. Sezi, A. Walter, R. Engl, A. Maltenberger, J. Schumann, M. Kund et C. Dehm, "Organic Materials for High-Density Non-Volatile Memory Applications", *IEEE IEDM Tech. Dig.*, pp. 10.2.1-10.2.4, 2003.
- [Steimle'04]** R.F. Steimle, R.Rao, C.T. Swift, K. Harber, S. Straub, R. Muralidhar, B. Hradsky, J.A. Yater, E.J. Prinz, W. Paulson, M.Sadd, C. Parikh, S.G.H. Anderson, T. Huynh, B. Acred, L. Grieve, M. Rossow, R. Mora, B. Darlington, Ko-Min Chang, B.E.White Jr., "Integration of Silicon Nanocrystals into a 6V 4Mb Nonvolatile Memory Array", *NVSMW Proc.*, pp. 73-74, 2004.
- [Steimle'07]** R.F. Steimle, R. Muralidhar, R. Rao, M. Sadd, C.T. Swift, J. Yater, B. Hradsky, S. Straub, H. Gasquet, L. Vishnubhotla, E.J. Prinz, T. Merchant, B. Acred, K. Chang, B.E. White Jr., "Silicon nanocrystal non-volatile memory for embedded memory scaling", *Microelectronics Reliability*, vol. 47 pp. 585-592, 2007.
- [Steimle'07]** R.F. Steimle, R. Muralidhar, R. Rao, M. Sadd, C.T. Swift, J. Yater, B. Hradsky, S. Straub, H. Gasquet, L. Vishnubhotla, E.J. Prinz, T. Merchant, B. Acred, K. Chang, B.E. White Jr., "Silicon nanocrystal non-volatile memory for embedded memory scaling", *Microelectronics Reliability*, vol. 47, pp. 585–592, 2007.
- [Swift'02]** C.T. Swift et al., "An embedded 90nm SONOS non-volatile memory utilizing hot electron programming and uniform tunnel erase", *IEEE IEDM Tech. Dig.*, pp. 927-930, 2002.
- [Tehari'06]** S. Tehari, "Status and Outlook of MRAM Memory Technology", (invité) *IEEE IEDM*, 2006.
- [Tiwari'95]** S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, D. Buchanan, "Volatile and nonvolatile memories in silicon with nano-crystal storage", *IEEE IEDM Tech. Dig.*, pp. 521-524, 1995.
- [VanDuuren'06]** M. Van Duuren et al "Pushing the scaling limits of embedded non-volatile memories with high-k materials", *Proc. of ICICDT*, pp.36-39, 2006.

- [Venkatesan'07]** R.K. Venkatesan, A.S. Al-Zawawi, K. Sivasubramanian, E. Rotenberg, "ZettaRAM: A Power-Scalable DRAM Alternative through Charge-Voltage Decoupling", IEEE Trans. on Computers, vol. 56, no. 2, p. 147, 2007.
- [Wann'95]** H.C. Wann, C. Hu, "High-endurance Ultra-Thin tunnel Oxide in MONOS Device Structure for Dynamic Memory Application", IEEE Elec. Dev. Lett., vol. 16, no. 11, pp. 491-493, 1995.
- [Wegener'67]** H.A.R. Wegener, A.J. Lincoln, H.C. Pao, M.R. O'Connell, and R.E. Oleksiak, "The variable threshold transistor, a new electrically alterable, non-destructive read-only storage device", IEEE IEDM Tech. Dig., 1967.
- [White'97]** M. H. White, Y. Yang, A. Purwar, M. L. French, "A Low Voltage SONOS Nonvolatile Semiconductor Memory Technology", IEEE Trans. Comp., pack. and manufact technology, vol. 20, no. 2, 1997.
- [Yim'03]** Y.S. Yim et al., "70nm NAND Flash technology with 0.025 μm^2 cell size for 4Gb Flash memory", IEEE IEDM Tech. Dig., pp.819-823, 2003.

Chapitre II : Fabrication des mémoires Flash à nanocristaux de silicium

Chapitre II : Fabrication des mémoires Flash à nanocristaux de silicium	61
II.1 Introduction.....	65
II.2 Méthodes de fabrication des nanocristaux de silicium	65
<i>II.2.1 Précipitation de silicium en excès</i>	<i>65</i>
<i>II.2.2 Synthèse sous forme d'aérosol</i>	<i>65</i>
<i>II.2.3 Technique de croissance par CVD.....</i>	<i>66</i>
II.2.3.1 Procédé à une étape	66
II.2.3.2 Procédé à deux étapes.....	67
II.2.3.3 Influence de la préparation de surface avant dépôt des nanocristaux de silicium.....	68
II.2.3.4 Nitruration des nanocristaux de silicium.....	69
II.3 Intégration des nanocristaux de silicium dans un produit Flash NOR 130 nm	70
<i>II.3.1 Organisation d'une mémoire Flash.....</i>	<i>70</i>
<i>II.3.2 Fabrication d'une mémoire Flash standard.....</i>	<i>71</i>
<i>II.3.3 Les différentes possibilités d'intégration des nanocristaux de silicium.....</i>	<i>80</i>
<i>II.3.4 Niveaux de masques.....</i>	<i>93</i>
II.4 Conclusion.....	95
ANNEXE : Autres procédés de fabrication possibles intégrant les nanocristaux de silicium.....	96
Bibliographie.....	106

II.1 Introduction

Dans ce deuxième chapitre, nous décrirons la fabrication des mémoires Flash à nanocristaux de silicium. La première partie sera consacrée à la description des méthodes de fabrication des Si-NCs. Dans la deuxième partie, nous verrons comment les Si-NCs peuvent être intégrés dans un produit Flash NOR ATMEL de technologie 130 nm. Après la description de l'organisation générale d'une mémoire Flash, nous présenterons le procédé de fabrication de la Flash NOR standard et enfin, nous verrons quelles sont les différentes possibilités d'intégrer les Si-NCs dans ce procédé.

II.2 Méthodes de fabrication des nanocristaux de silicium

Pour des applications industrielles les Si-NCs doivent avoir des caractéristiques morphologiques maîtrisées. Nous allons ici décrire les principales méthodes reportées dans la littérature pour l'élaboration des Si-NCs.

II.2.1 Précipitation de silicium en excès

Cette méthode utilise un oxyde de silicium enrichi en silicium (SiO_x avec $x < 2$). Le SiO_x peut être obtenu par deux techniques :

- Implantation basse énergie de silicium dans du SiO_2 [Lopez'02].
- Dépôt direct par CVD [Yun'00][Buffet'02].

Ensuite, une phase de recuit de la couche de SiO_x provoque la précipitation du silicium en excès pour former les Si-NCs. La taille et la densité des Si-NCs sont contrôlées par l'enrichissement de la couche en silicium, la température et la durée des recuits.

Cette technique est pénalisée par le budget thermique qu'elle requiert : les recuits de la couche SiO_x sont généralement effectués à haute température (1000°C) et peuvent être longs (typiquement : 30 minutes). D'autre part, il est délicat de contrôler la morphologie des Si-NCs, ce qui induit une dispersion en taille de ceux-ci. De plus, la nucléation des Si-NCs ayant lieu dans le volume de la couche déposée, la distance entre les Si-NCs et le canal est difficile à maîtriser et peu homogène.

II.2.2 Synthèse sous forme d'aérosol

La synthèse sous forme d'aérosol est un procédé en trois étapes :

Tout d'abord, un aérosol de nanocristaux de silicium est élaboré par pyrolyse de silane à haute température [Ostraat'01] [Boer'01]. La taille des Si-NCs est contrôlée par la température d'élaboration, la concentration de silane dans la phase gazeuse et le temps de résidence dans le réacteur.

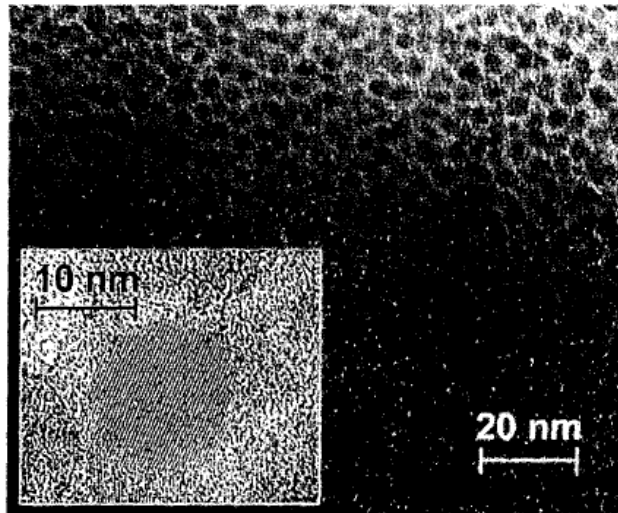


Fig. II-1 : Vue planaire d'une mono-couche de nanocristaux fabriqués avec la technique de dépôt en aérosol. La densité obtenue est de $6.10^{12}/\text{cm}^2$ et le diamètre moyen est de 4-5 nm. Insert : Observation de la cristallinité d'un filot de silicium. [Ostraat'01].

Dans la deuxième étape, les nanocristaux sont oxydés thermiquement sur une épaisseur variable, entre 1,5 et 2 nm suivant les conditions. Cette étape permet de bien isoler les Si-NCs entre eux. Enfin, les particules de silicium en suspension sont collectées sur le substrat par différents moyens (thermophorèse en appliquant un gradient de température [DeBlauwe'00] ou chargement des Si-NCs grâce à une source radioactive et application d'un champ électrique [Boer'01]).

Les auteurs [DeBlauwe'00][Ostraat'01] montrent que l'on peut former une monocouche dense de Si-NCs, sans former d'agrégats (**Fig. II-1**), avec des densités très élevées (environ $10^{13}/\text{cm}^2$).

Néanmoins, cette méthode paraît peu compatible avec les procédés technologiques utilisés aujourd'hui dans l'industrie de la microélectronique.

II.2.3 Technique de croissance par CVD

Dans cette technique, on dépose des îlots de silicium polycristallin sur l'oxyde tunnel par LPCVD (Low Pressure Chemical Vapor Deposition). Ce procédé peut être réalisé avec deux méthodes différentes décrites ci-dessous.

II.2.3.1 Procédé à une étape

Le gaz précurseur utilisé est le silane (SiH_4). La réaction conduisant au dépôt de silicium est la réaction de dissociation du silane dont l'équation bilan est la suivante :



Des îlots de silicium se forment (partant des sites de nucléation) et croissent en taille tout au long du temps de déposition. La température varie entre 500 et 600°C et la pression partielle de silane varie entre 35 et 200 mTorr.

La densité et la taille des Si-NCs peuvent être modifiées sur un large spectre par les conditions de dépôt (pression, température et temps, **Fig. II-2**), mais aussi par des traitements de surface de l'oxyde tunnel [**Mazen'03a**].

L'inconvénient majeur du procédé à une étape est la dispersion en taille des Si-NCs (**Fig. II-3**). En effet, pendant toute la durée du dépôt, certains Si-NCs croissent, tandis que d'autres nucléent [**Nicotra'03**].

II.2.3.2 Procédé à deux étapes

C'est ce procédé qui a été utilisé dans la fabrication des dispositifs réalisés dans le cadre de cette thèse. Ce procédé permet de dissocier l'étape de nucléation de celle de croissance (**Fig. II-4**) [**Mazen'03these**]. Dans la première étape (très brève), des nuclei de silicium se forment à la surface du substrat par CVD en chimie silane. Dans la deuxième étape on utilise du dichlorosilane (SiH_2Cl_2), ce qui permet une croissance sélective du silicium sur les nuclei (formés durant l'étape de nucléation) sans formation de nouveaux Si-NCs. Comme le montre la **Fig. II-5**, la densité de Si-NCs reste pratiquement constante au cours de la deuxième étape. Le fait de séparer la nucléation et la croissance des Si-NCs permet ainsi de mieux maîtriser la taille, la densité et l'homogénéité des Si-NCs.

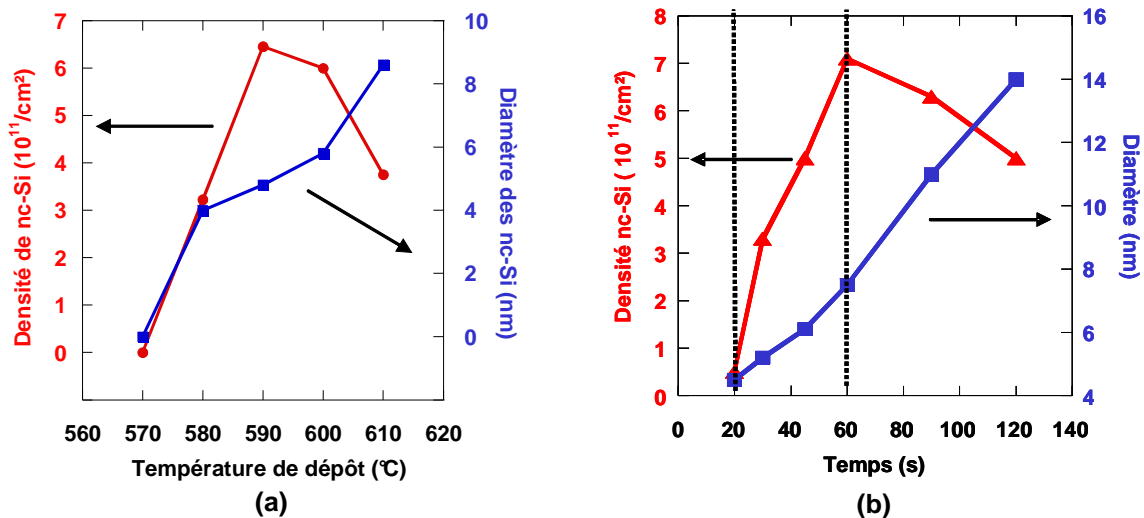


Fig. II-2 : Evolution de la densité et du diamètre moyen des Si-NCs en fonction de la température (a) et du temps de dépôt (b) d'après [**Mazen'03these**].

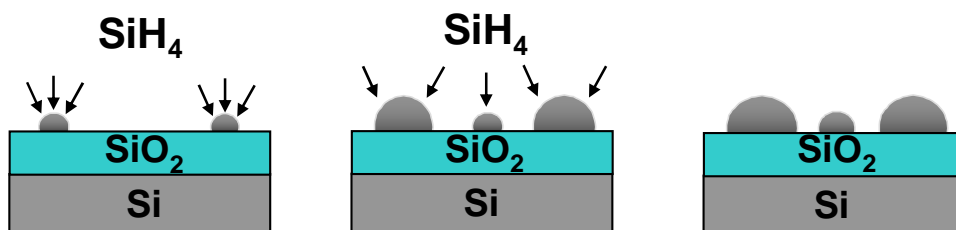


Fig. II-3 : Représentation schématique de l'origine de la dispersion en taille avec le procédé à une étape.

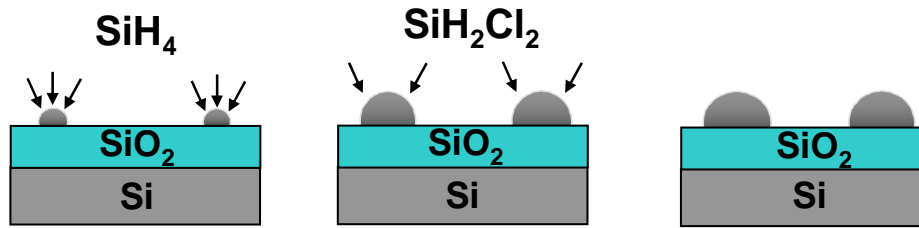


Fig. II-4 : Représentation schématique du procédé à deux étapes.

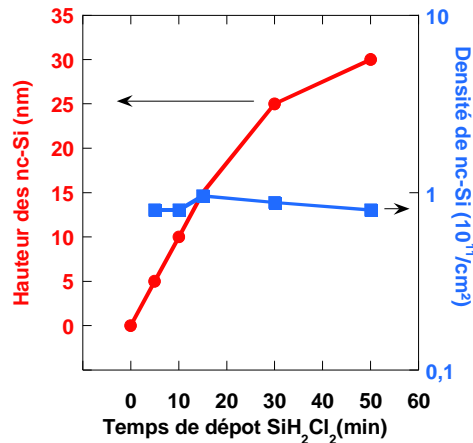


Fig. II-5 : Evolution de la densité et de la taille des Si-NCs en fonction du temps de dépôt au cours de la deuxième étape d'après [Mazen'03these].

II.2.3.3 Influence de la préparation de surface avant dépôt des nanocristaux de silicium

Afin d'augmenter la densité des Si-NCs, il est possible de faire un traitement de surface de l'oxyde tunnel donnant lieu à une forte densité de groupements silanols (Si-OH). En effet, il a été démontré que les groupements silanols agissent comme sites de nucléation pour les Si-NCs (Fig. II-6).

On utilise pour cela une solution diluée d'acide fluorhydrique (HF) qui permet d'obtenir une surface d'oxyde fortement hydroxylée. Le principe est décrit sur la Fig. II-7.

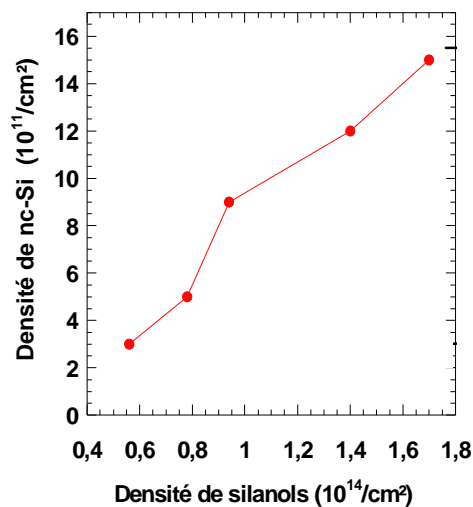


Fig. II-6 : Augmentation de la densité de Si-NCs avec la densité de groupements silanols d'après [Mazen'03these]. Le dépôt de Si-NCs est réalisé à 600°C.

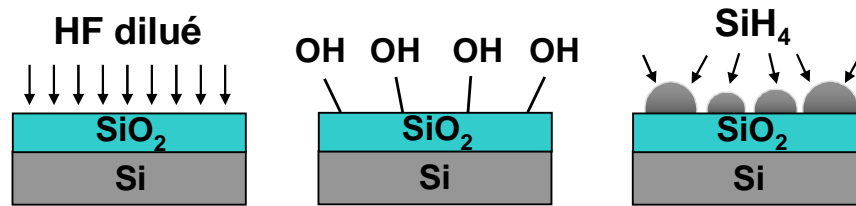


Fig. II-7 : Représentation schématique de l'hydroxylation de la surface de l'oxyde tunnel avant le dépôt de Si-NCs.

II.2.3.4 Nitruration des nanocristaux de silicium

Un des problèmes lorsque l'on veut intégrer des Si-NCs dans un procédé de fabrication complet d'une mémoire non-volatile est l'oxydation que peuvent subir les Si-NCs au cours de la fabrication (recuits, oxydes thermiques des transistors de la périphérie...).

Une solution consiste à protéger les nanocristaux par une fine couche d'oxynitrure qui constitue une barrière à l'oxydation. On doit pour cela, incorporer de l'azote par un recuit sous espèce nitrurante. Dans notre cas, deux espèces ont été utilisées : NH₃ et NO. En effet, plusieurs papiers [Dufourcq'07][Kamath'97][Scheer'03] ont démontré que ces espèces permettent de retarder l'oxydation des Si-NCs (**Fig. II-8**).

Le recuit sous NH₃ se fait à une température comprise entre 650°C et 800°C. Le recuit sous NO s'effectue à des températures plus élevées (autour de 1000°C). Néanmoins, l'étape de nitruration consomme aussi du silicium ; il faut donc incorporer assez d'azote pour avoir une barrière suffisante à l'oxydation mais en évitant de consommer complètement les Si-NCs.

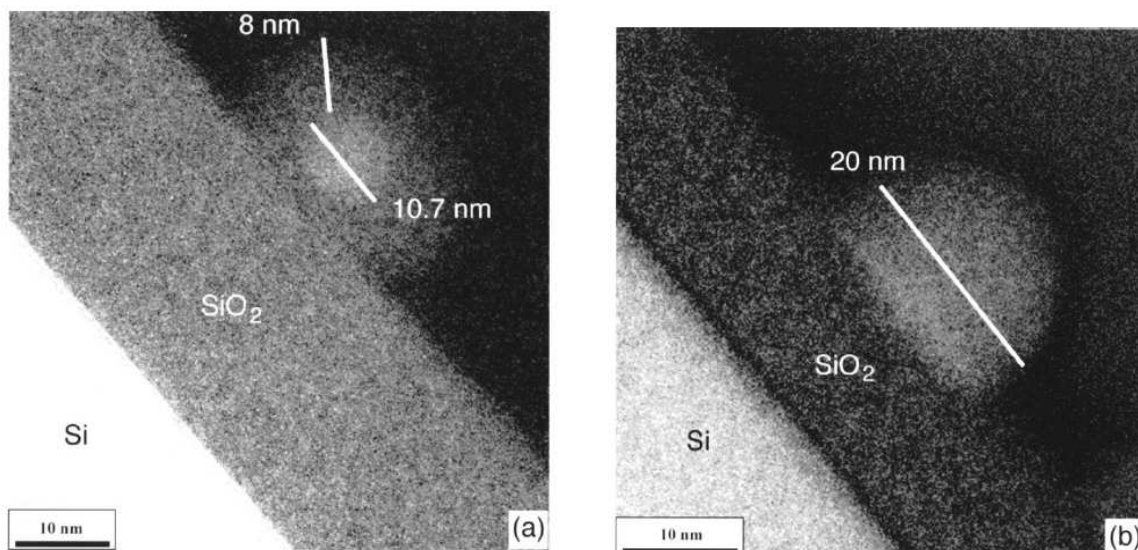


Fig. II-8 : Images TEM après une oxydation sous O₂ à 1 atm à 1050°C pendant 5 minutes d'un Si-NC non nitruré (a) et d'un Si-NC nitruré sous NO (b) (1 atm, 1050°C, 30 min). L'échantillon (a) présente une couche de 8 nm d'oxyde autour du cœur de silicium contrairement à l'échantillon (b) qui n'a pas été oxydé [Scheer'03].

II.3 Intégration des nanocristaux de silicium dans un produit Flash NOR 130 nm

Les Si-NCs ont été intégrés sur un produit Atmel Flash NOR basé sur une technologie 130 nm. Le démonstrateur, d'une capacité de 32 Mb, a été obtenu à partir d'un produit Flash standard, dont le procédé de fabrication a été modifié afin de pouvoir intégrer les Si-NCs. Nous verrons d'abord comment s'organise un produit mémoire Flash de façon générale. Ensuite, nous détaillerons le procédé de fabrication de la mémoire standard et nous présenterons les différentes possibilités d'intégration des nanocristaux de silicium.

II.3.1 Organisation d'une mémoire Flash

L'architecture d'une mémoire Flash est assez complexe. La **Fig. II-9** présente les différents blocs composant un produit mémoire. Les éléments principaux sont le cœur, les décodeurs lignes et colonnes qui permettent l'adressage d'une ligne (« wordline ») et d'une colonne (« bitline ») données, les pompes de charge (capacités) qui permettent de générer la haute tension V_{pp} (HV : High Voltage), l'amplificateur de lecture (« sense amplifier ») qui permet la lecture des bits. Les tensions nécessaires pour les opérations de la mémoire sont appliquées sur les bitlines et wordlines par l'intermédiaire de transistors CMOS haute tension (tensions d'écriture et d'effacement) et basse tension (tensions de lecture).

Le cœur, composé du plan mémoire, est divisé en plusieurs secteurs. Dans le cas du produit dans lequel les Si-NCs ont été intégrés, la matrice de 32 Mb est divisée en 64 secteurs de 512 Kb chacun.

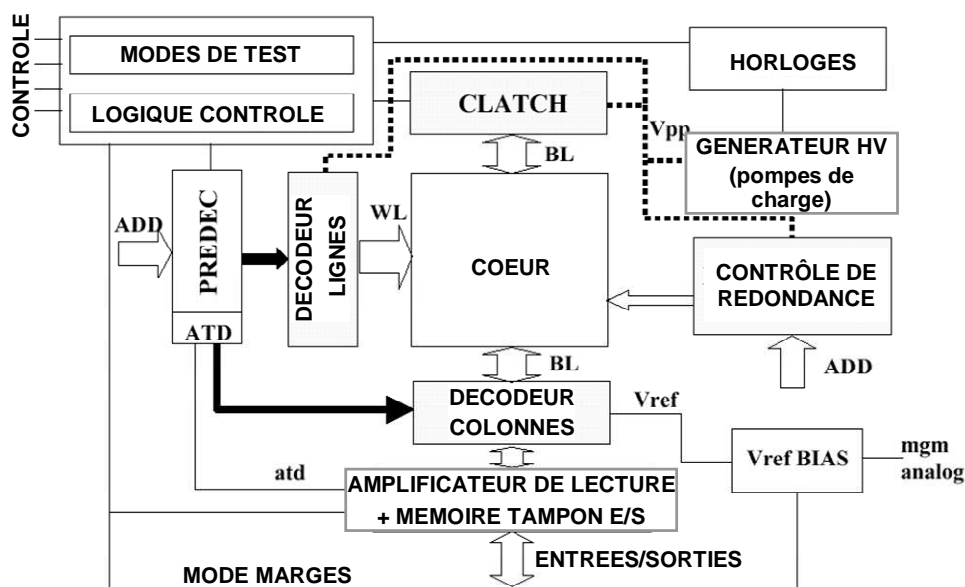


Fig. II-9 : Schéma illustrant l'architecture des blocs mémoires.

II.3.2 Fabrication d'une mémoire Flash standard

Nous allons décrire les principales étapes du procédé de fabrication du plan mémoire et de la périphérie du produit Flash NOR standard. En particulier, nous décrirons la fabrication des transistors NMOS et PMOS basse et haute tension et de deux cellules mémoires. On s'arrêtera au procédé front end (de la définition des zones actives à la définition des contacts) et nous ne décrirons pas le procédé de fabrication back end (niveaux métalliques et isolants), celui-ci n'ayant aucune conséquence sur l'intégration des nanocristaux.

Afin de faciliter la compréhension du procédé de fabrication, nous avons représenté des dessins des cellules mémoires standard sur la **Fig. II-10** selon différentes vues. On peut voir que la grille flottante est en deux parties. Selon la largeur W , la partie supérieure, de forme trapézoïdale, permet d'augmenter le couplage grille de contrôle/grille flottante. La dernière couche de polysilicium qui recouvre plusieurs cellules constitue la wordline (**Fig. II-10(a)**). Les cellules mémoires et les transistors sont isolés par des tranchées dans le silicium remplies d'oxyde (STI : Shallow Trench Isolation). Il faut noter que dans ce procédé de fabrication, les transistors de la logique en périphérie et les cellules mémoires sont réalisés grâce au procédé SASTI (Self Aligned STI) : les grilles polysilicium et les zones actives sont auto-alignées, ce qui permet d'augmenter la tolérance d'alignement et d'augmenter la densité de composants par des dimensions plus agressives.

Enfin, comme on peut le voir sur la **Fig. II-10(b)**, l'architecture NOR impose à deux cellules mémoires voisines de partager leur source (voir Chapitre I).

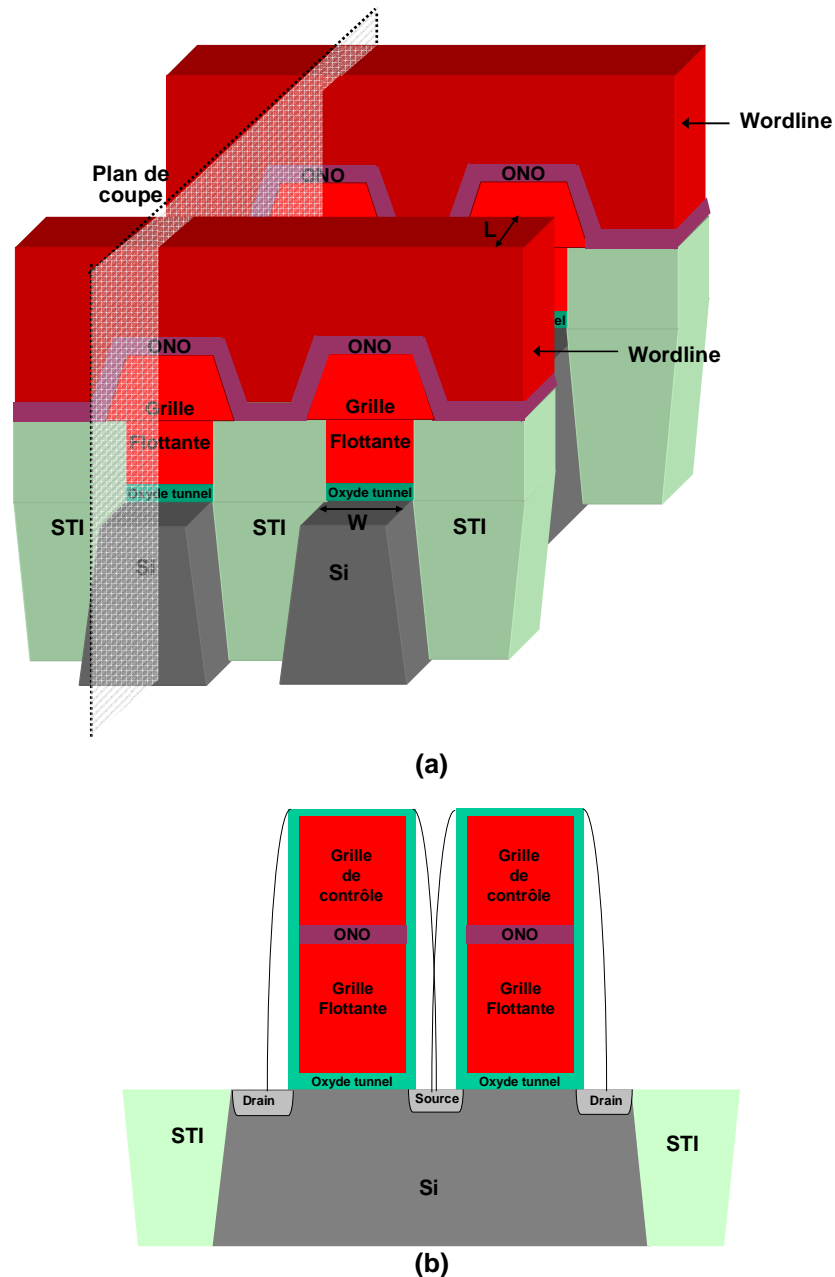


Fig. II-10 : Dessins du plan mémoire standard selon les wordlines (a) et selon le plan de coupe perpendiculaire aux wordlines (b).

La fabrication commence depuis un substrat silicium de type P. Un implant Nwell profond (« Deep NWell ») est d'abord réalisé dans le plan mémoire. Cet implant va permettre d'isoler le caisson de la mémoire du substrat des transistors de la périphérie. Ainsi, il sera possible de polariser le caisson de la mémoire (négativement pour l'écriture, positivement pour l'effacement) indépendamment du substrat des transistors haute et basse tension qui est à la masse. Ensuite, on réalise les implants caissons N et P des transistors et du plan mémoire (**Fig. II-11**).

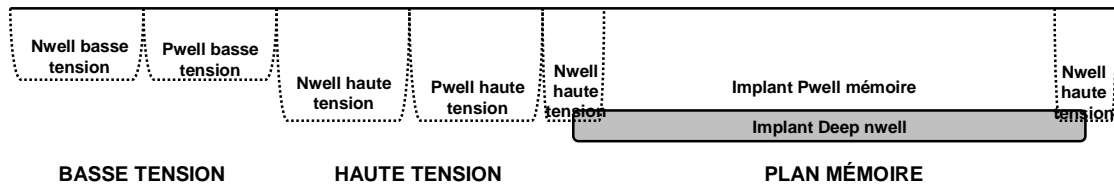


Fig. II-11 : Implants caissons transistors et mémoire.

La séquence suivante est la croissance des oxydes de grille et de l'oxyde tunnel du plan mémoire (**Fig. II-12**). On fait d'abord croître un oxyde d'environ 120 Å (**Fig. II-12(a)**). Ensuite, l'oxyde est gravé dans la partie mémoire et basse tension à l'aide d'un masque adapté (**Fig. II-12(b)**). Un oxyde est à nouveau réalisé (**Fig. II-12(c)**) puis gravé dans la partie basse tension (**Fig. II-12(d)**). Enfin, une dernière oxydation est effectuée (**Fig. II-12(e)**), qui aboutit aux épaisseurs finales : 30 Å pour l'oxyde de grille des transistors basse tension, 180 Å pour l'oxyde de grille des transistors haute tension et environ 100 Å pour l'oxyde tunnel dans le plan mémoire. L'épaisseur fine de l'oxyde de grille des transistors basse tension permet à ces dispositifs d'être rapides. Les transistors haute tension, eux, doivent être robustes, d'où un oxyde de grille épais.

Ensuite, une première couche de polysilicium est déposée (**Fig. II-13**).

La séquence de formation des tranchées isolantes (STI) est alors réalisée (**Fig. II-14**). Comme nous l'avons fait remarquer précédemment, le procédé est auto-aligné. La gravure des isolations intervient après la formation des différents oxydes et le dépôt d'une couche de polysilicium. Pour réaliser la gravure, on fait d'abord croître un oxyde sur le polysilicium, puis on dépose une couche de nitrure et d'oxyde qui joue le rôle de masque dur et enfin on procède à l'étape de lithographie des zones actives (**Fig. II-14(a)**). Les tranchées sont ensuite gravées (**Fig. II-14(b)**), puis le masque dur en s'arrêtant sur le nitrure. L'étape suivante est une oxydation qui permet d'obtenir ce que l'on appelle un « liner » le long des zones actives, puis un dépôt d'oxyde HDP (High Density Plasma) pour « remplir » les tranchées STI que l'on densifie sous O₂. Un polissage CMP est réalisé pour planariser la surface. (**Fig. II-14(c)**). Le liner étant un oxyde thermique (donc de bonne qualité) permet d'avoir également une interface active/oxyde HDP de bonne qualité. L'oxyde HDP est gravé jusqu'au bas du nitrure (**Fig. II-14(d)**), puis les couches de nitrure et d'oxyde thermique sont enlevées afin de revenir sur le polysilicium 1 (**Fig. II-14(e)**).

Une deuxième couche de polysilicium est déposée pour former la wordline (**Fig. II-15**) puis dopée N⁺ dans le plan mémoire et pour les transistors NMOS de la logique (**Fig. II-16**). Les couches de polysilicium 1 et 2 vont constituer la grille flottante des cellules mémoires. La partie supérieure de la grille flottante est gravée selon la **Fig. II-17**. L'empilement de grille des cellules mémoires est achevé avec le dépôt du diélectrique interpoly ONO et d'une troisième couche de polysilicium (**Fig. II-18**). Cette couche formant la grille de contrôle de la mémoire est ensuite dopée (**Fig. II-19**).

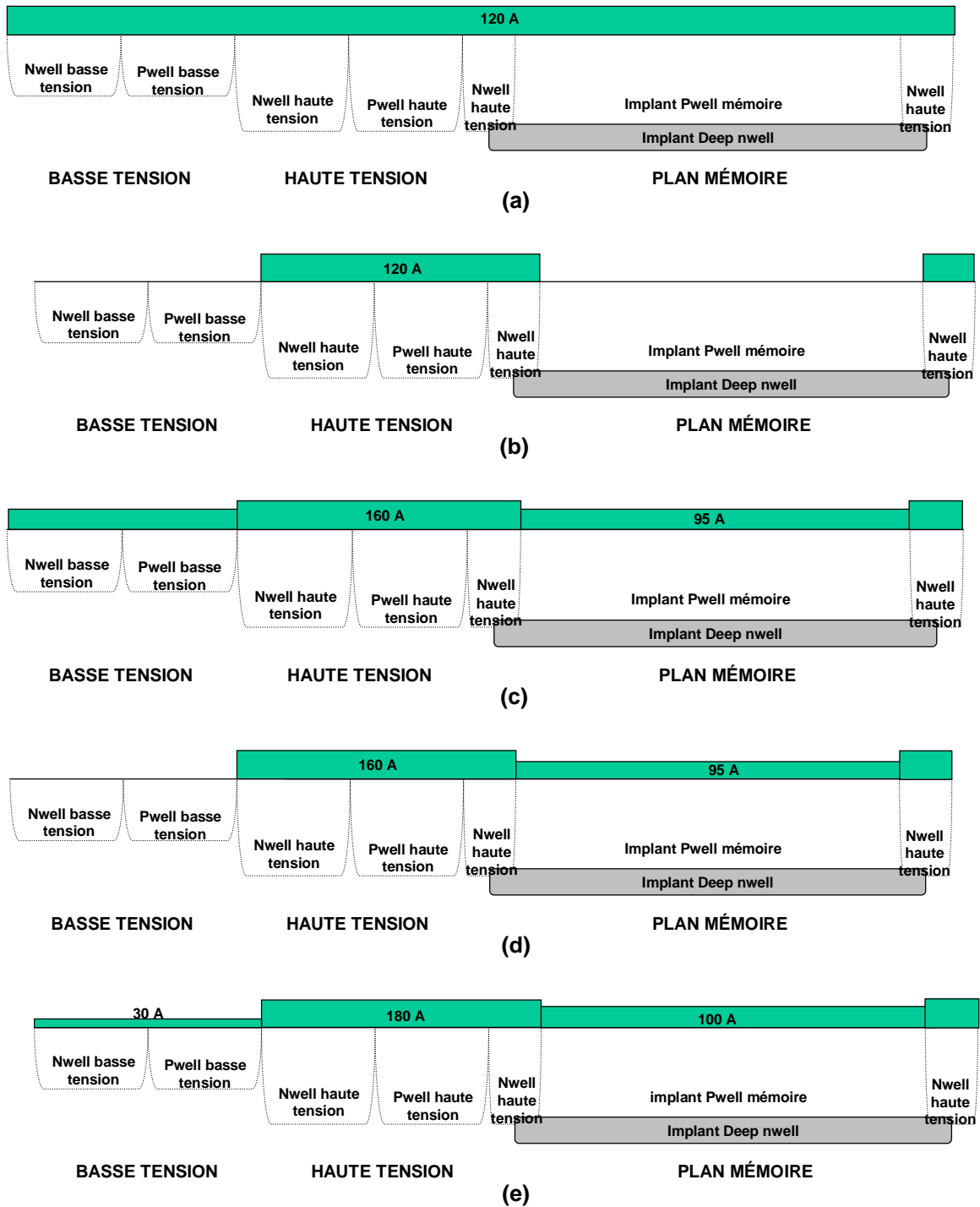


Fig. II-12 : Croissance des oxydes de grilles des transistors de la périphérie et de l'oxyde tunnel des cellules mémoires.

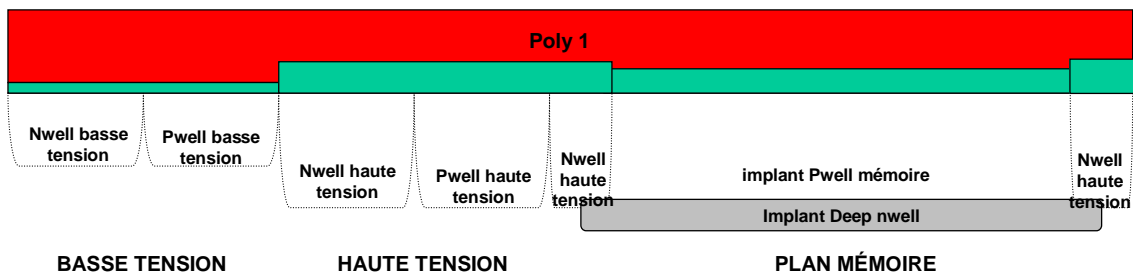


Fig. II-13 : Dépôt de la première couche de polysilicium.

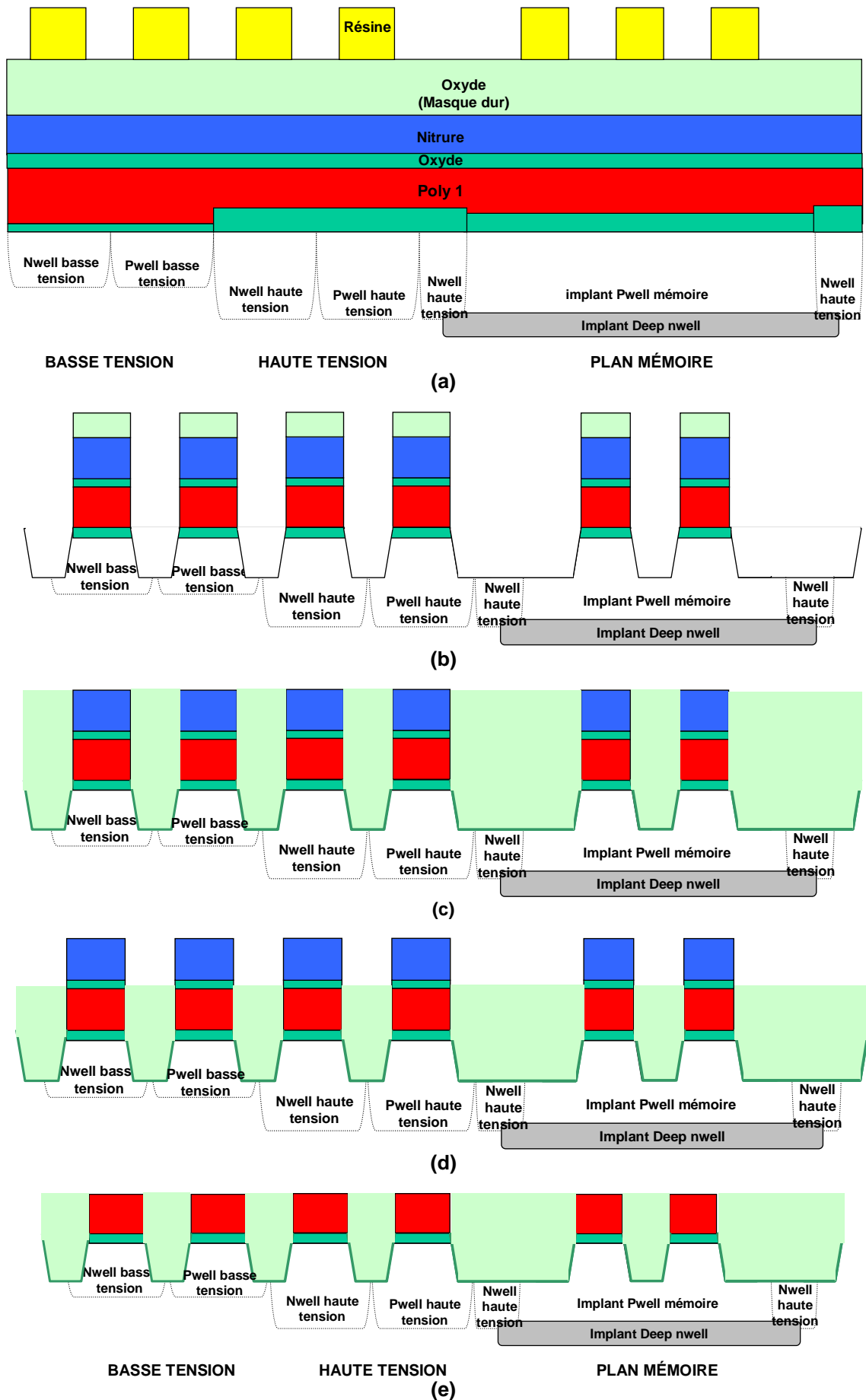


Fig. II-14 : Formation des STI (Shallow Trench Isolation).

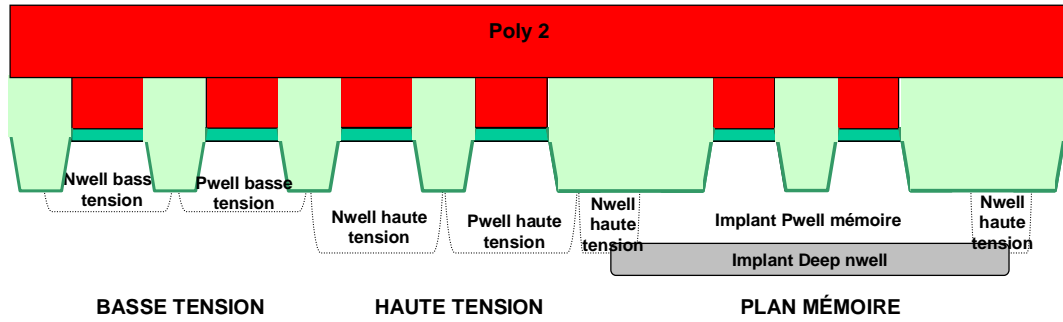


Fig. II-15 : Dépôt du polysilicium 2.

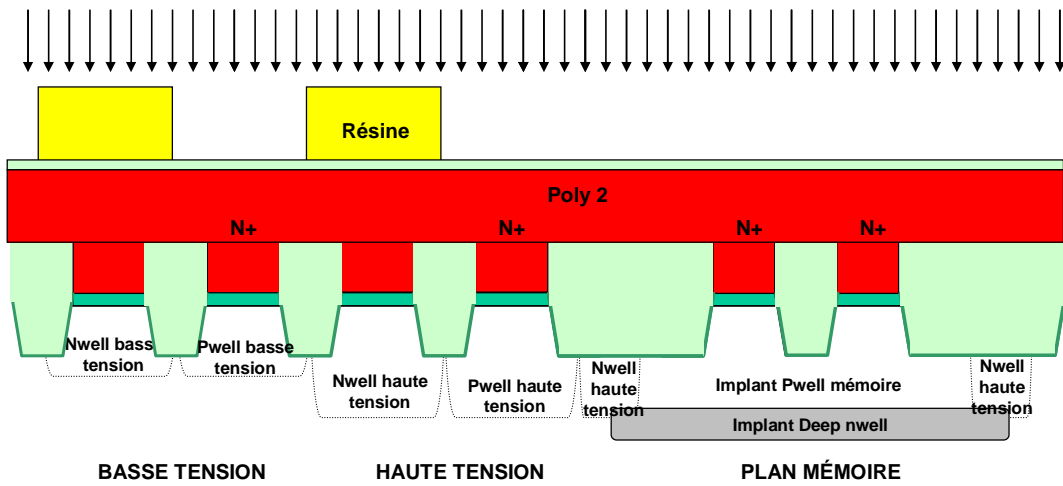


Fig. II-16 : Implantation N+ des grilles des NMOS de périphérie et des grilles flottantes des cellules mémoires.

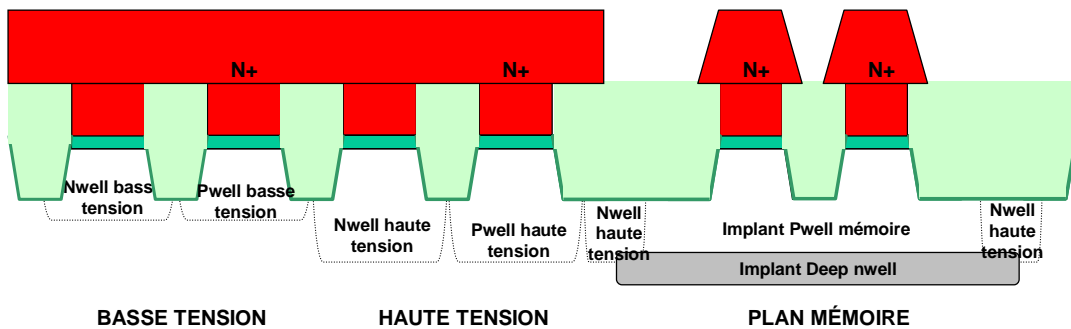


Fig. II-17 : Gravure de la partie supérieure de la grille flottante.

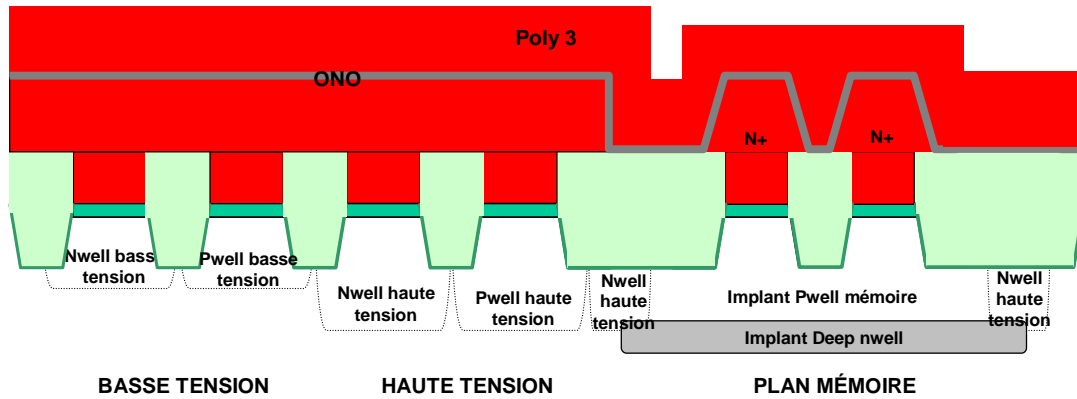


Fig. II-18 : Fin de l'empilement de grille : diélectrique de grille ONO de la mémoire + polysilicium 3 (grille de contrôle de la mémoire).

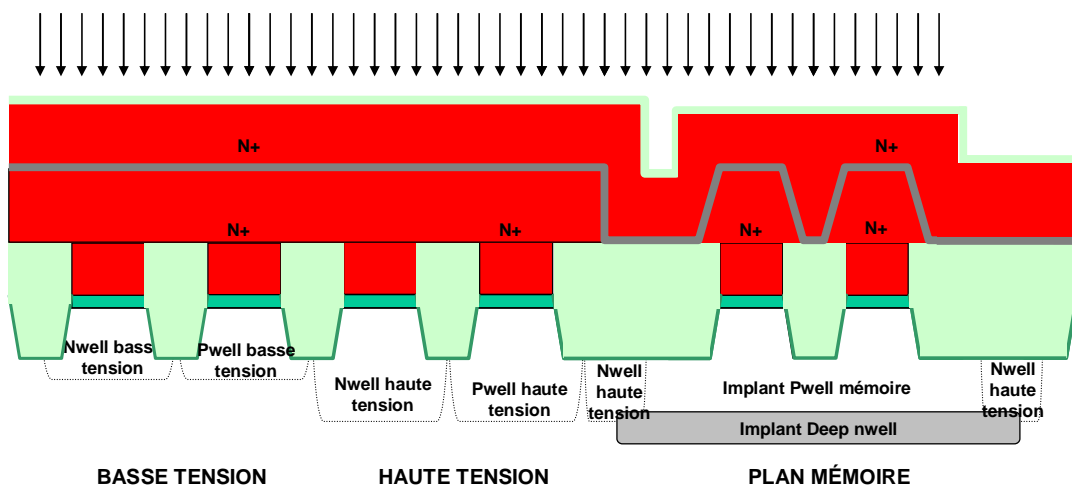


Fig. II-19 : Implantation N+ du polysilicium 3 (grille de contrôle de la mémoire).

Dans les étapes suivantes, le plan mémoire et la périphérie seront représentés dans le plan perpendiculaire à la wordline. La séquence qui suit est la formation des cellules mémoires. Une lithographie suivie de la gravure de l'empilement Polysilicium 3/ONO/Polysilicium 2/Polysilicium 1 sont réalisées afin de définir les cellules mémoires selon la longueur (**Fig. II-20(a)**). Ensuite, un premier implant drain est réalisé. Les lignes source sont gravées et l'implant source est réalisé (**Fig. II-20(b)**). Enfin, on réalise une oxydation thermique afin d'obtenir une couche d'oxyde autour de l'empilement de grille des cellules mémoires (**Fig. II-20(c)**). Cette couche d'oxyde permet d'isoler latéralement la grille flottante afin de ne pas perdre la charge stockée.

Le polysilicium 3 et l'ONO sont gravés dans la périphérie (**Fig. II-21**). Puis c'est la formation des transistors haute et basse tension. Une étape de lithographie définissant les grilles de contrôle des transistors selon leur longueur précède la gravure des polysilicium 2 puis 1 (**Fig. II-22(a)**). Ensuite, on fait croître un oxyde autour de la grille de contrôle et on réalise les implants LDD (Low-Doped Drain) N (pour les NMOS) et P (pour les PMOS) (**Fig. II-22(b)**).

La dernière séquence front-end est la formation des espaces en nitrure des transistors et des cellules mémoires (**Fig. II-23(a)**), suivie des implants N source et drain des NMOS et de l'implant drain des cellules mémoires et des implants P source et drain des transistors PMOS.

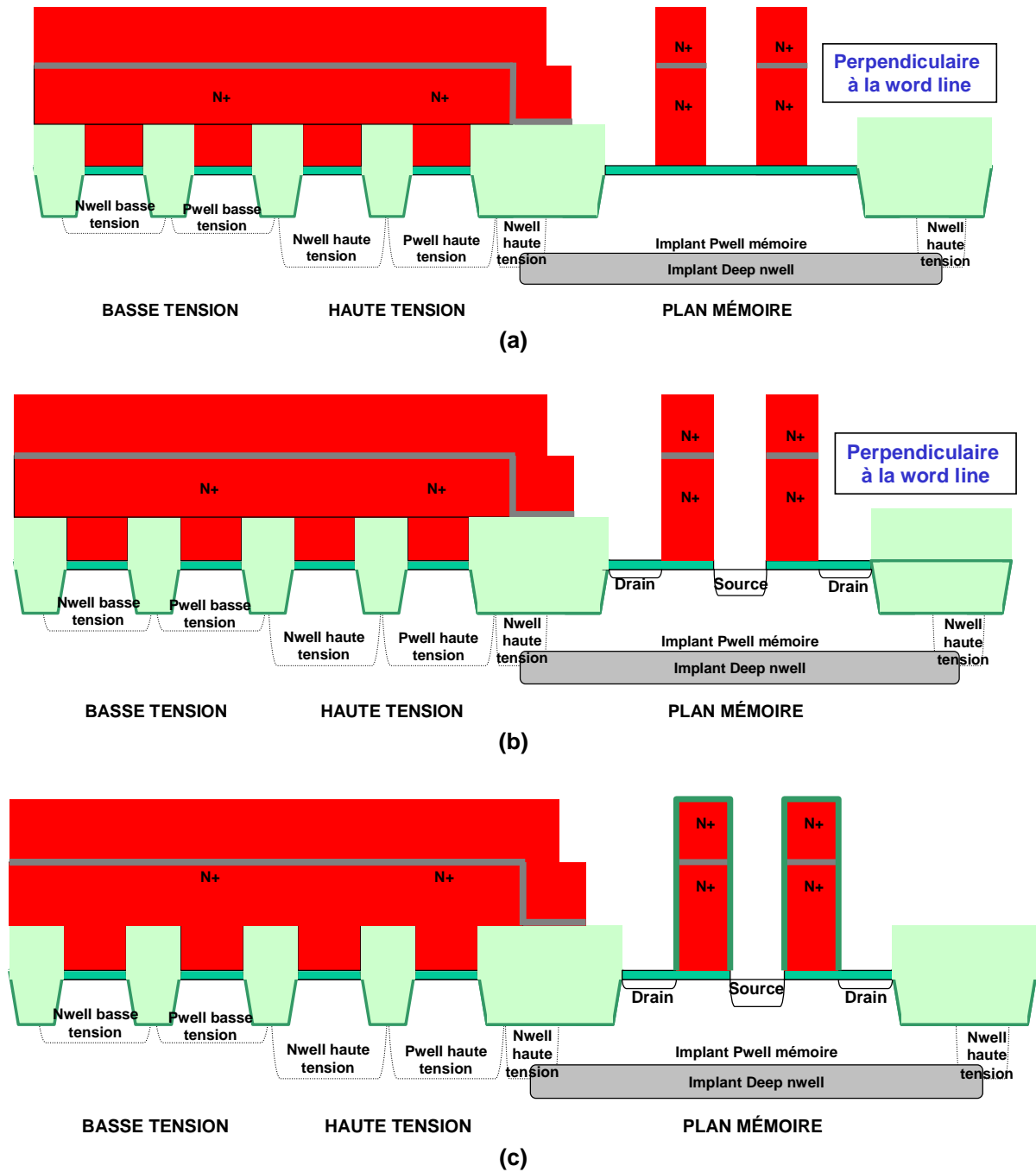


Fig. II-20 : Définition des cellules mémoires.

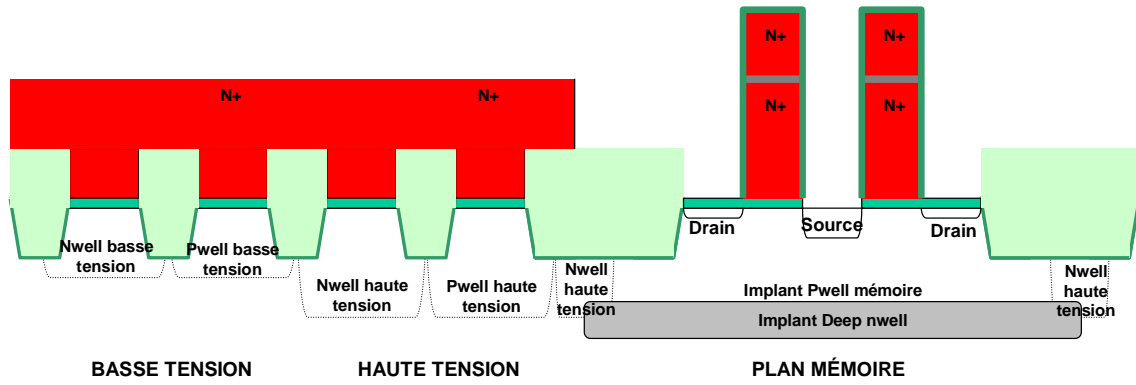
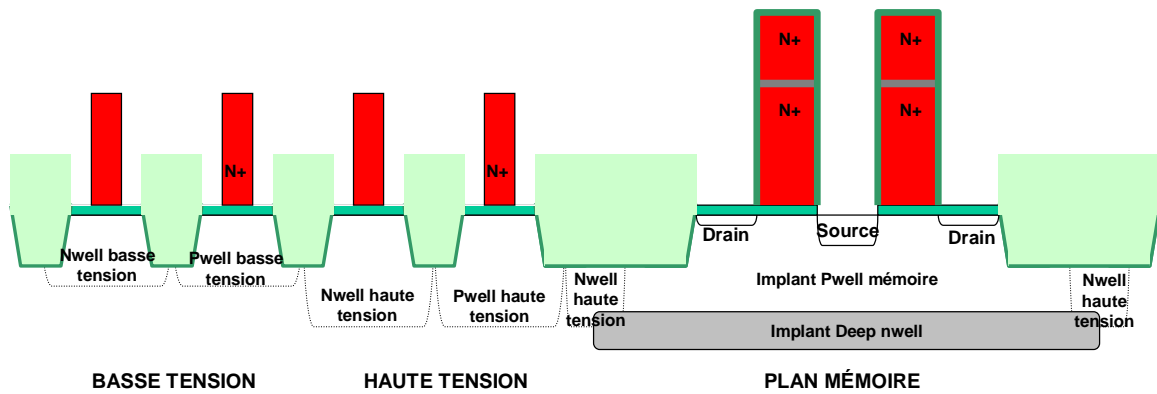
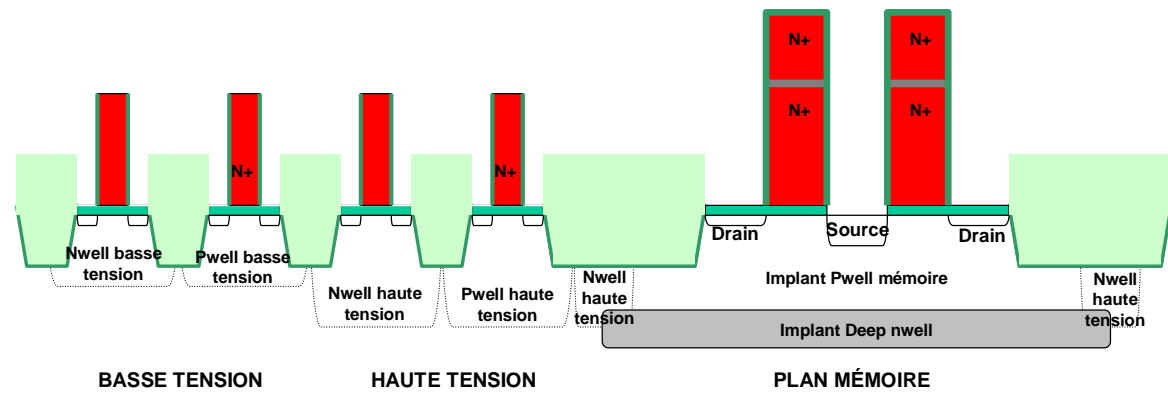


Fig. II-21 : Gravure Polysilicium 3 + ONO en périphérie.



(a)



(b)

Fig. II-22 : Définition des transistors de périphérie : Gravure polysilicium (a) et implantations LDD des transistors de périphérie (b).

La première possibilité est la plus simple à mettre en oeuvre car c'est la plus proche du procédé standard. Après avoir réalisé les implants caissons de manière identique au procédé standard, on fait croître l'oxyde tunnel de la mémoire. Ensuite, les Si-NCs puis l'oxyde de contrôle sont déposés (**Fig. II-24**). L'oxyde de contrôle peut être soit un tri-couche ONO, soit un HTO.

Ensuite, l'empilement est gravé dans la périphérie (**Fig. II-25**) et on procède à la séquence de croissance des oxydes de grille des transistors haute et basse tension d'une façon proche du procédé standard (**Fig. II-26**).

Les étapes suivantes rejoignent le procédé de fabrication standard, la gravure de la grille flottante en moins : dépôt du polysilicium 1 (**Fig. II-27**), formation des STI (**Fig. II-28**), dépôt et dopage de la couche de polysilicium 2 (**Fig. II-29**), formation des cellules mémoires selon la longueur (**Fig. II-30**), des transistors de périphérie et fin du procédé front end (**Fig. II-31**).

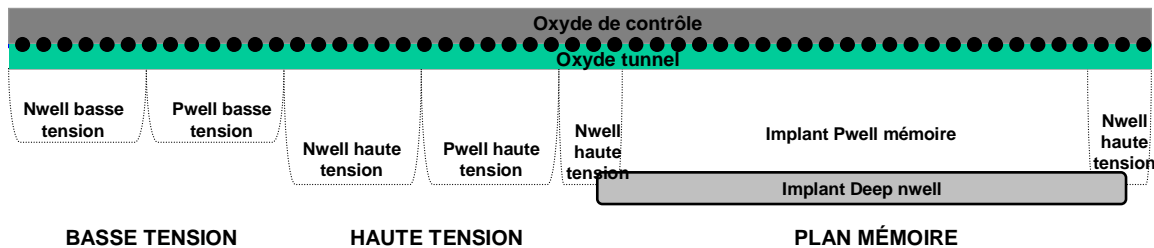


Fig. II-24 : Croissance de l'oxyde tunnel, dépôt des Si-NCs et de l'oxyde de contrôle.

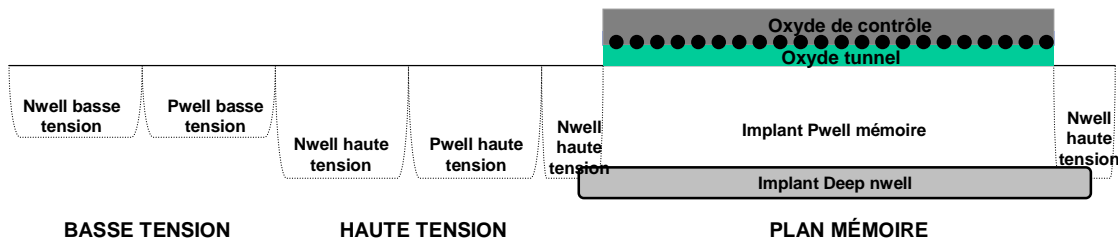


Fig. II-25 : Gravure de l'empilement mémoire dans la périphérie.

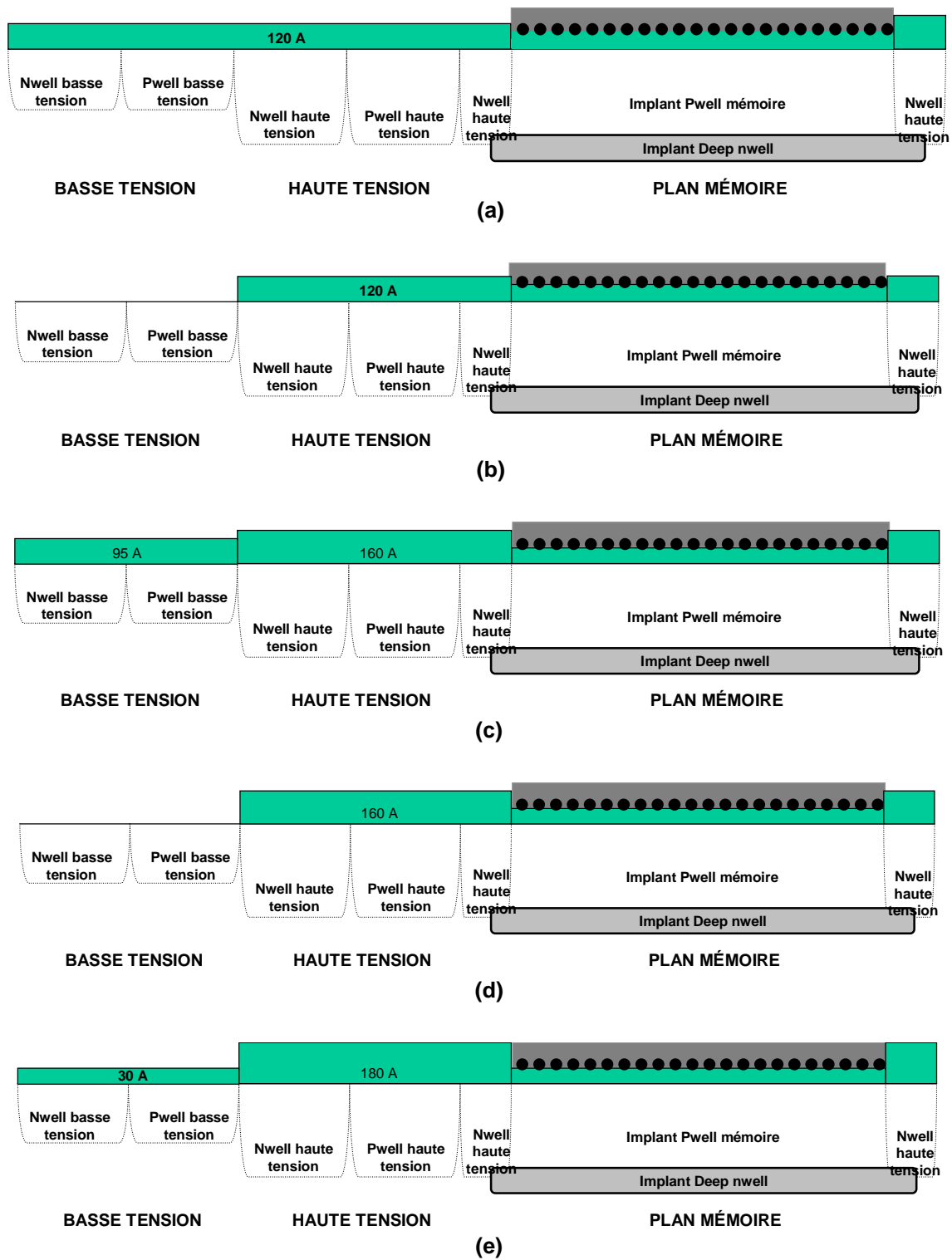


Fig. II-26 : Croissance des oxydes de grille des transistors haute et basse tension. Croissance d'un oxyde de 120 Å (a), gravure dans la partie basse tension (b), croissance d'un oxyde (160 Å dans la partie haute tension) (c), gravure dans la partie basse tension (d), croissance de l'oxyde de grille basse tension (e).

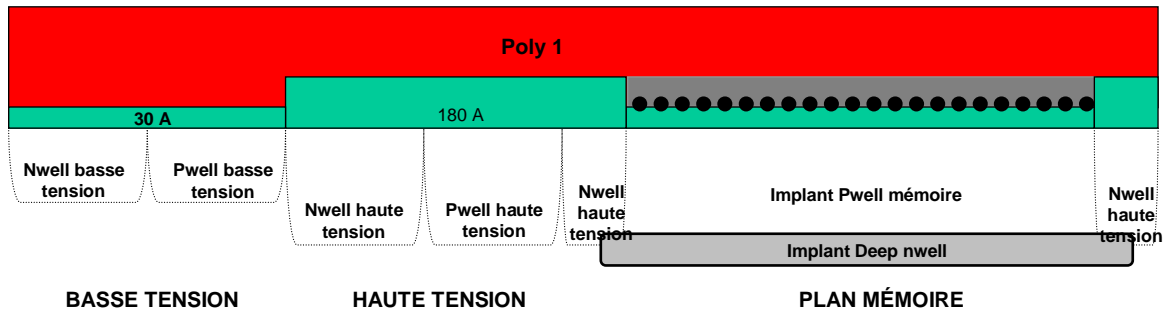


Fig. II-27 : Dépôt de la couche de polysilicium 1.

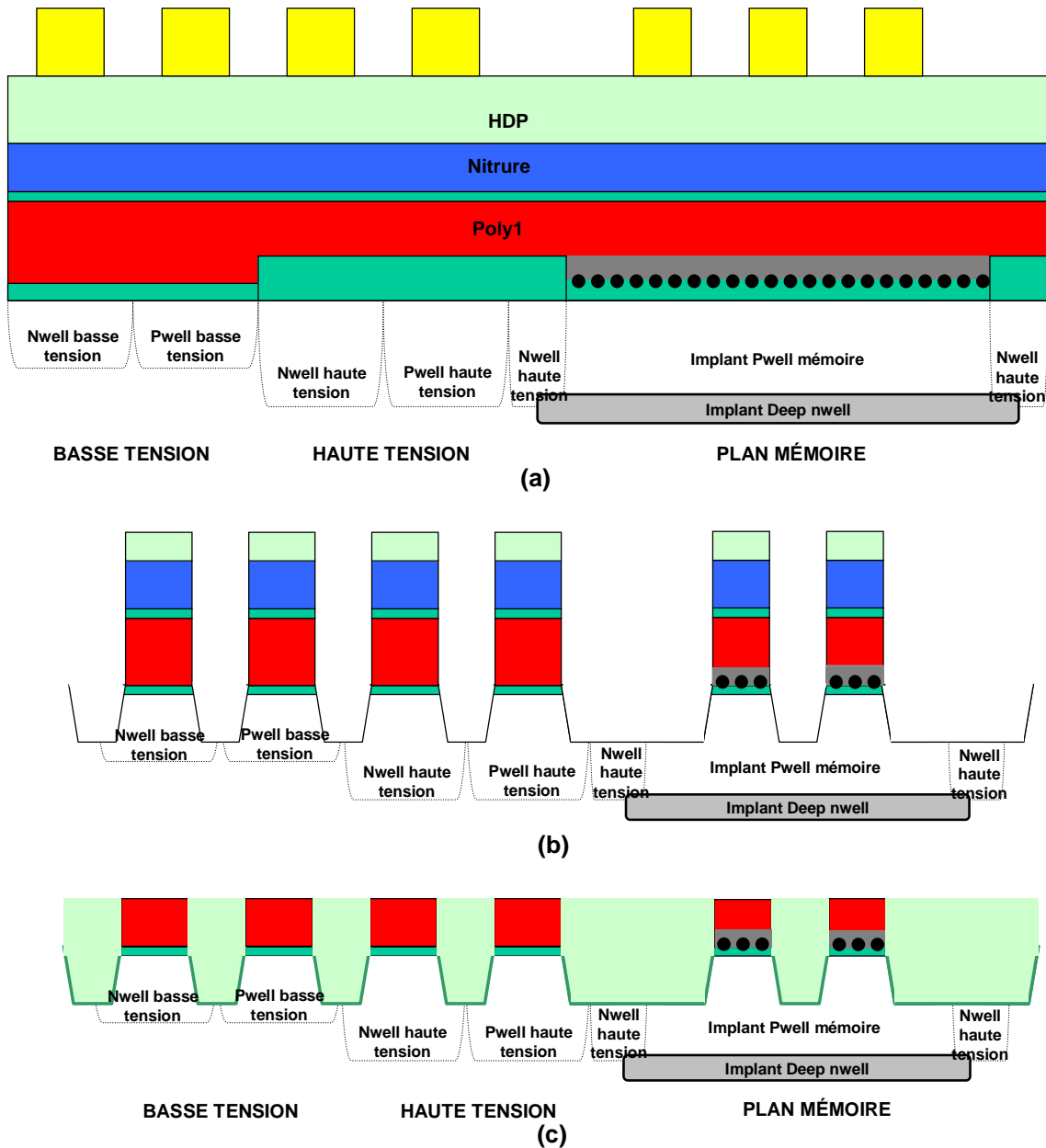


Fig. II-28 : Formation des STI. Croissance d'un oxyde, dépôt d'une couche de nitrure, d'un oxyde HDP et de la résine puis lithographie des zones actives (a). Gravure des zones actives (b). Oxydation liner, remplissage des STI par un oxyde HDP, densification HDP puis gravure jusqu'au polysilicium (c).

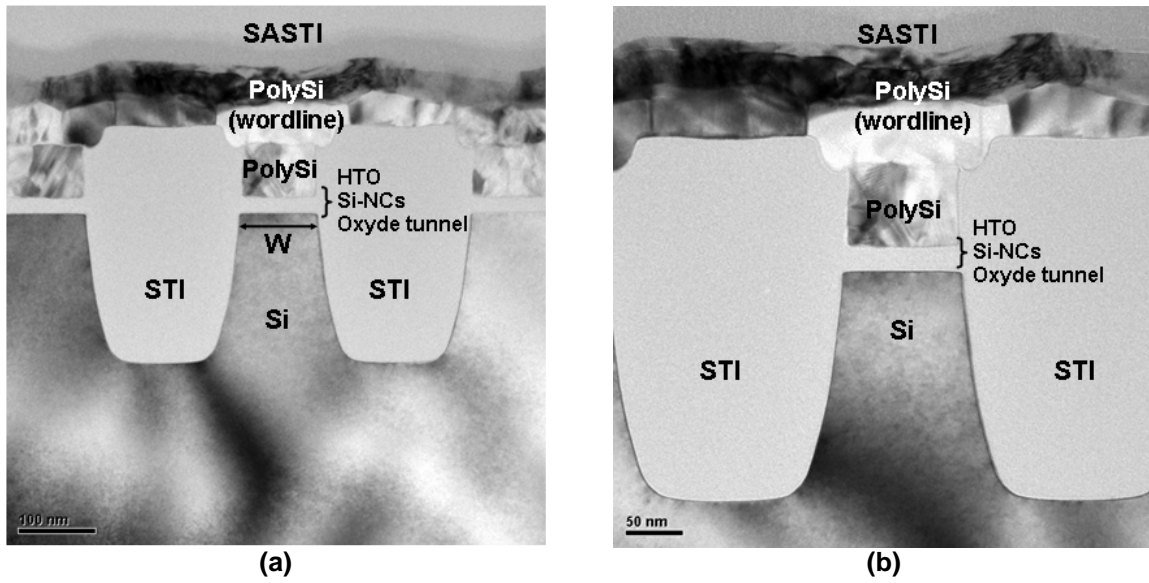


Fig. II-32 : Coupes TEM dans le plan mémoire selon la largeur des cellules avec un HTO comme oxyde de grille. (b) : agrandissement de (a).

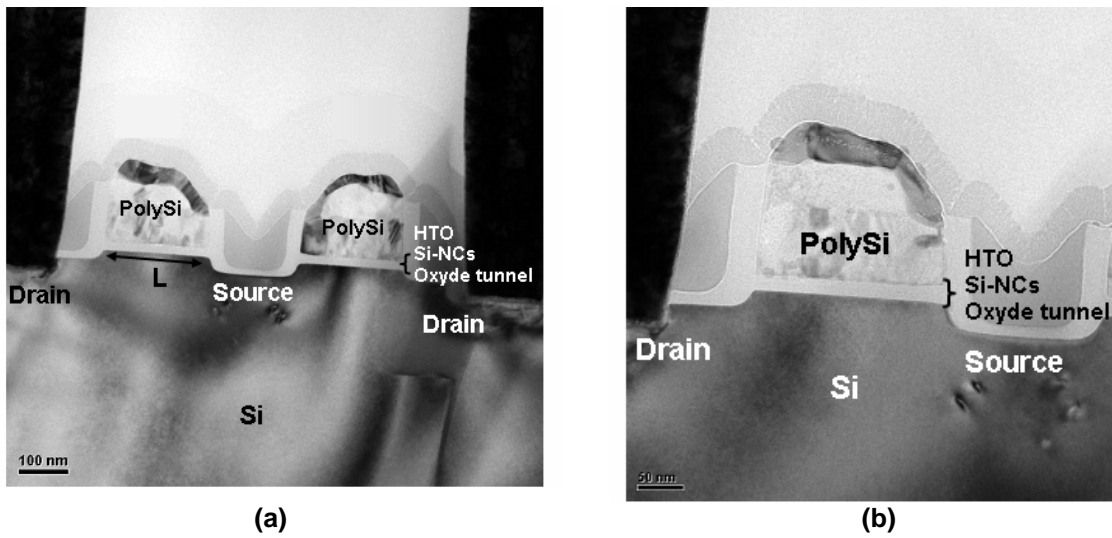


Fig. II-33 : Coupes TEM dans le plan mémoire selon la longueur des cellules avec un HTO comme oxyde de grille. (b) : agrandissement de (a).

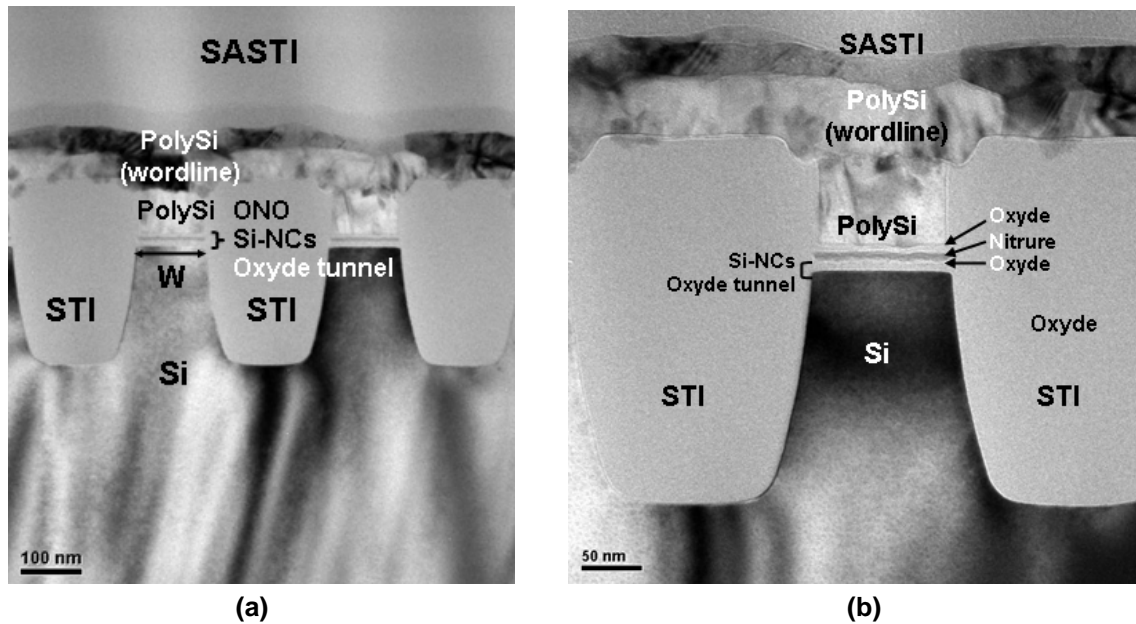


Fig. II-34 : Coupes TEM dans le plan mémoire selon la largeur des cellules avec un ONO comme oxyde de grille. (b) : agrandissement de (a).

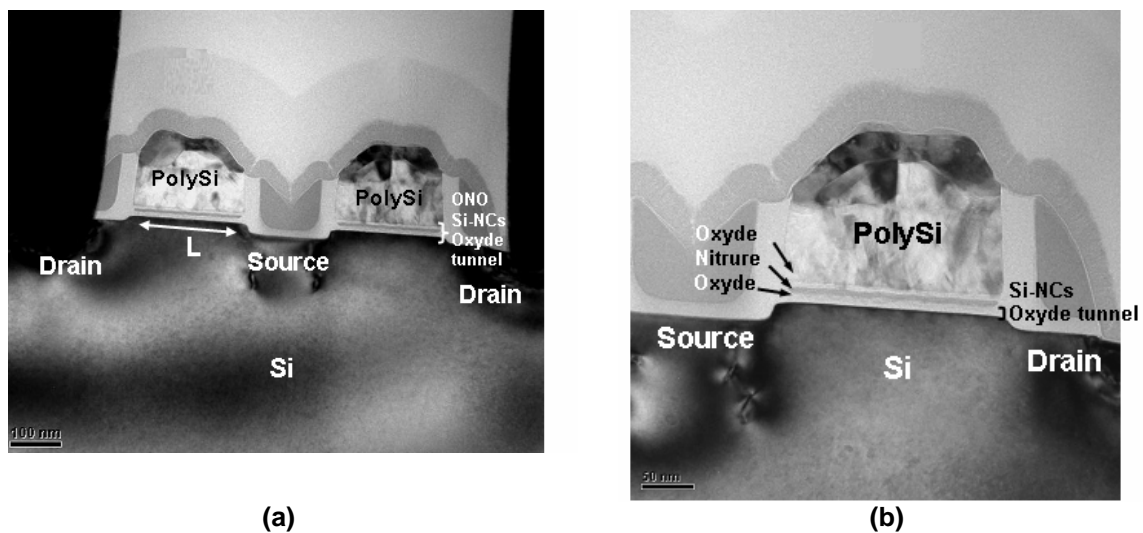


Fig. II-35 : Coupes TEM dans le plan mémoire selon la longueur des cellules avec un ONO comme oxyde de grille. (b) : agrandissement de (a).

Ce procédé est simple et très proche du procédé standard, mais il a un inconvénient majeur qui est le fait de réaliser les oxydes de grille de la périphérie après les Si-NCs qui ne sont couverts que par l'oxyde de grille. Or, ces oxydations sont très fortes et dans le cas d'un HTO comme oxyde de contrôle, les Si-NCs seront certainement partiellement ou totalement oxydés. Nous montrerons dans le chapitre III, qu'une oxydation correspondant à la croissance un oxyde de 140 Å (rappelons que l'oxyde de grille haute tension fait environ 180 Å), oxyde déjà fortement les nanocristaux. Il a donc été nécessaire de réfléchir à un deuxième procédé de fabrication dans lequel les oxydes de la périphérie seraient réalisés avant les Si-NCs.

Ce procédé de fabrication commence de façon strictement identique au procédé standard, avec notamment, la séquence de croissance des oxydes de grille haute et basse tension et celle de l'oxyde tunnel de la mémoire flash à grille flottante continue (**Fig. II-36**), le dépôt d'une couche de polysilicium (**Fig. II-37**) et la séquence de formation des tranchées isolantes de la gravure des zones actives, jusqu'au retour sur polysilicium (**Fig. II-38**).

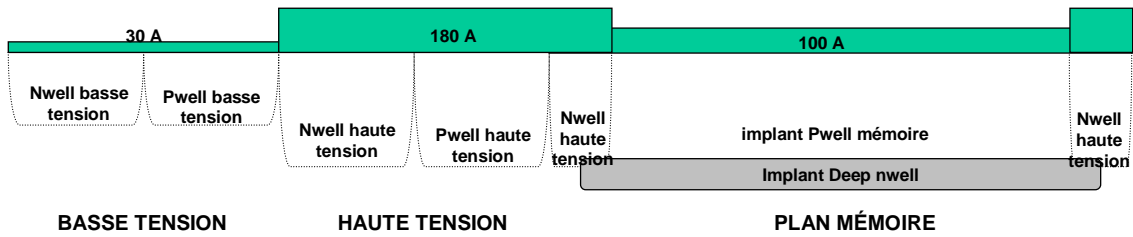


Fig. II-36 : Croissance des oxydes de grille.

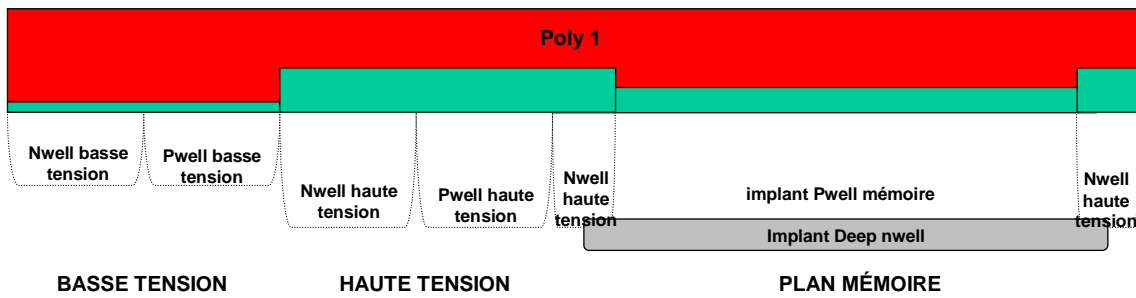


Fig. II-37 : Dépôt de la couche de polysilicium 1.

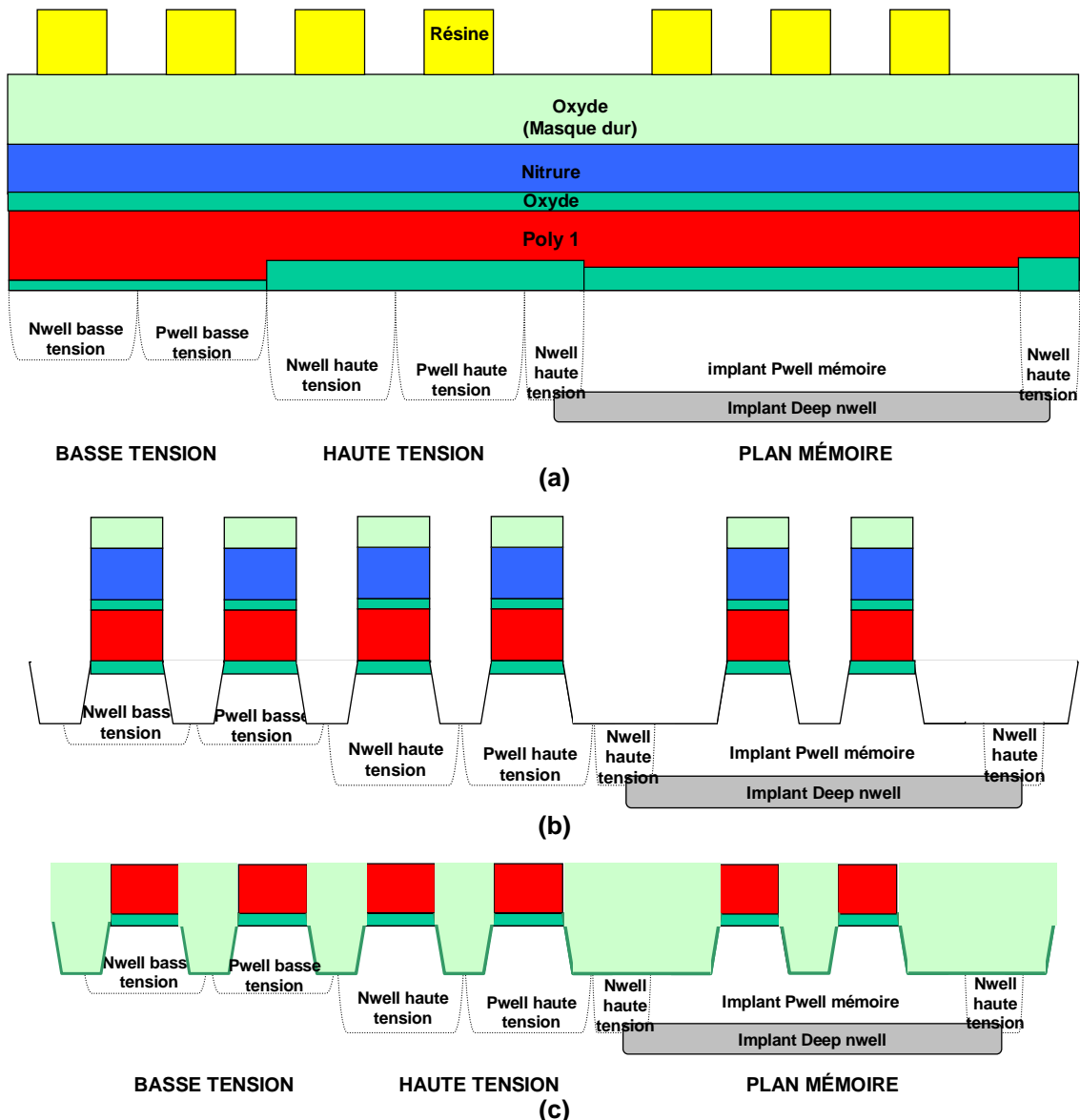


Fig. II-38 : Formation des STI. Croissance d'un oxyde, dépôt d'une couche de nitrure, d'un oxyde HDP et de la résine puis lithographie des zones actives (a). Gravure des zones actives (b). Oxydation liner, remplissage des STI par un oxyde HDP, densification HDP puis gravure jusqu'au polysilicium (c).

C'est ensuite que le procédé doit être modifié, afin de réaliser l'empilement de grille du plan mémoire. Pour cela, il faut tout d'abord enlever le polysilicium puis l'oxyde dans le plan mémoire. Ceci est réalisé au moyen d'une gravure de type damascène (**Fig. II-39**). Ensuite, on peut alors réaliser l'empilement de grille : croissance de l'oxyde tunnel (plus fin que l'oxyde tunnel de la mémoire standard), dépôt des Si-NCs puis de l'oxyde de contrôle HTO ou ONO (**Fig. II-40**). Evidemment, l'empilement de grille se retrouve également sur la périphérie. Il faut donc le graver dans cette partie afin de revenir sur le polysilicium 1 tel que le montre la **Fig. II-41**.

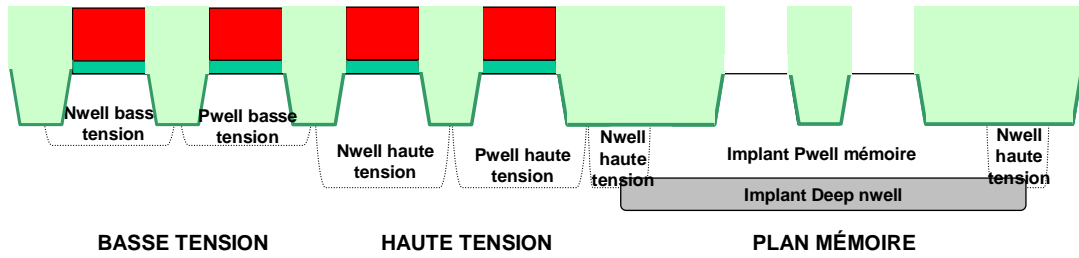


Fig. II-39 : Gravure damascène polysilicium + oxyde selon les actives dans le plan mémoire.

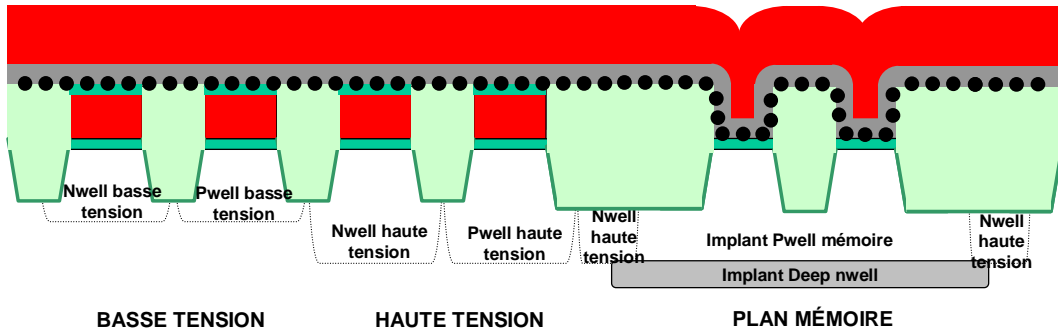


Fig. II-40 : Croissance de l'oxyde tunnel, dépôt des Si-NCs, de l'oxyde de contrôle et d'une couche de polysilicium 1.

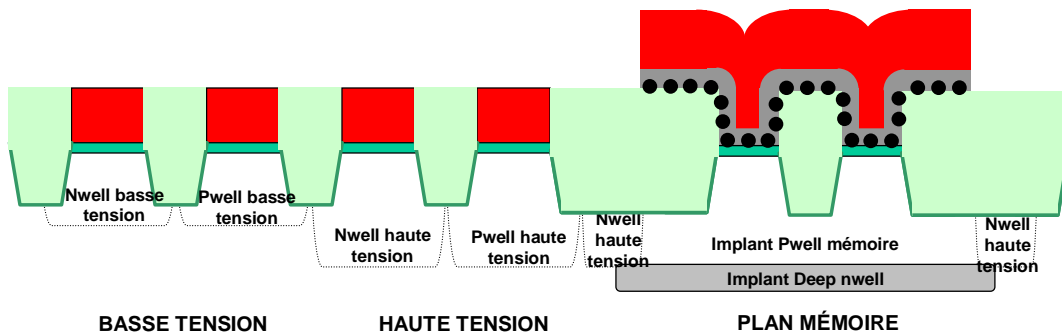


Fig. II-41 : Gravure de l'empilement mémoire (oxyde/Si-NCs/oxyde de contrôle/polysilicium) dans la périphérie.

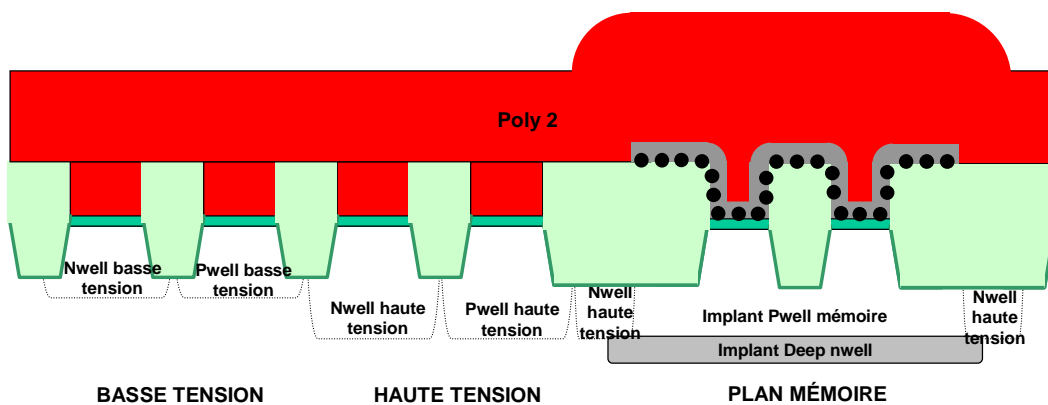


Fig. II-42 : Dépôt de la couche de polysilicium 2.

Après cette gravure, on dépose une deuxième couche de polysilicium (**Fig. II-42**) et on rejoint à nouveau le procédé standard (sauf la gravure de la grille flottante), comme c'était le cas dans le procédé précédent pour parvenir de la même manière à la **Fig. II-43**.

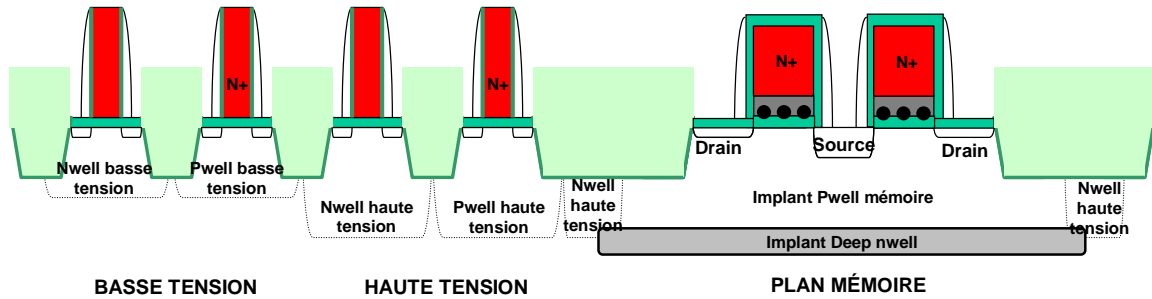


Fig. II-43 : Fin du procédé front-end.

Ce procédé a donc l'avantage de contourner le problème de l'oxydation des Si-NCs. On peut remarquer une autre particularité de ce procédé. La gravure des zones actives est réalisée après l'empilement de grille des transistors haute et basse tension, mais avant l'empilement de grille de la mémoire. Cela implique que les transistors de la périphérie sont toujours auto-alignés (SASTI), comme dans le procédé standard, alors que les cellules mémoires ne le sont plus (non SASTI). Ceci apparaît clairement sur les coupes TEM réalisées le long de la largeur des **Fig. II-44** et **Fig. II-45** (qui correspondent à la **Fig. II-42**).

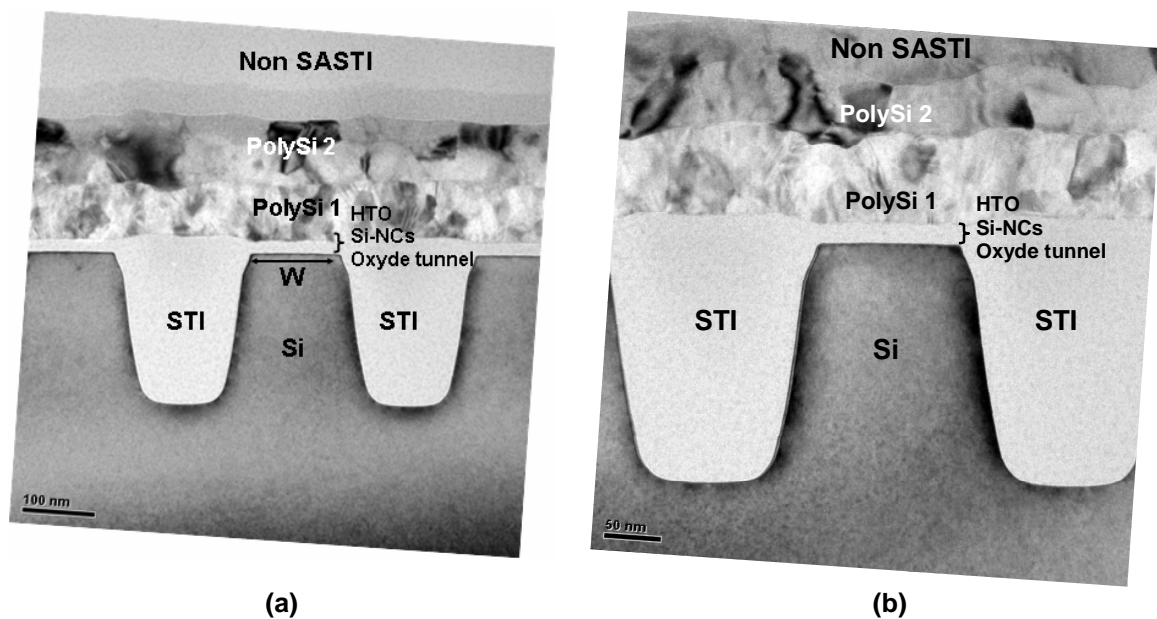


Fig. II-44 : Coupes TEM dans le plan mémoire selon la largeur des cellules avec un HTO comme oxyde de grille. (b) : agrandissement de (a).

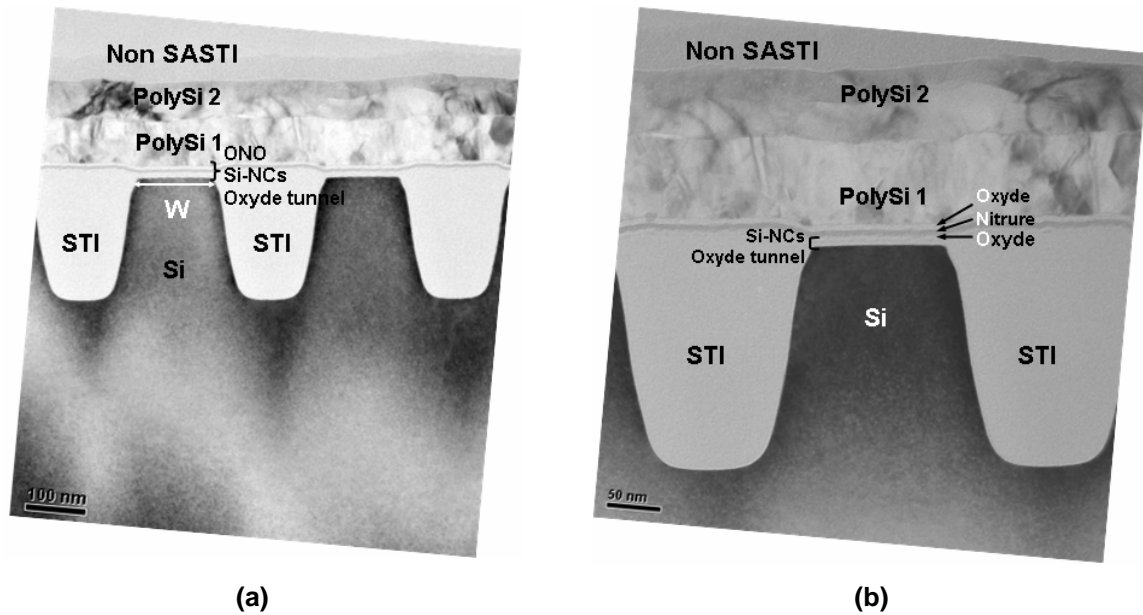


Fig. II-45 : Coupes TEM dans le plan mémoire selon la largeur des cellules avec un ONO comme oxyde de grille. (b) : agrandissement de (a).

Deux procédés différents ont donc été réalisés dans le but d'intégrer les nanocristaux de silicium. Le premier est le plus simple et semblable au procédé standard. Les transistors de la périphérie et les cellules mémoires sont auto-alignés (**Fig. II-46(a)**). Cependant, les risques d'oxydation des Si-NCs sont très forts. Le deuxième procédé évite toute oxydation parasite des Si-NCs. Dans ce procédé, comme nous l'avons déjà expliqué, la périphérie est auto-alignée et les cellules mémoires ne le sont pas (**Fig. II-46(a)**). Pour les cellules à grille flottante continue, l'alignement du polysilicium de la grille flottante avec l'active devient critique lors de la réduction des dimensions de la cellule. L'auto-alignement joue donc un rôle important. Lorsque l'on remplace la grille flottante par des Si-NCs, l'auto-alignement n'est plus nécessaire. De plus, si les cellules sont auto-alignées, cela implique d'effectuer les oxydations liées à la gravure des STI (oxydation liner et densification de l'oxyde HDP) après avoir déposé les Si-NCs (**Fig. II-46(c)**). Nous avons donc fait le choix d'un procédé non auto-aligné pour la mémoire qui paraissait le plus sûr pour les Si-NCs. On peut cependant noter que des résultats électriques ont depuis montré que ces deux oxydations n'étaient pas critiques pour les Si-NCs (**Fig. II-47**).

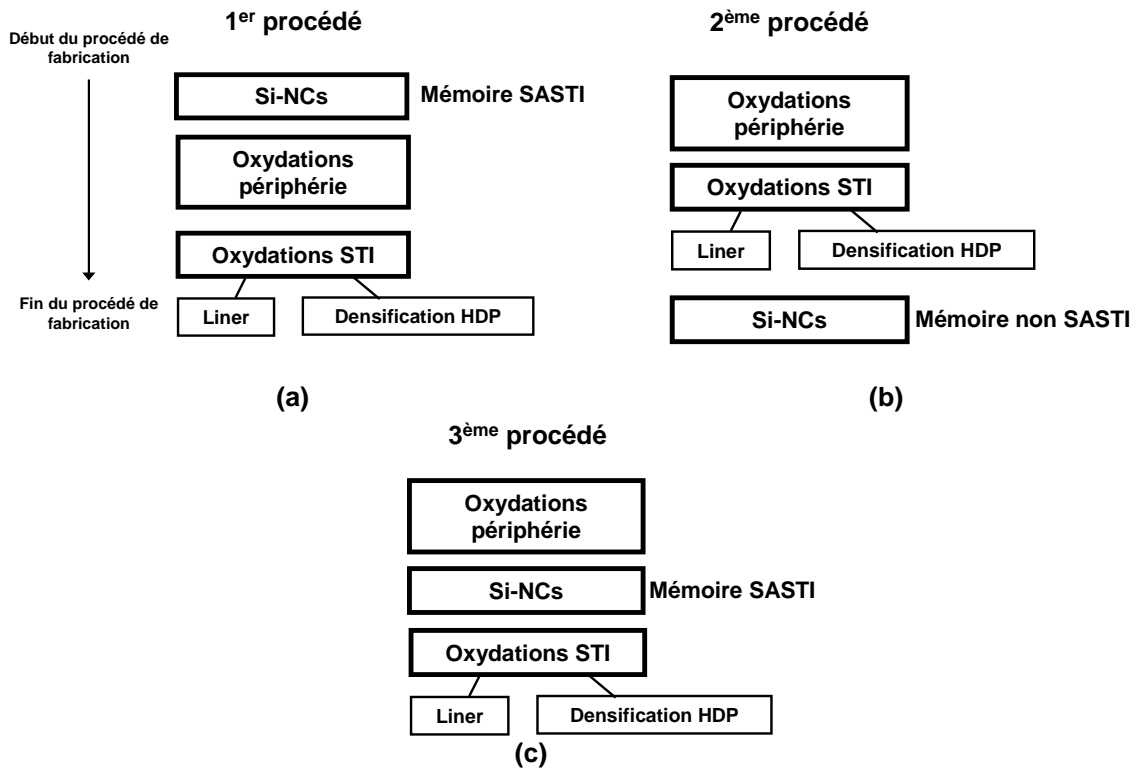


Fig. II-46 : Représentation schématique de la succession des étapes d'oxydation en fonction des différents procédés de fabrication intégrant les Si-NCs.

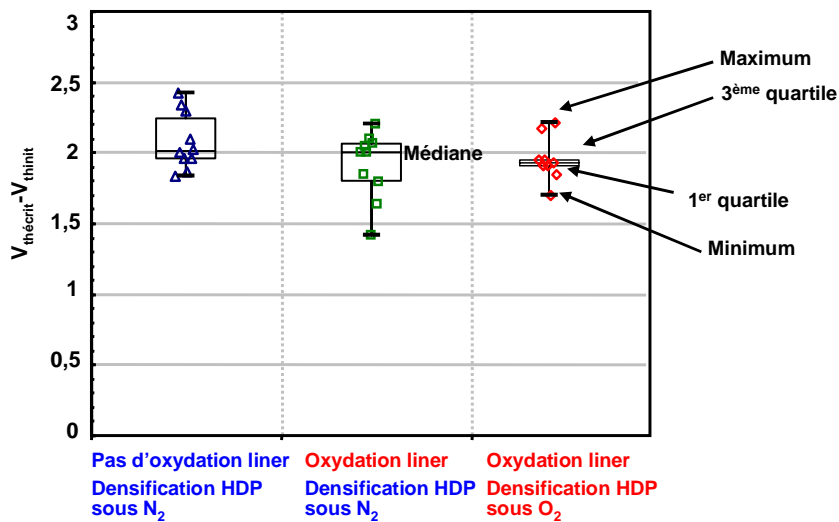


Fig. II-47 : Mesures de la fenêtre de programmation sur 10 cellules en fonction des oxydations réalisées. L'oxydation liner est réalisée sous O_2 ou pas du tout. La densification HDP est réalisée sous oxygène ou bien sous azote (non oxydante). Les symboles représentent les mesures expérimentales des 10 dispositifs. Ecriture : $V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10$ μ s. Lecture : $V_d=1$ V et $I_d=1$ μ A.

Il aurait donc été possible d'imaginer un procédé dans lequel la périphérie et la mémoire étaient auto-alignées, tout en réalisant la croissance des oxydes de grille de la périphérie avant les Si-NCs. Ce procédé est décrit dans le **paragraphe A** de l'**Annexe 1**

Enfin, habituellement, ce sont plutôt les cellules mémoires qui sont auto-alignées alors que la périphérie ne l'est pas forcément, car on désire être plus agressif pour les dimensions de la mémoire et pas forcément pour la périphérie. Il peut donc paraître surprenant de voir une périphérie auto-alignée alors que la partie mémoire ne l'est pas. Une alternative au deuxième procédé aurait donc été possible en réalisant les transistors de périphérie et la mémoire non auto-alignés. Ce procédé est présenté dans le **paragraphe B** de l'**Annexe 1**. Cependant, il nous a semblé préférable de garder la périphérie identique au procédé standard (et donc auto-alignée) afin que les dispositifs de périphérie aient les mêmes performances que dans le procédé standard à grille flottante continue.

II.3.4 Niveaux de masques

Nous avons vu dans le chapitre I que Freescale annonçait que plusieurs niveaux de masques pouvaient être économisés grâce à la technologie Si-NCs. Dans cette partie, nous allons évaluer le nombre de masques que l'intégration des Si-NCs, permet d'économiser par rapport au procédé standard présenté précédemment.

De façon générale, un procédé de fabrication d'une Flash standard compte entre 10 et 13 masques additionnels par rapport au procédé CMOS de la logique [Muralidhar'03]. Considérons dans un premier temps le deuxième procédé Si-NCs qui a été réalisé (oxyde de grille de la périphérie avant les Si-NCs). Les étapes de masquage qui ne sont plus nécessaires sont les suivantes :

- La gravure de la grille flottante (**Fig. II-17**)
- La gravure Polysilicium +ONO en périphérie (**Fig. II-21**)

On économise donc deux masques. Cependant, de nouvelles étapes de masquage sont nécessaires pour le procédé Si-NCs :

- La gravure damascène Polysilicium+oxyde dans le plan mémoire (**Fig. II-39**)
- La gravure de l'empilement mémoire dans la périphérie (**Fig. II-41**)

Finalement, dans le cas du procédé Si-NCs qui a été réalisé on obtient le même nombre de masques.

On peut imaginer améliorer encore l'intégration afin d'économiser des niveaux de masques supplémentaires.

Premièrement, il serait possible de profiter du masque de la gravure damascène polysilicium+oxyde dans le plan mémoire pour réaliser juste après, l'implant caisson de la mémoire. Il faudrait cependant veiller à ce que le bilan thermique vu par l'implant reste le même afin d'avoir la même tension de seuil, ce qui nécessiterait quelques ajustements.

Deuxièmement, on peut imaginer réaliser la gravure des grilles polysilicium des transistors de la logique et des cellules mémoires (**Fig. II-20** et **Fig. II-22**) avec un seul masque. En effet,

dans le procédé standard, il est nécessaire de faire ces gravures en deux étapes distinctes, les épaisseurs de polysilicium étant différentes en périphérie (grille unique) et dans le plan mémoire (grille empilée). Comme le procédé Si-NCs, l'épaisseur de polysilicium est identique en périphérie et en mémoire, une seule étape photo-gravure est suffisante pour définir toutes les grilles.

Troisièmement, il serait possible de simplifier la séquence de croissance des oxydes de grilles des transistors de la périphérie (**Fig. II-12**). L'oxyde de 100 Å réalisé dans le plan mémoire mais aussi dans le design (capacités) n'est plus utile pour la mémoire. En adaptant le design, il serait donc possible de supprimer l'étape de gravure de l'oxyde dans le plan mémoire et dans la partie basse tension (**Fig. II-12(b)**) pour ne faire que la séquence suivante : croissance oxyde haute tension → gravure oxyde dans la partie basse tension → croissance de l'oxyde de grille basse tension.

Finalement, on voit donc qu'il serait possible d'économiser trois masques par rapport au procédé de la Flash standard, avec un design adapté aux Si-NCs et une optimisation du procédé de fabrication. Ceci représente une économie de masque d'environ 25 % sur la dizaine de masques additionnels à la logique qui sont utilisés dans un procédé standard. L'intégration des Si-NCs peut donc permettre une réduction du coût de fabrication par plaque non négligeable.

II.4 Conclusion

Dans la première partie de ce chapitre, nous avons présenté les différentes méthodes de fabrication des nanocristaux de silicium. La technique que nous avons choisie pour la fabrication des nanocristaux de silicium dans nos dispositifs est la croissance par LPCVD (Low Pressure Chemical Vapor Deposition). Cette méthode semble être la plus proche des procédés industriels couramment utilisés pour la fabrication des mémoires Flash conventionnelles. De plus, nous avons montré qu'en utilisant un procédé à deux étapes qui dissocie la nucléation de la croissance des Si-NCs, il est possible de mieux contrôler leur densité et leur taille. Nous avons également vu qu'il était possible de former une fine couche d'oxynitride autour des Si-NCs (grâce à un recuit sous NH_3 ou sous NO) afin de limiter l'oxydation des Si-NCs.

Dans la deuxième partie du chapitre, nous avons présenté l'intégration des Si-NCs dans un produit ATMEL Flash NOR 32 Mb, basé sur une technologie 130 nm. Nous avons détaillé l'organisation d'un produit mémoire Flash, incluant la matrice mémoire (divisée en 64 secteurs de 512 Kb) et les transistors de périphérie CMOS haute et basse tension.

A partir du procédé de fabrication d'une mémoire Flash standard à grille flottante continue, nous avons étudié les différentes façons d'intégrer les nanocristaux de silicium dans le procédé de fabrication à la place de la grille flottante en polysilicium. Un des points clés de l'intégration est de limiter l'oxydation des Si-NCs, les étapes d'oxydation les plus critiques étant la croissance des oxydes de grille des transistors de périphérie. Deux procédés différents ont été réalisés. Dans le premier, les oxydes de grille des transistors sont réalisés après le dépôt des Si-NCs. Cette solution se rapproche le plus du procédé standard et elle est donc la plus simple, mais les Si-NCs risquent d'être oxydés. Un deuxième procédé, plus complexe mais permettant d'éviter toute oxydation parasite des Si-NCs a donc été mis en place. Dans ce cas, les nanocristaux sont déposés après avoir réalisé les oxydes de grille des transistors de la périphérie. Enfin, nous avons mis en évidence que d'autres procédés d'intégration des Si-NCs étaient possibles, indépendamment des oxydations de la périphérie et des Si-NCs, selon que la périphérie et/ou la mémoire soient SASTI (« auto-alignées ») ou non.

Enfin, nous avons démontré que trois niveaux de masques pourraient être économisés, entraînant une réduction non négligeable des coûts de fabrication.

ANNEXE : Autres procédés de fabrication possibles intégrant les nanocristaux de silicium

A. Procédé « périphérie et mémoire auto-alignées/oxydes de périphérie avant les Si-NCs »

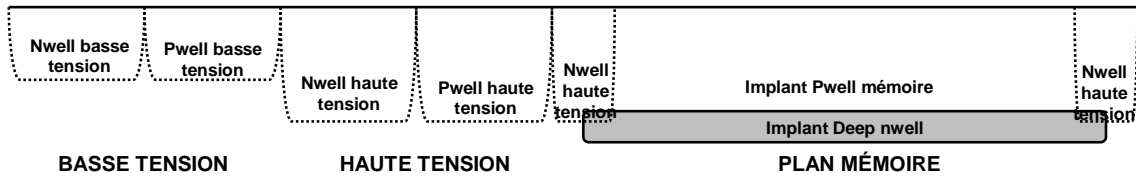


Fig. II-48 : Implants caissons transistors et mémoire.

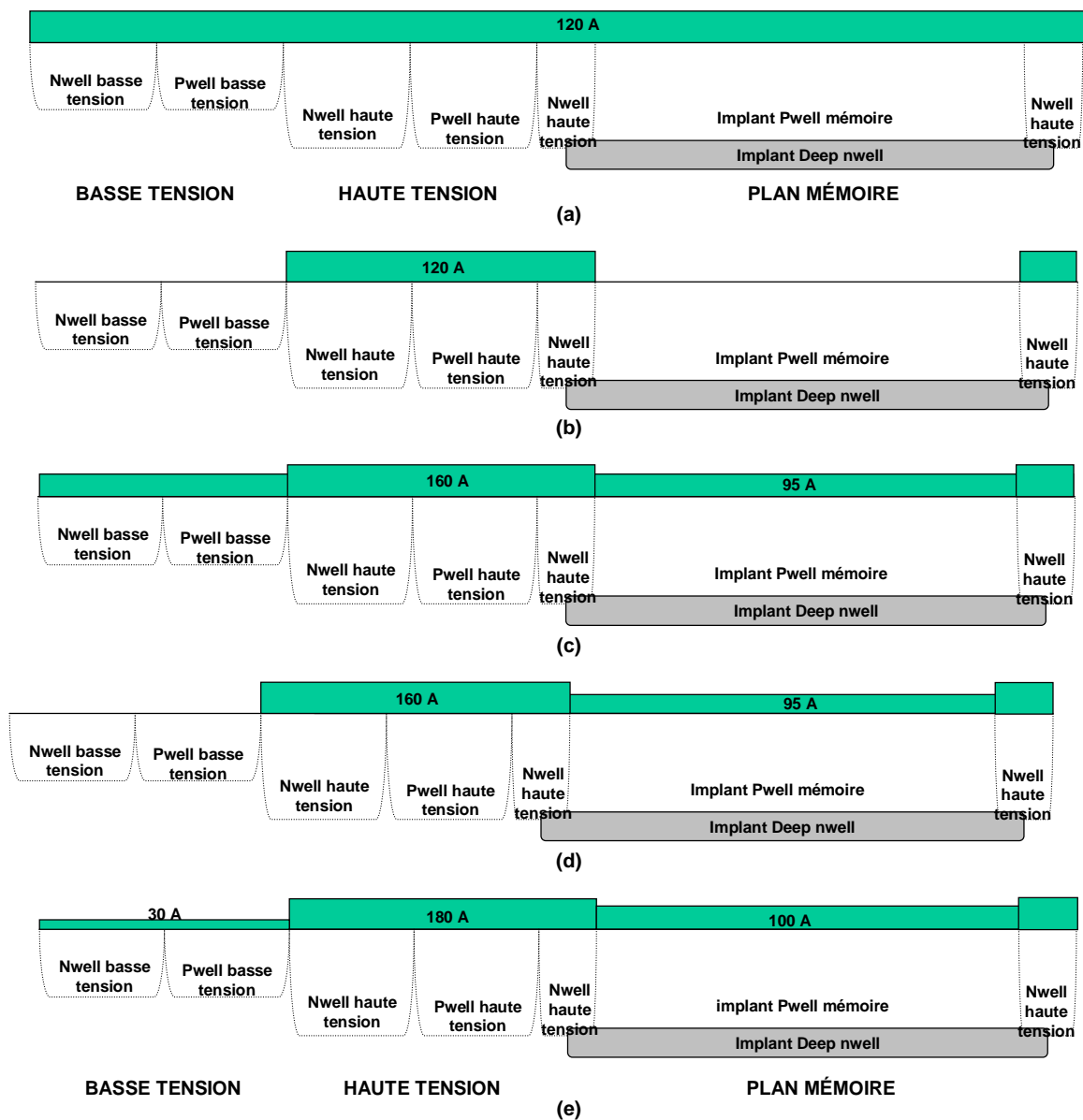


Fig. II-49 : Croissance des oxydes de grille des transistors haute et basse tension. Croissance d'un oxyde de 120 Å (a), gravure dans la partie basse tension et dans le plan mémoire (b), croissance d'un oxyde (160 Å dans la partie haute tension) (c), gravure dans la partie basse tension (d), croissance de l'oxyde de grille basse tension (e).

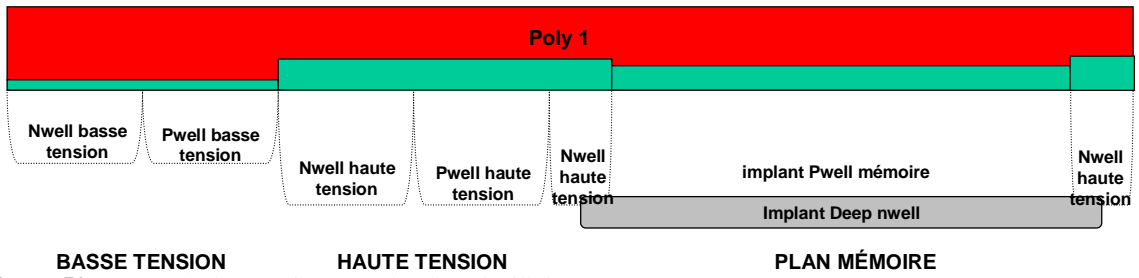


Fig. II-50 : Dépôt de la première couche de polysilicium.

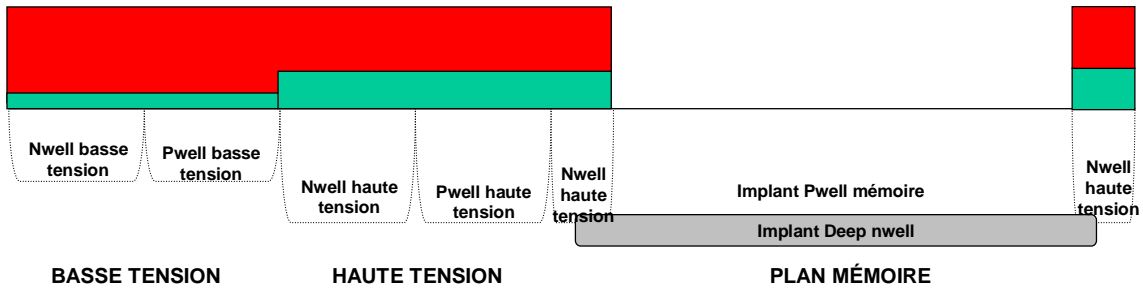


Fig. II-51 : Gravure de l'oxyde et du polysilicium dans la partie mémoire.

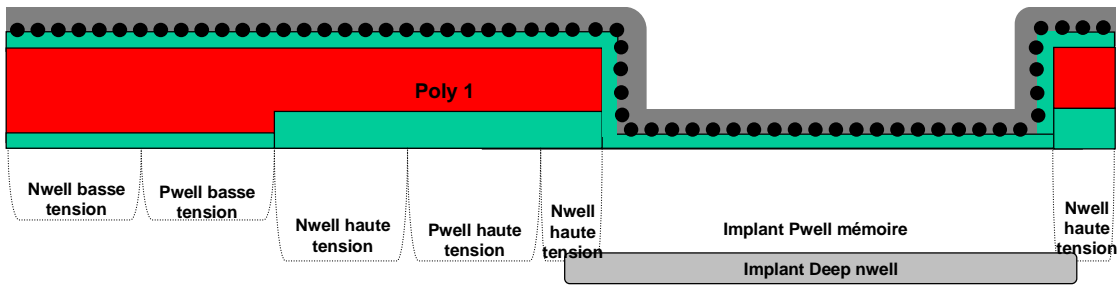


Fig. II-52 : Croissance de l'oxyde tunnel, dépôt des Si-NCs et de l'oxyde de contrôle.

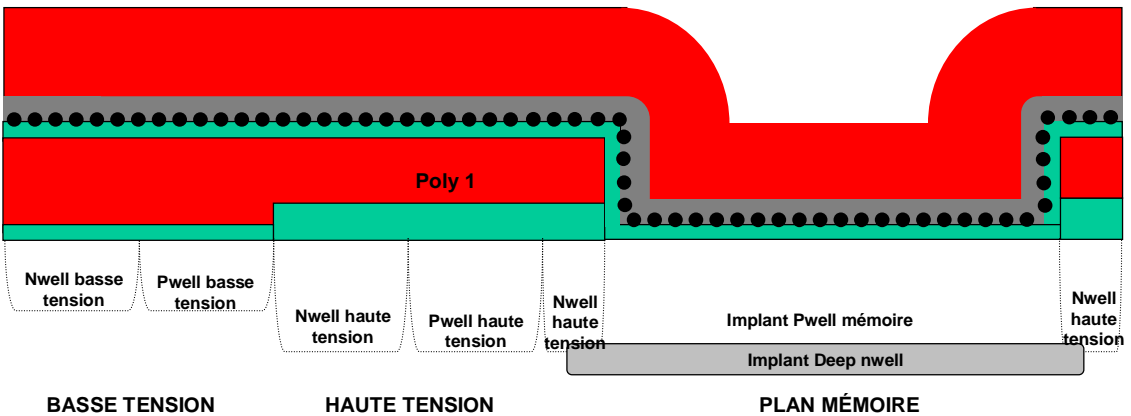


Fig. II-53 : Dépôt de la deuxième couche de polysilicium.

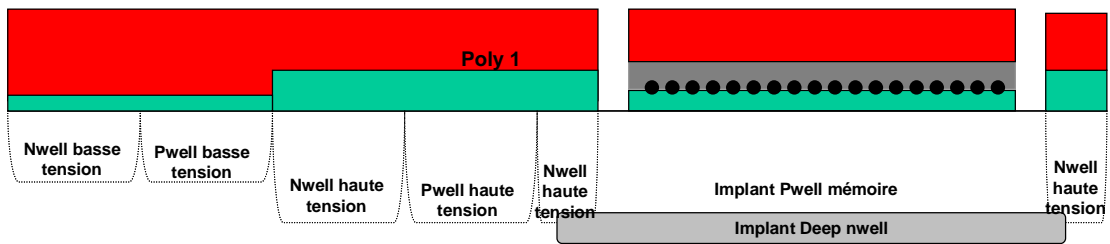


Fig. II-54 : Gravure de l'empilement mémoire (oxyde/Si-NCs/oxyde de contrôle/polySi) dans la périphérie et sur les bords du plan mémoire.

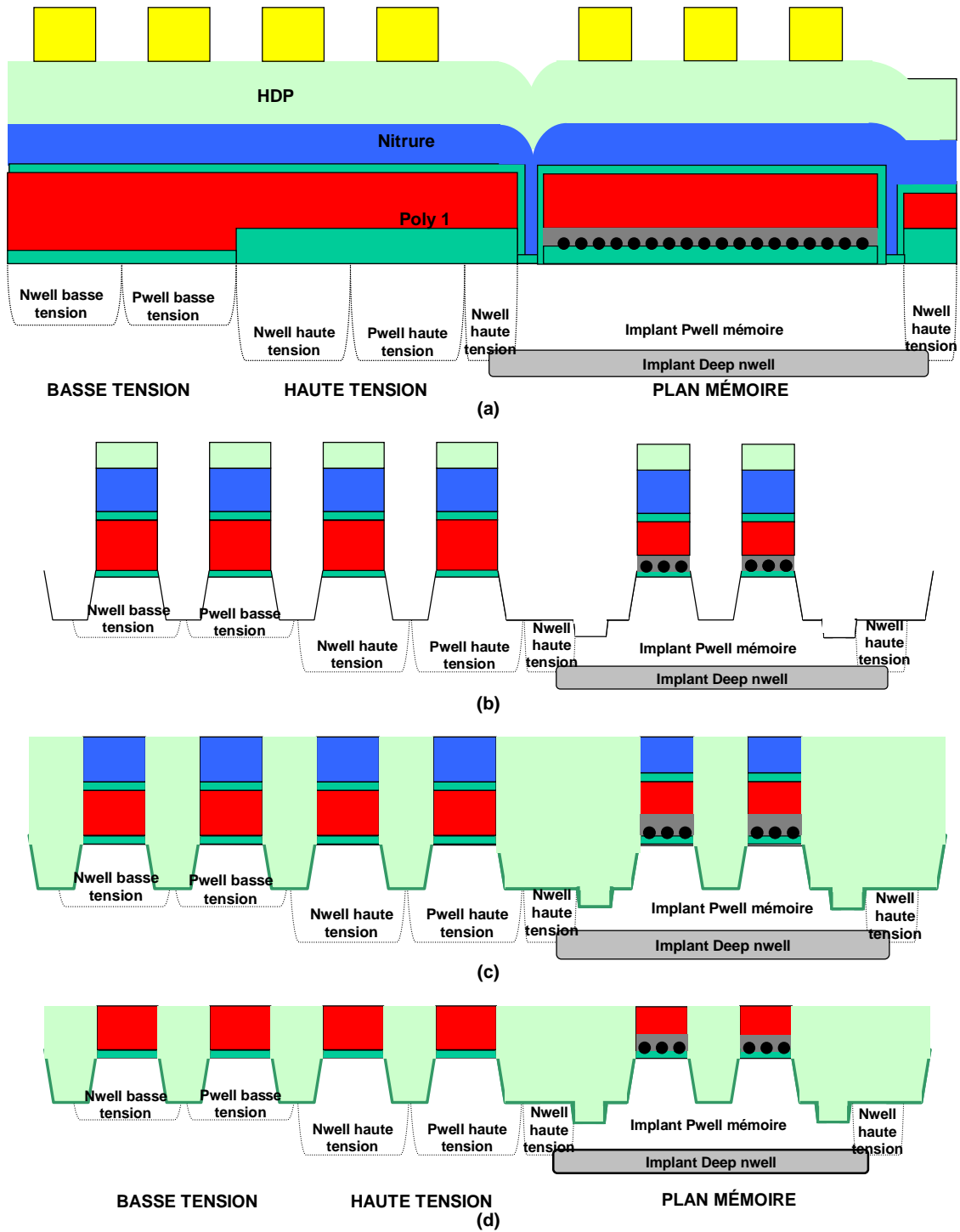


Fig. II-55 : Formation des STI. Croissance d'un oxyde, dépôt d'une couche de nitrure, d'un oxyde HDP et de la résine puis lithographie des zones actives (a). (b) : gravure des zones actives. (c). Oxydation liner, remplissage des STI par un oxyde HDP, densification HDP, puis gravure jusqu'au polysilicium (d).

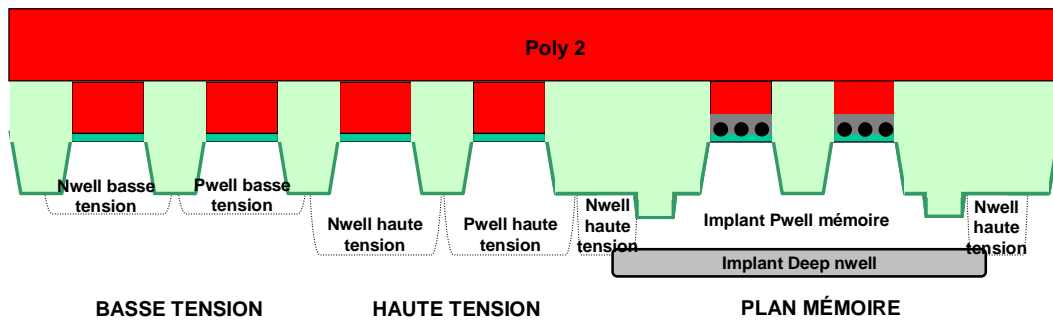


Fig. II-56 : Dépôt de la couche de polysilicium 2.

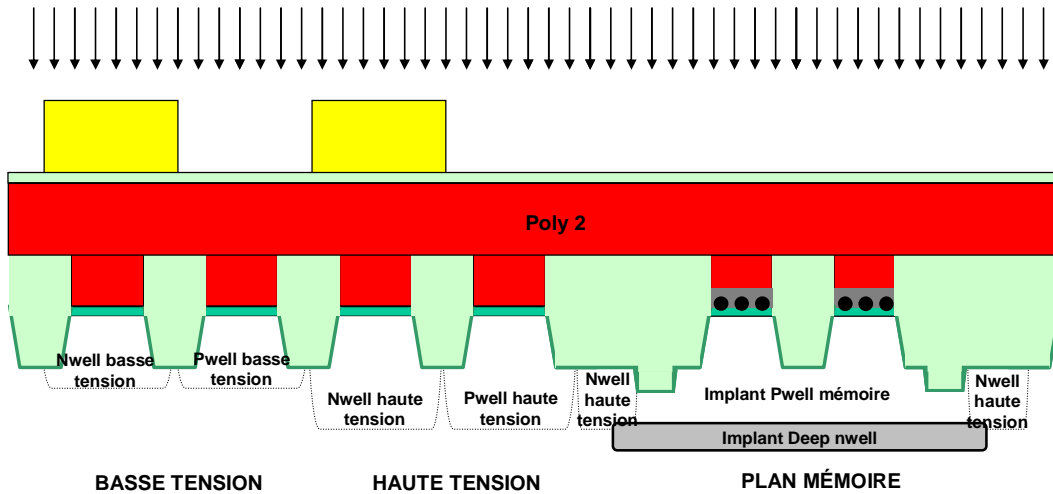


Fig. II-57 : Implantation N+ des grilles des NMOS de périphérie.

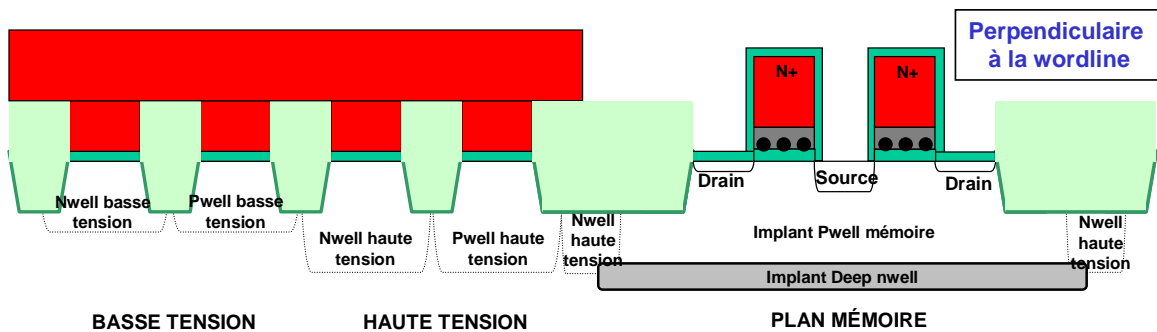


Fig. II-58 : Gravure des cellules mémoires (définition de la longueur des cellules), premier dopage drain, gravure des lignes de source et dépôt d'un oxyde autour des cellules mémoires, dopage source et drain.

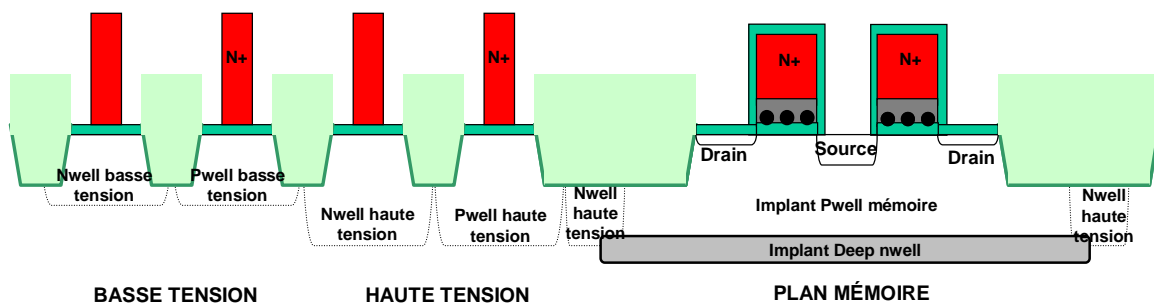


Fig. II-59 : Gravure du polysilicium des transistors de périphérie haute et basse tension (définition de la longueur des transistors).

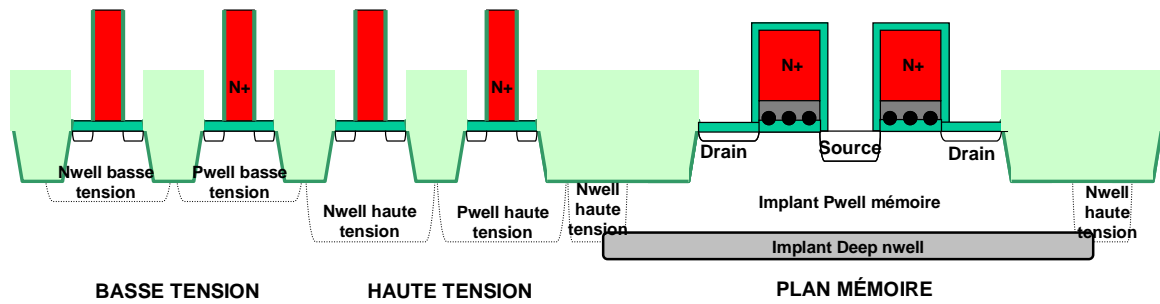


Fig. II-60 : Croissance d'un oxyde autour de la grille de contrôle des transistors et implants LDD des transistors de périphérie.

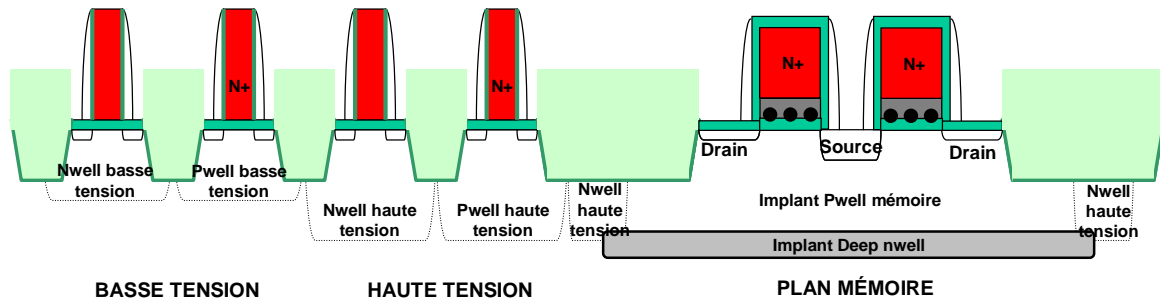


Fig. II-61 : Formation des espaceurs en nitrure.

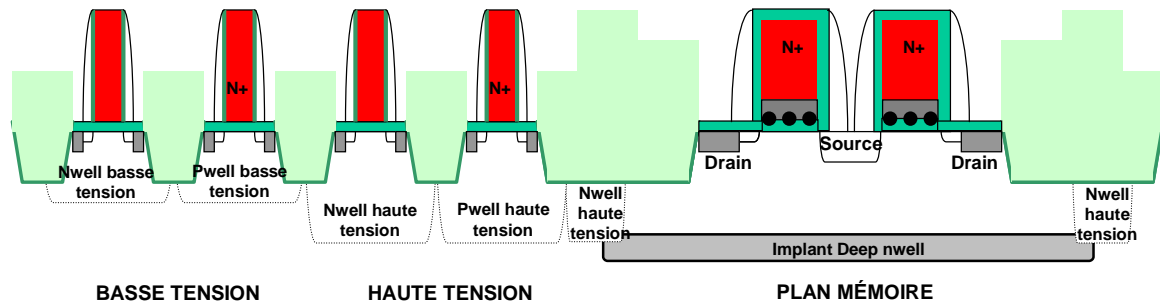


Fig. II-62 : Dopage N+ source/drain des NMOS de la périphérie et drain des cellules mémoires et dopage P+ source/drain des PMOS de la périphérie.

B. Procédé « périphérie et mémoire non auto-alignées/oxydes de périphérie avant les Si-NCs »

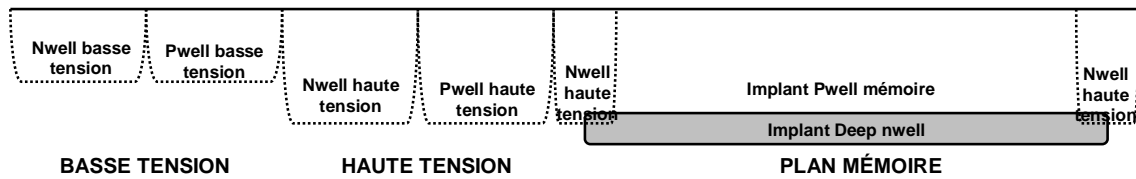


Fig. II-63 : Implants caissons transistors et mémoire.

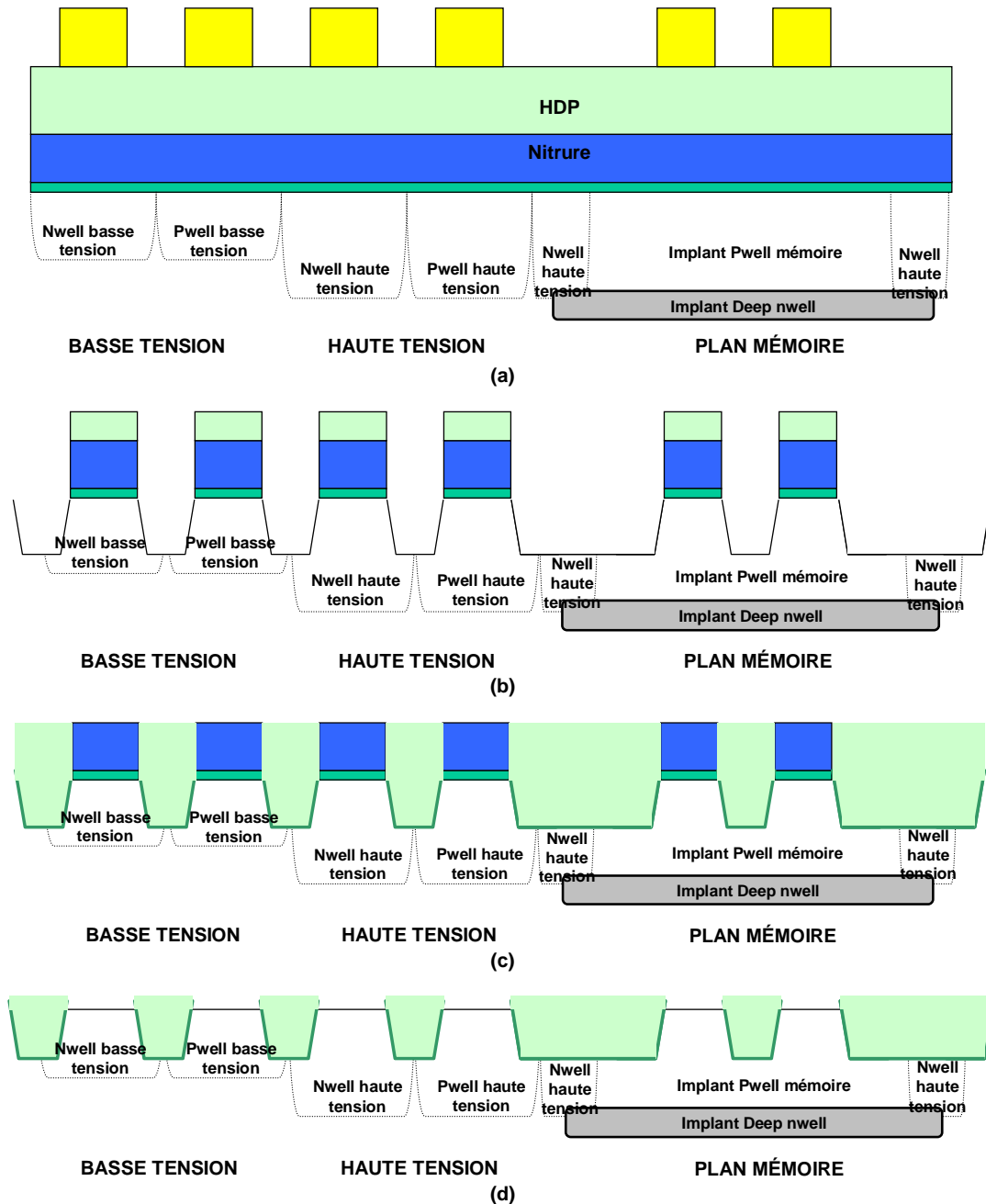


Fig. II-64 : Formation des STI. Croissance d'un oxyde, dépôt d'une couche de nitrure, d'un oxyde HDP et de la résine puis lithographie des zones actives (a). (b) : gravure des zones actives. (c). Oxydation liner, remplissage des STI par un oxyde HDP, densification HDP, puis gravure jusqu'au polysilicium (d).

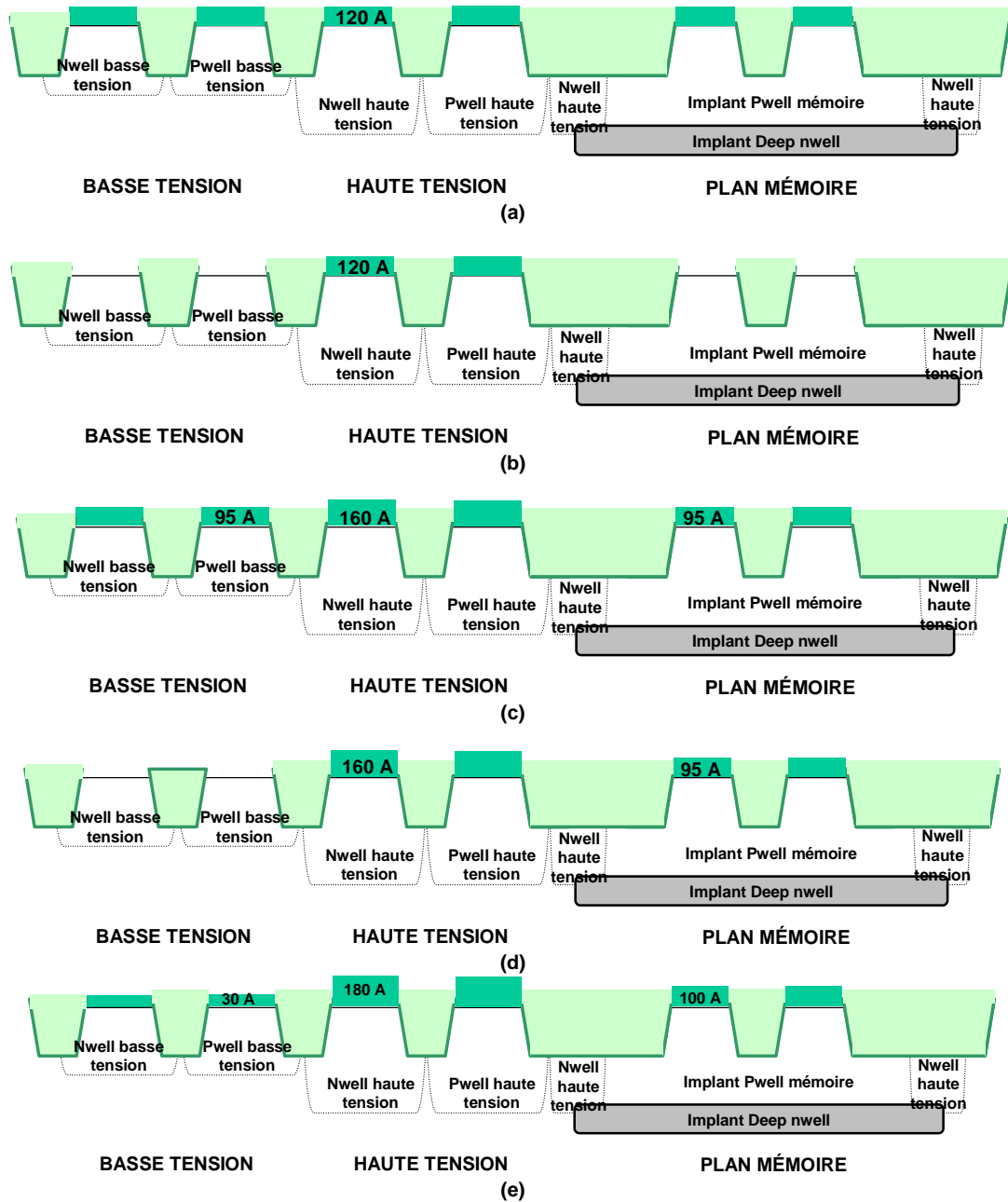


Fig. II-65 : Croissance des oxydes de grille des transistors haute et basse tension. Croissance d'un oxyde de 120 Å (a), gravure dans la partie basse tension et dans le plan mémoire (b), croissance d'un oxyde (160 Å dans la partie haute tension) (c), gravure dans la partie basse tension (d), croissance de l'oxyde de grille basse tension (e).

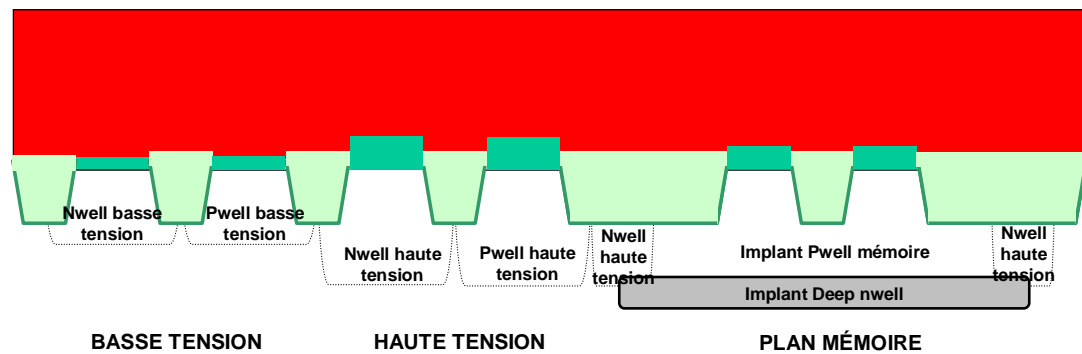


Fig. II-66 : Dépôt d'une couche de polysilicium d'épaisseur équivalente à la somme des épaisseurs Poly 1 et Poly 2.

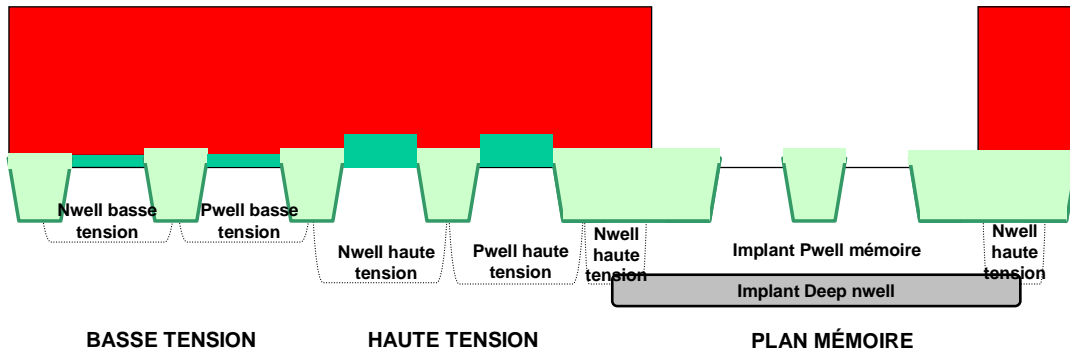


Fig. II-67 : Gravure polysilicium+oxyde dans le plan mémoire.

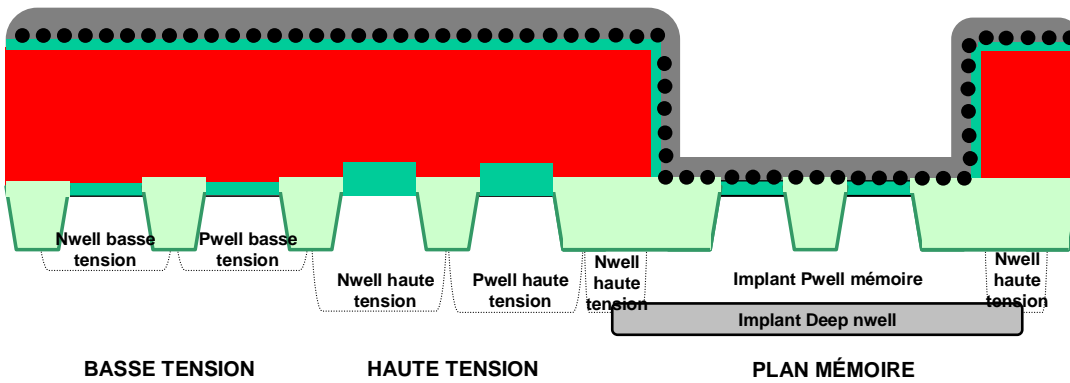


Fig. II-68 : Croissance de l'oxyde tunnel, dépôt des Si-NCs et de l'oxyde de contrôle.

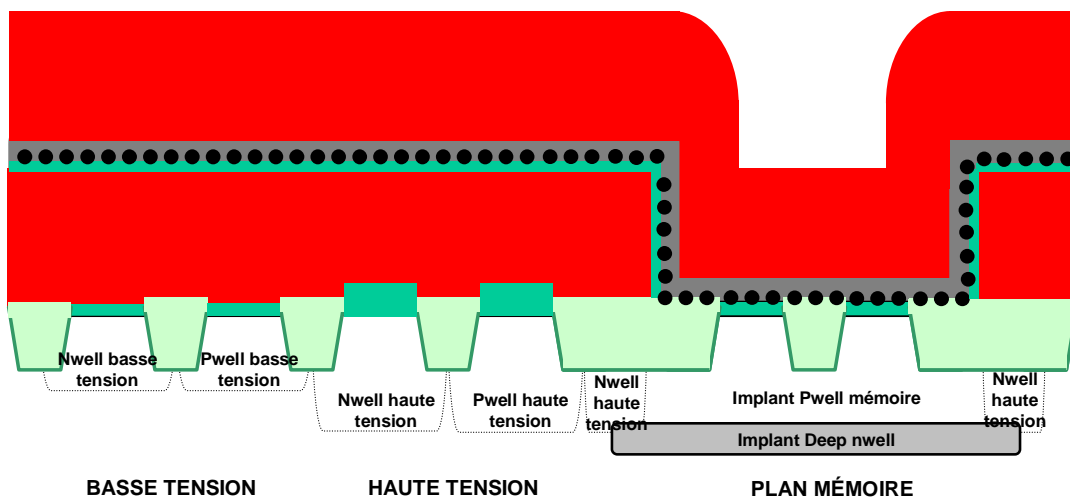


Fig. II-69 : Dépôt d'une couche de polysilicium d'épaisseur équivalente à la somme des épaisseurs Poly 1 et Poly 2.

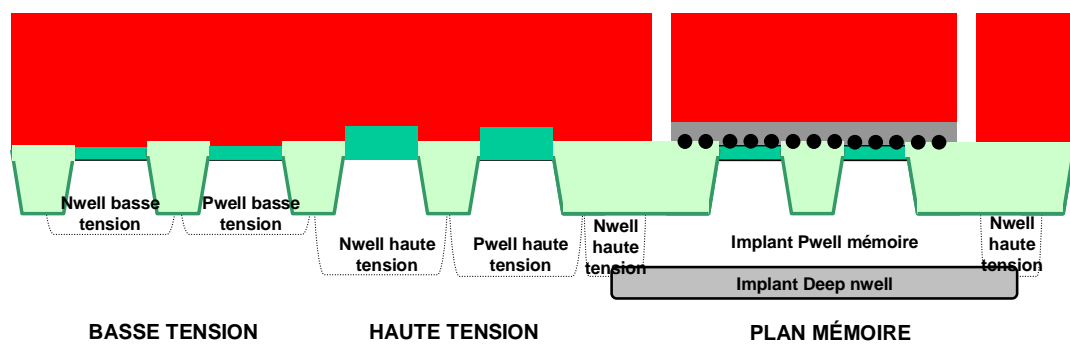


Fig. II-70 : Gravure de l'empilement mémoire (oxyde/Si-NCs/oxyde de contrôle/polySi) dans la périphérie et sur les bords du plan mémoire.

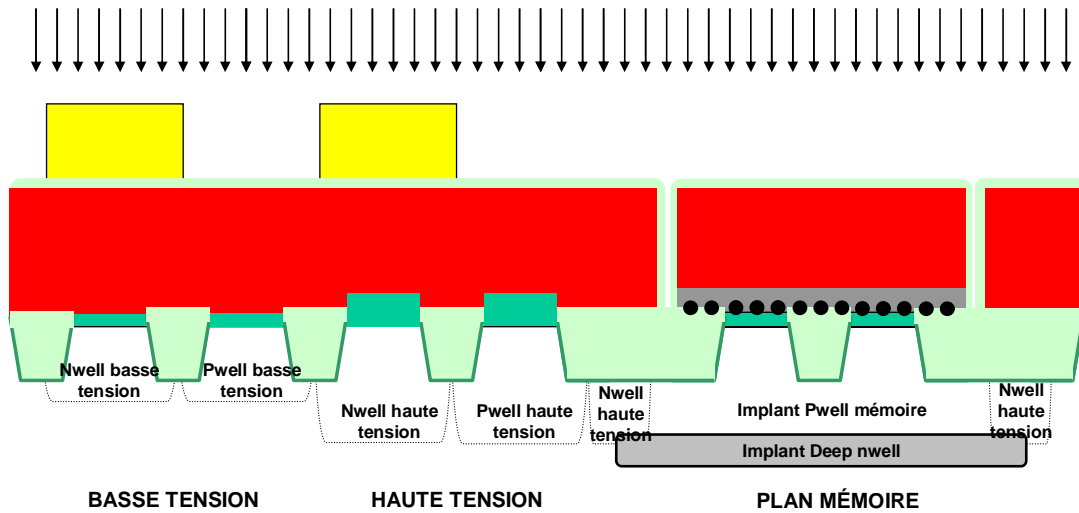


Fig. II-71 : Implantation N+ des grilles des NMOS de périphérie.

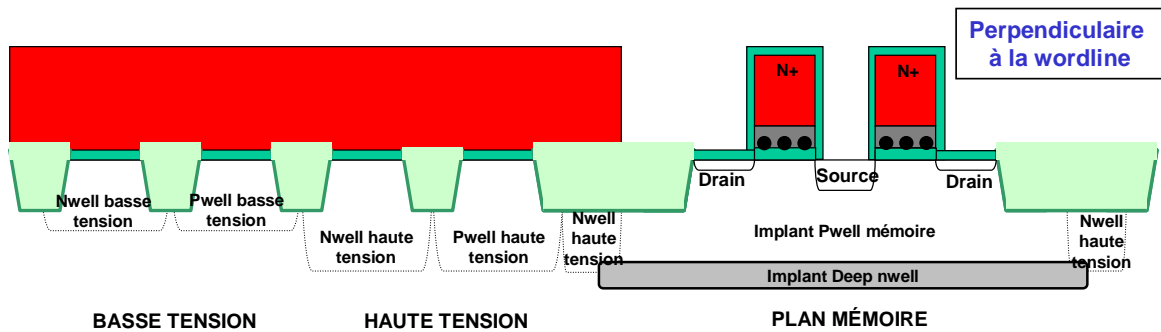


Fig. II-72 : Gravure des cellules mémoires (définition de la longueur des cellules), premier dopage drain, gravure des lignes de source et dépôt d'un oxyde autour des cellules mémoires, dopage source et drain.

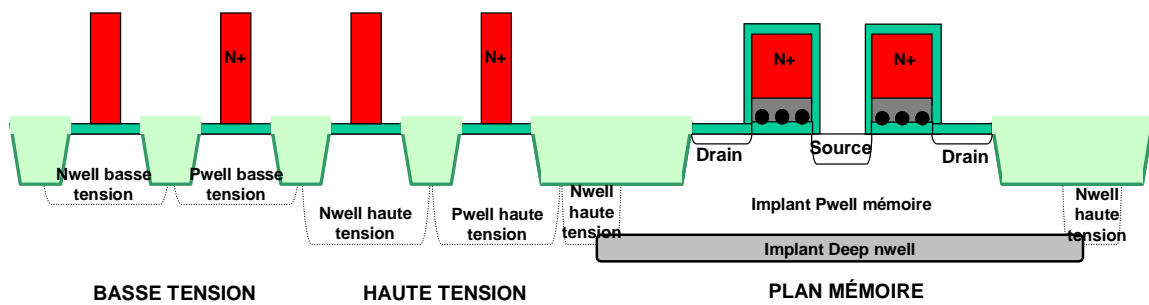


Fig. II-73 : Gravure du polysilicium des transistors de périphérie haute et basse tension (définition de la longueur des transistors).

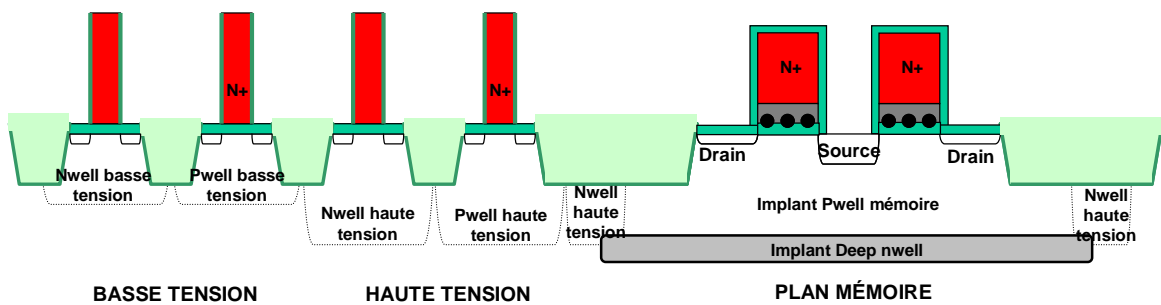


Fig. II-74 : Croissance d'un oxyde autour de la grille de contrôle des transistors et implants LDD des transistors de périphérie.

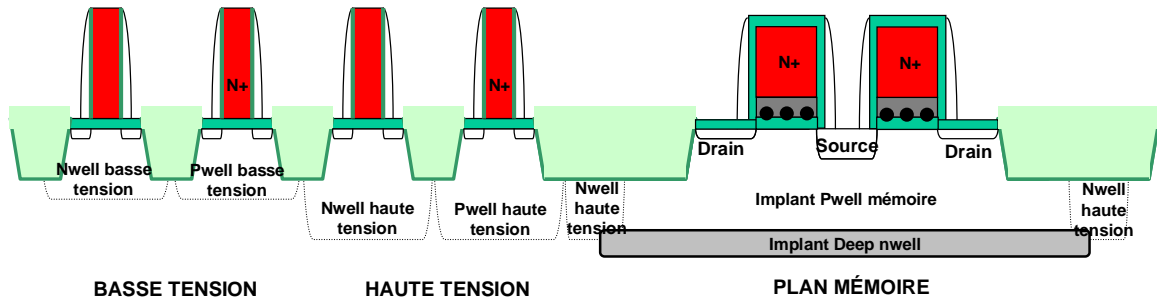


Fig. II-75 : Formation des espaceurs en nitrure.

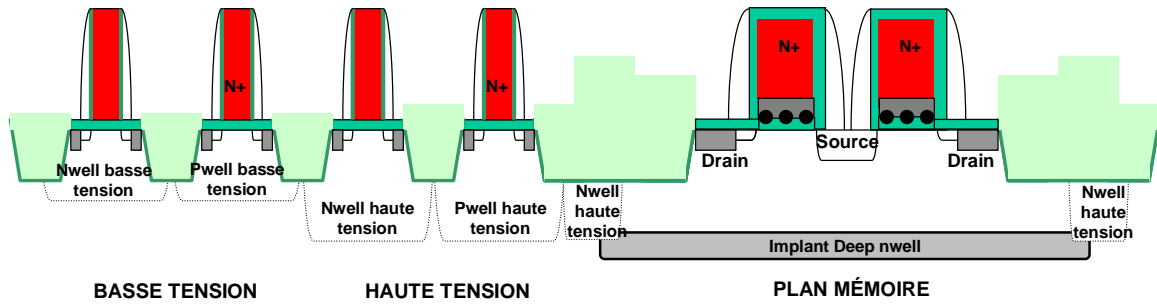


Fig. II-76 : Dopage N+ source/drain des NMOS de la périphérie et drain des cellules mémoires et dopage P+ source/drain des PMOS de la périphérie.

Bibliographie

- [Baron'00] T Baron, F. Martin., P. Mur, C. Wyon, M. Dupuy, "Silicon quantum dot nucleation on Si₃N₄, SiO₂ and SiO_xN_y substrates for nanoelectronic devices", J. Crys. Gro., vol. 209, no 4, pp. 1404-1408, Fev. 2000.
- [Boer'01] E. A. Boer., "Synthesis, passivation and charging of silicon nanocrystals", PhD thesis, California Institute of Technology, 2001.
- [Buffet'02] N. Buffet, P. Mur, B. De Salvo, M. N. Séméria, "Silicon nanocrystals precipitation in a SiO/sub 2/ matrix elaborated from the decomposition of LPCVD SiO/sub x/", IEEE-NANO 2002, Proceedings of the 2002 2nd IEEE Conference on nanotechnology, p.269.
- [DeBlauwe'00] J. De Blauwe, M. Ostraat, M.L. Green, G. Weber, T. Sorch, A. Kerber, F. Klemens, R. Cireli, E. Ferry, J.L. Grazul, F. Braumann, Y. Kim, W. Mansfield, J. Bude, J.T.C. Lee, S.J. Hillenius, R.C. Flagan, H.A. Atwater, "A novel aerosol-nanocrystal floating gate device for non-volatile memory applications", IEDM Tech. Dig., pp. 115-118, 2000.
- [Dufourcq'07] J. Dufourcq, O. Renault, C. Crotti, N. Barrett, S. Bodnar, G. Festes, P. Mur, T. Baron, R. Coppard, "High-Resolution X-ray Photoelectron Spectroscopy Investigations on the Core-shell Interface of Nitrided Si Nanocrystals", MRS proceedings, avril 2007.
- [Kamath'97] A. Kamath, D. L. Kwong, Y. M. Sun, P. M. Blass, S. Whaley, J. M. White, "Oxidation of Si(100) in nitric oxide at low pressures: An x-ray photoelectron spectroscopy study", Appl. Phys. Lett., vol. 70, p. 63, 1997.
- [Lopez'02] M. López, B. Garrido, C. García, P. Pellegrino, A. Pérez-Rodríguez, J. R. Morante, C. Bonafos, M. Carrada, A. Claverie, « Elucidation of the surface passivation role on the photoluminescence emission yield of silicon nanocrystals embedded in SiO₂ », App. Phys. Lett., vol. 80, p. 1637, 2002.
- [Mazen'03a] F. Mazen, T. Baron, G. Brémond, N. Buffet, N. Rochat, P. Mur, M.N. Séméria, "Influence of the chemical properties of the substrate on silicon quantum dot nucleation", J. Electrochem. Society, vol. 150, no 3, pp. 203-208, 2003.
- [Mazen'03b] F. Mazen, T. Baron, A.M. Papon, R. Truche, J.M. Hartmann, "A two steps CVD process for the growth of silicon nano-crystal", Appl. Surf. Sci., vol. 214, pp. 359-363, 2003.
- [Mazen'03these] F. Mazen, "Etude de la nucléation et de la croissance de nanocristaux de silicium élaborés par dépôt chimique en phase vapeur pour dispositifs nanoélectroniques", Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, soutenue le 7 octobre 2003.
- [Muralidhar'03] R. Muralidhar, R.F. Steimle, M. Sadd, R.Rao, C.T. Swift, E.J. Prinz, J. Yater, L. Grieve, K Hradsky, S. Straub, B. Acred, W. Paulson, W. Chen, L. Parker, S.G.H. Anderson, M. Rossow, M. Paransky, T. Huynh, D. Hadad, KO-Min Chang, B.E. White Jr., "A 6V Embedded 90nm Silicon Nanocrystal Nonvolatile Memory", IEEE IEDM Tech. Dig., p. 26.2.1, 2003.

- [Nicotra'03]** G. Nicotra, S. Lombardo, C. Spinella, G. Ammendola, C. Gerardi, C. Demuro, « Observation of the nucleation kinetics of Si quantum dots on SiO₂ by energy filtered transmission electron microscopy », Applied Surface Science, vol. 205, p.304, 2003.
- [Ostraat'01]** M. Ostraat, J. De Blauwe, "Future Silicon Nanocrystal Nonvolatile Memory Technology" Materials Research Society Proceedings, vol. 686, 2001.
- [Scheer'03]** K. C. Scheer, R.A. Rao, R. Muralidhar, S. Bagchi, J. Conner, L. Lozano, C. Perez, M. Sadd, B.E. White Jr, "Thermal oxidation of silicon nanocrystals in O₂ and NO ambient", J. App. Phys., vol. 93, no. 9, p.5637.
- [Yun'00]** F.Yun, B.J. Hinds, S. Hatatani, S.Oda, Q.X. Zhao, M. Willander., "Study of structural and optical properties of nanocrystalline silicon embedded in SiO₂, vol. 375, p.137, 2000.

Chapitre III : Caractérisation électrique des mémoires Flash à nanocristaux de silicium

Chapitre III : Caractérisation électrique des mémoires Flash à nanocristaux de silicium.....	109
III.1 Introduction.....	113
III.2 Caractérisation électrique des cellules mémoires simples à nanocristaux de silicium.....	113
III.2.1 Influence des méthodes de programmation.....	113
III.2.1.1 Ecriture par porteurs chauds.....	113
III.2.1.2 Effacement par Fowler-Nordheim	114
III.2.2 Influence des paramètres technologiques	115
III.2.2.1 Influence de l'implant canal.....	117
III.2.2.2 Influence de l'épaisseur de l'oxyde tunnel.....	117
III.2.2.3 Influence de la taille et de la densité des nanocristaux	120
III.2.2.3.A Influence du procédé de fabrication	120
III.2.2.3.B Influence de la taille des nanocristaux	120
III.2.2.4 Influence de la passivation des nanocristaux	124
III.2.2.4.A Passivation avec NH ₃ et NO	124
III.2.2.4.B Dépôt d'une couche de Si ₃ N ₄	126
III.2.2.5 Influence du diélectrique de contrôle.....	127
III.2.2.5.A Influence de la nature du diélectrique : HTO ou ONO	127
III.2.2.5.B Influence de l'épaisseur, densification du HTO.....	129
III.3 Caractérisation électrique des matrices mémoires à nanocristaux de silicium	135
III.3.1 Introduction	135
III.3.2 Influence du procédé de fabrication des nanocristaux de silicium.....	136
III.3.2.1 Influence de la taille des nanocristaux	136
III.3.2.2 Comparaison des procédés une étape/deux étapes	137
III.3.3 Influence des conditions de programmation	137
III.3.3.1 Amélioration de la distribution des tensions écrites.....	138
III.3.3.1.A Influence du temps d'écriture.....	138
III.3.3.1.B Influence de la polarisation du substrat.....	139
III.3.3.2 Amélioration de la distribution des tensions effacées.....	140
III.4 Fiabilité.....	141
III.4.1 Endurance	141
III.4.2 Rétention	143
III.4.3 Rétention après endurance	144
III.4.4 « Gate disturb »	146
III.5 Conclusion.....	148
Bibliographie.....	150

III.1 Introduction

L'objectif de ce troisième chapitre est de présenter les résultats obtenus lors de la caractérisation électrique des mémoires à nanocristaux de silicium présentés dans le chapitre II. Dans un premier temps, nous étudierons les cellules mémoires simples en vue de l'optimisation des paramètres technologiques et électriques de la cellule. Ensuite, nous présenterons les résultats électriques obtenus sur le démonstrateur ATMEL Flash NOR 32 Mb à nanocristaux. L'objectif est de compléter l'étude réalisée sur les cellules unitaires par des tests électriques statistiques sur des matrices mémoires.

III.2 Caractérisation électrique des cellules mémoires simples à nanocristaux de silicium

Nous allons d'abord étudier l'influence des conditions de programmations sur les performances électriques de la cellule mémoire. Ensuite, les résultats électriques seront présentés en fonction des variantes technologiques des dispositifs réalisées sur les différents lots.

III.2.1 Influence des méthodes de programmation

III.2.1.1 Ecriture par porteurs chauds

Tel qu'il est montré dans la **Fig.III-1**, les paramètres électriques intervenant dans l'écriture par porteurs chauds sont les suivants : tensions appliquées sur la grille de contrôle, sur le drain et le substrat ainsi que le temps de programmation.

Tout d'abord, on observe une saturation de la tension de seuil lorsque le temps d'écriture augmente. De plus, on peut voir sur la **Fig. III-1(a)** que plus la tension de grille est élevée, plus l'écriture est rapide et plus le niveau de saturation est élevé. Les tensions de stress appliquées au substrat (**Fig. III-1(b)**) et au drain (**Fig. III-1(c)**) ont une influence sur la vitesse d'écriture, mais pas sur le niveau de saturation. On peut voir qu'il existe un optimum pour la tension de drain (sur la **Fig. III-1(d)**, par exemple, pour des tensions de drain supérieures à 4/4,5V, la tension de seuil diminue car l'injection devient moins efficace).

L'influence des tensions de grille et de substrat et du temps de programmation sera expliquée par la suite à partir des simulations TCAD présentées dans le Chapitre IV.

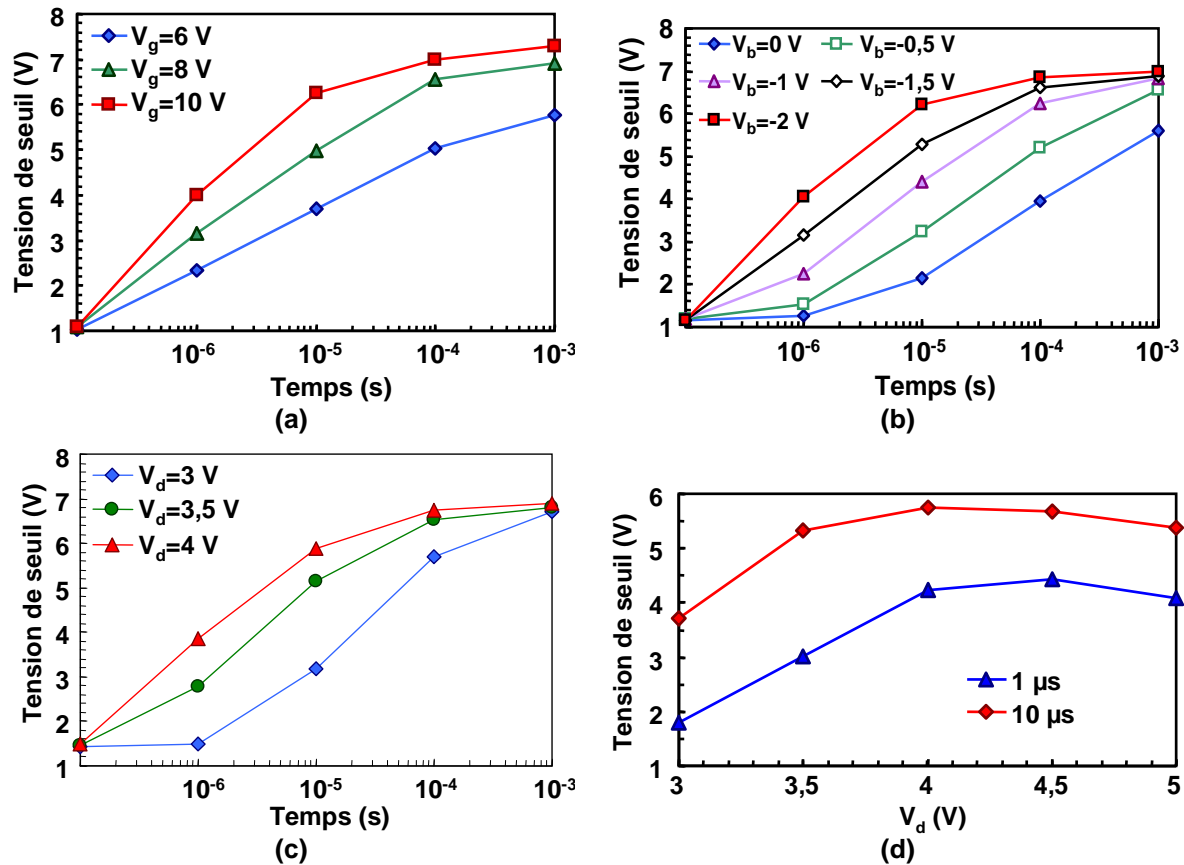


Fig. III-1 : Caractéristiques d'écriture en porteurs chauds en fonction du temps cumulé de stress pour différentes tensions de grille de contrôle (a) ($V_d=3,75\text{ V}$, $V_b=-1,5\text{ V}$, $V_s=0\text{ V}$), de substrat (b) ($V_g=8\text{ V}$, $V_d=3,75\text{ V}$, $V_s=0\text{ V}$) et de drain (c) ($V_g=8\text{ V}$, $V_b=-1,5\text{ V}$, $V_s=0\text{ V}$). Evolution de la tension de seuil en fonction de la tension de stress appliqué sur le drain pour deux temps différents (d) ($V_g=8\text{ V}$, $V_b=-2\text{ V}$, $V_s=0\text{ V}$). Dans (a), (b) et (c), la lecture est effectuée à $V_d=1\text{ V}$ et $I_d=0,3\text{ nA}$ ($t_{\text{tun}}=5\text{ nm}$, $t_{\text{HTO}}=10\text{ nm}$, $\Phi_{\text{dot}}=6,5\text{ nm}$, $R_{\text{dot}}\approx 0,25$). Dans (d) la lecture est effectuée à $V_d=1\text{ V}$ et $I_d=1\ \mu\text{A}$ ($t_{\text{tun}}=5\text{ nm}$, $t_{\text{HTO}}=8\text{ nm}$, $\Phi_{\text{dot}}=6,5\text{ nm}$, $R_{\text{dot}}\approx 0,2$).

III.2.1.2 Effacement par Fowler-Nordheim

La **Fig. III-2(a)** montre les dynamiques d'effacement en Fowler-Nordheim. La vitesse d'effacement augmente lorsque $|V_g|$ augmente. Pendant l'effacement, la tension de seuil sature au niveau initial (i.e. vierge) des dispositifs, ce qui est une différence par rapport au comportement de la grille flottante continue avec laquelle on peut observer un sur-effacement. De plus, l'effacement est beaucoup plus rapide que pour une grille flottante continue (**Fig. III-3(a)-(b)**). Par exemple, la tension de seuil diminue d'environ 3 V en 1 ms ($V_g=-16\text{ V}$). On voit sur la **Fig. III-3(a)** qu'à 1 ms, le niveau initial est atteint pour la cellule à Si-NCs alors qu'il reste encore plus 1,5 V de fenêtre de programmation pour le dispositif à grille flottante continue.

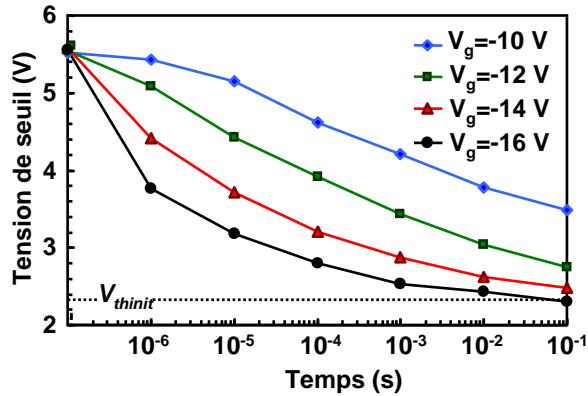


Fig. III-2 : Caractéristiques d'effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour différentes tensions de grille ($V_d=V_s=V_b=0$ V). Lecture : $V_d=1$ V et $I_d=0,3$ nA. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=8$ nm, $\Phi_{\text{dot}}=5$ nm, $R_{\text{dot}}\approx 0,15$.

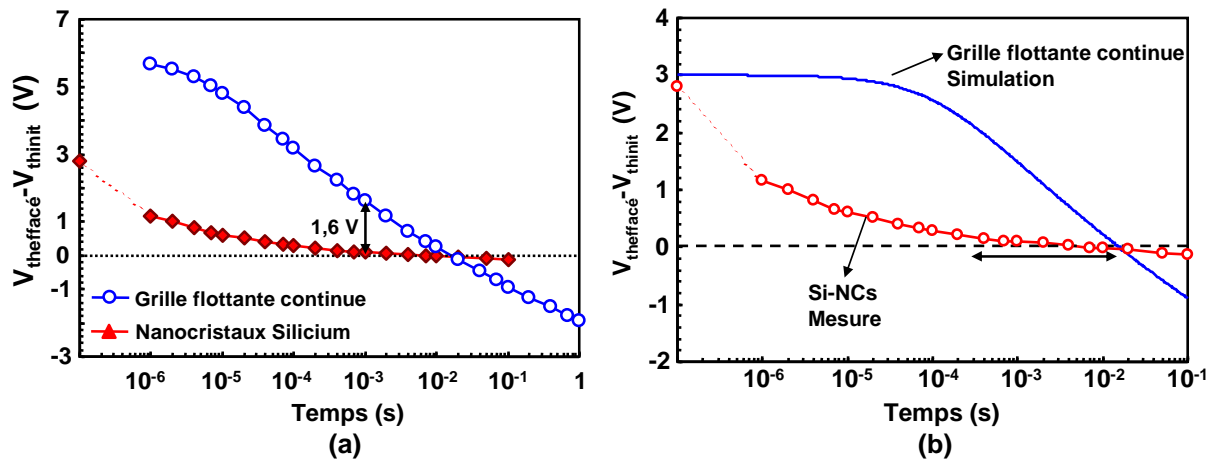


Fig. III-3 : (a) : Comparaison des caractéristiques d'effacement mesurées en Fowler-Nordheim ($V_g=-16$ V) en fonction du temps cumulé de stress d'une cellule à grille flottante continue et d'une cellule à Si-NCs. Lecture : $V_d=1$ V et $I_d=1$ μ A. (b) : Comparaison des dynamiques d'effacement ($V_g=-16$ V) d'une cellule à Si-NCs (mesurée) et d'une grille flottante continue (simulée avec le modèle de la quasi-grille flottante [DeSalvo'01]) à fenêtre de programmation identique. Grille flottante continue : $t_{\text{tun}}\approx 10$ nm, $t_{\text{ONO}}\approx 13$ nm. Si-NCs : $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=5$ nm, $R_{\text{dot}}\approx 0,15$.

III.2.2 Influence des paramètres technologiques

Dans cette partie, nous allons étudier l'influence des différents paramètres technologiques de la cellule mémoire sur ses caractéristiques électriques. Les paramètres technologiques et leurs variantes sont représentés dans la **Fig. III-4** et **Tableau III-1**.

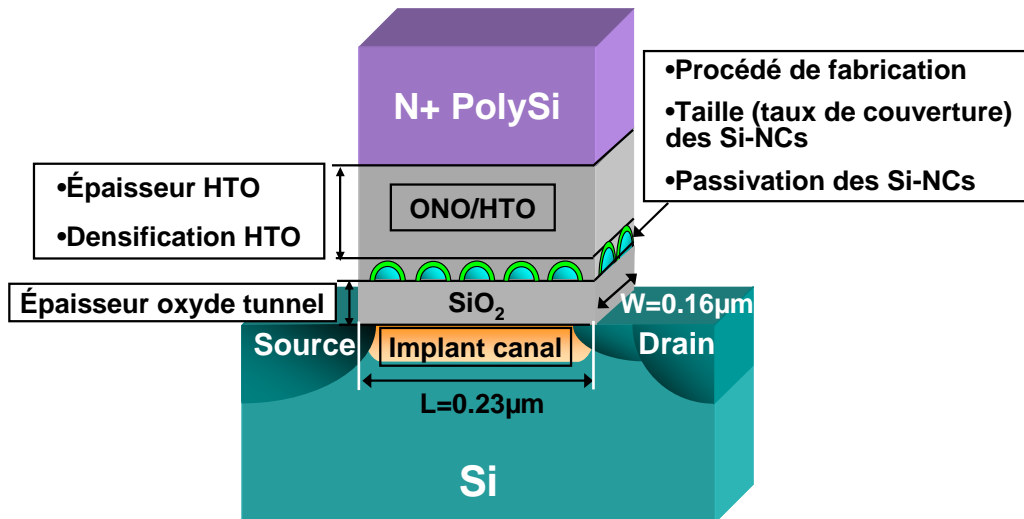


Fig. III-4 : Schéma de la cellule mémoire à nanocristaux de silicium montrant les paramètres technologiques que l'on a fait varier lors de cette étude.

PARAMÈTRES TECHNOLOGIQUES	
Implant canal (dose)	Bore $3.10^{13}/\text{cm}^2$, $6.10^{13}/\text{cm}^2$, $9.10^{13}/\text{cm}^2$
Épaisseur oxyde tunnel t_{tun}	$t_{\text{tun}} = 4/5/6$ nm
Procédé de fabrication des Si-NCs	<ul style="list-style-type: none"> •LPCVD 1 étape (nucléation et croissance dans la même étape) •LPCVD 2 étapes (nucléation et croissance dissociées)
Taille (taux de couverture R_{dot}) des Si-NCs	Étape de nucléation identique Diamètre $\Phi_{\text{dot}} = 4/6,5/8/12$ nm Densité = $7,5.10^{11}/5,5.10^{11}/5.10^{11}/2,7.10^{11}$ Taux de couverture $R_{\text{dot}} = 0,11/0,25/0,34/0,43$
Passivation des Si-NCs	<ul style="list-style-type: none"> •Pas de passivation •Passivation par dépôt d'une couche de nitrure •Passivation sous NO •Passivation sous NH_3
Type du diélectrique de contrôle	HTO/ONO
Densification du HTO	<ul style="list-style-type: none"> •Pas de densification •Densification sous N_2 à 850°C, 950°C et 1000°C •Densification sous O_2 à 850°C
Épaisseur du HTO t_{HTO}	$t_{\text{HTO}} = 8/10/12/14$ nm

Tableau III-1 : Variantes technologiques de cette étude.

III.2.2.1 Influence de l'implant canal

On peut voir dans les **Fig. III-5(a)** et **(b)** que la dose de l'implant canal a une forte influence sur la dynamique d'écriture par porteurs chauds. En particulier, plus la dose est élevée, plus l'écriture est rapide. La fenêtre de programmation peut être jusqu'à 2,5 V supérieure pour un temps de programmation identique avec une dose 50 % plus élevée. Il a été démontré que l'augmentation de la dose de l'implant canal permettait d'augmenter l'efficacité d'injection [Esseni'00]. Ceci peut s'expliquer par le fait qu'en augmentant la dose de l'implant canal, on rend la jonction canal/drain plus abrupte et le champ électrique latéral plus fort, facilitant ainsi la génération d'électrons chauds.

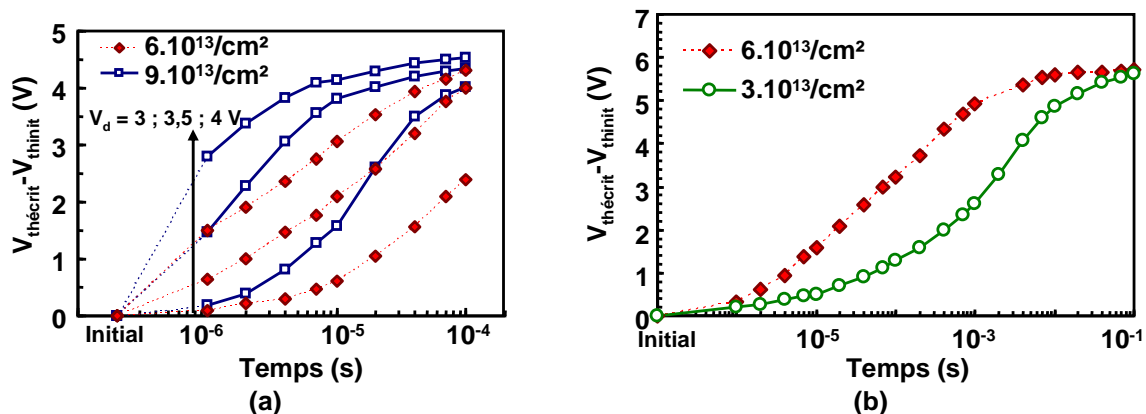


Fig. III-5 : Caractéristiques d'écriture en porteurs chauds en fonction du temps cumulé de stress pour différentes doses de Bore lors de l'implant canal : (a) $6 \cdot 10^{13}/\text{cm}^2$ et $9 \cdot 10^{13}/\text{cm}^2$ ($V_g=8$ V, $V_b=-1$ V, $V_d=3$ V ; 3,5V; 4V, $V_s=0$ V) et (b) $6 \cdot 10^{13}/\text{cm}^2$ et $3 \cdot 10^{13}/\text{cm}^2$ ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V). Lecture : $V_d=1$ V et $I_d=1$ μ A. (a) : $t_{tun}=5$ nm, $t_{HTO}=8$ nm, $\Phi_{dot}=6,5$ nm, $R_{dot}\approx 0,2$. (b) : $t_{tun}=5$ nm, $t_{HTO}=10$ nm, $\Phi_{dot}=6,5$ nm, $R_{dot}\approx 0,25$.

III.2.2.2 Influence de l'épaisseur de l'oxyde tunnel

Les **Fig. III-6(a)-(b)** montrent une légère influence de l'épaisseur de l'oxyde tunnel, lorsque les dispositifs sont écrits par porteurs chauds. L'écriture est plus rapide lorsque l'épaisseur de l'oxyde tunnel diminue. A temps d'écriture identique, on constate une augmentation de la tension de seuil comprise entre 100 et 600 mV (selon les conditions de programmation) pour un oxyde de 4 nm par rapport à un oxyde de 5 nm.

Concernant l'effacement réalisé par Fowler-Nordheim, on observe également que l'effacement est plus rapide pour un oxyde plus fin (**Fig. III-6(c)-(d)**). On peut donc conclure de manière générale que la fenêtre de programmation augmente lorsque l'épaisseur de l'oxyde tunnel diminue. Il faut noter que les épaisseurs de l'oxyde tunnel données sont les épaisseurs nominales visées, avec une incertitude de $\pm 0,3$ nm qui peut expliquer les différences entre lots.

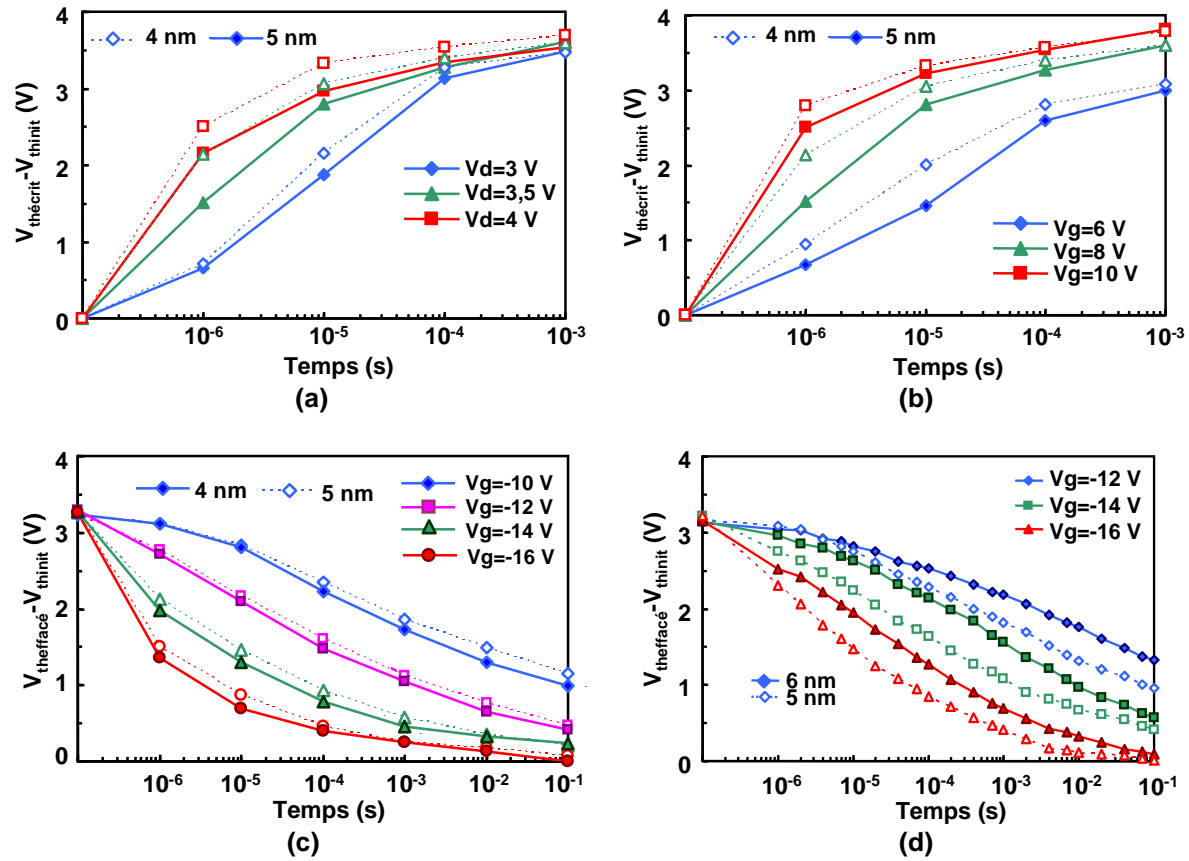


Fig. III-6 : (a), (b) : Caractéristiques d'écriture en porteurs chauds en fonction du temps cumulé de stress pour deux épaisseurs nominales de l'oxyde tunnel : 4 et 5 nm ((a) : $V_g=8$ V, $V_b=-1.5$ V, $V_d=3$ V ; 3,5 V ; 4 V , $V_s=0$ V ; (b) : $V_g=6, 8, 10$ V, $V_b=-1.5$ V, $V_d=3,5$, $V_s=0$ V). (c) : Caractéristiques d'effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour deux épaisseurs nominales de l'oxyde tunnel : 4 et 5 nm et différentes tensions de grille d'effacement. Lecture : $V_d=1$ V et $I_d=0,3$ nA.

(d) : Caractéristiques d'effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour deux épaisseurs nominales de l'oxyde tunnel : 5 et 6 nm et différentes tensions de grille d'effacement. Lecture : $V_d=1$ V et $I_d=1$ μ A. (a), (b) et (c) : $t_{\text{tun}} = 5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=5$ nm, $R_{\text{dot}}\approx 0,15$. (d) : $t_{\text{tun}} = 5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=6,5$ nm, $R_{\text{dot}}\approx 0,25$.

Nous avons également réalisé des mesures d'endurance sur des dispositifs avec différentes épaisseurs d'oxyde tunnel pour des conditions d'écriture/effacement identiques. La **Fig. III-7** montre que la dérive de la tension de seuil au cours de l'endurance est identique quelque soit l'épaisseur de l'oxyde tunnel. Il apparaît néanmoins que la fenêtre de programmation de l'oxyde tunnel de 4 nm qui est légèrement plus importante, conformément à ce qui a été montré dans le paragraphe précédent.

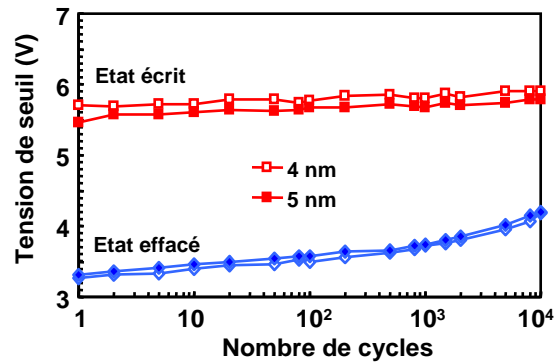


Fig. III-7 : Caractéristiques d'endurance (moyennée sur 4 dispositifs) pour deux épaisseurs d'oxyde tunnel 4 et 5 nm. Ecriture : $V_g=6,5$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$. Effacement : $V_g=-8$ V, $V_d=V_s=V_b=8$ V, $t=1$ ms. Lecture : $V_d=1$ V et $I_d=1$ μA . $t_{HTO}=10$ nm, $\Phi_{dot}=5$ nm, $R_{dot}\approx 0,15$.

Des mesures de rétention sur l'état programmé à 150°C ont été effectuées sur des dispositifs ayant une épaisseur d'oxyde tunnel de 4, 5 et 6 nm. La **Fig. III-8** montre la diminution de la tension de seuil mesurée après 168 h. On observe une dépendance de la perte de charge en fonction de l'épaisseur de l'oxyde tunnel. Par contre, lorsque l'on compare les oxydes de 5 et 6 nm (**Fig. III-8(b)-(d)**), on constate que la perte de charge est similaire. On peut donc en conclure qu'une épaisseur d'oxyde tunnel de 5 nm est suffisante pour garantir la rétention de la charge.

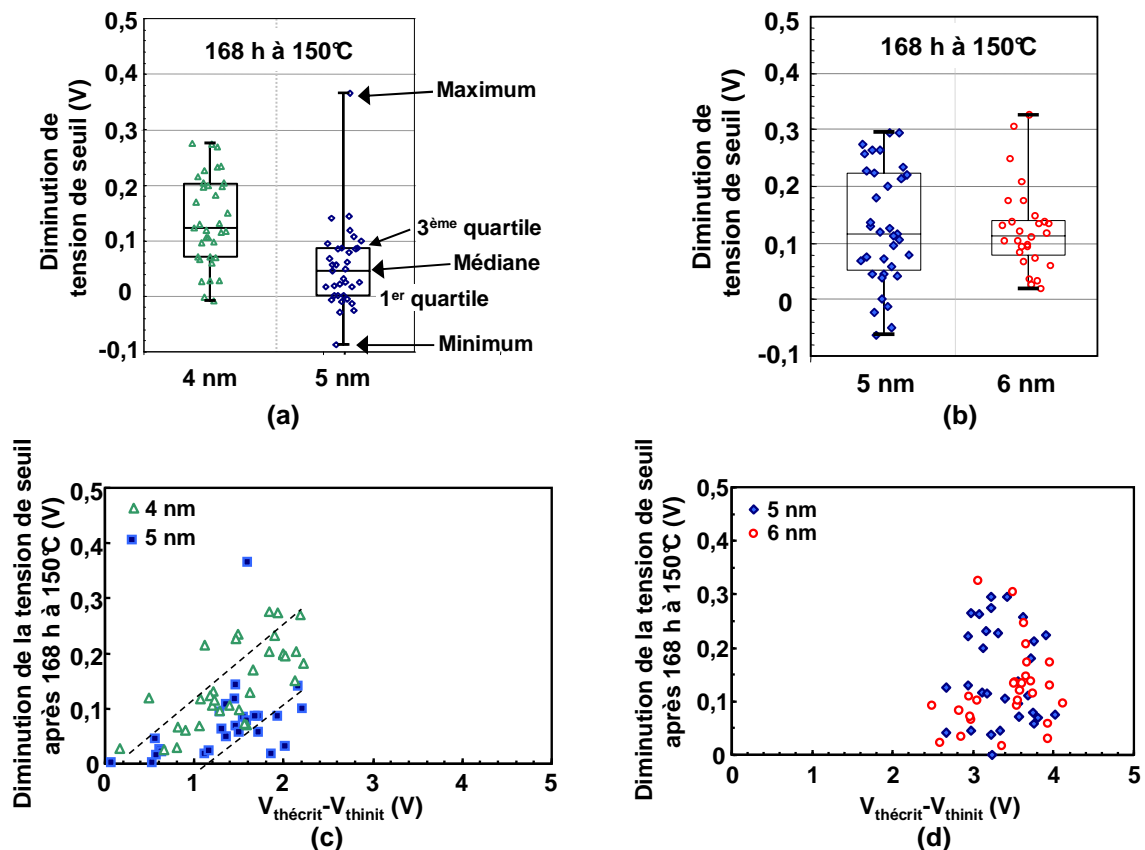


Fig. III-8 : Diminution de la tension de seuil pendant la rétention à 150°C après 168 h mesurée sur des dispositifs avec des épaisseurs d'oxyde tunnel de 4 nm et 5 nm (a) et 5 nm et 6 nm (b) et en fonction de la fenêtre de programmation initiale (c) et (d). Les symboles représentent les valeurs des tensions de seuil des 35 cellules testées pour chaque variante. Lecture : $V_d=1$ V et $I_d=1\mu A$. (a) : $t_{HTO}=10$ nm, $\Phi_{dot}=5$ nm, $R_{dot}\approx 0,15$. (b) : $t_{HTO}=10$ nm, $\Phi_{dot}=6,5$ nm, $R_{dot}\approx 0,25$.

III.2.2.3 Influence de la taille et de la densité des nanocristaux

III.2.2.3.A Influence du procédé de fabrication

Nous avons d'abord étudié l'influence du procédé de fabrication des Si-NCs. Comme il a été dit au Chapitre précédent, il est en effet possible de fabriquer des Si-NCs par LPCVD avec un procédé en une seule étape ou bien un procédé en deux étapes (la phase de nucléation des Si-NCs étant dissociée de la croissance dans ce dernier procédé [Mazen'03]). Concernant les caractéristiques d'écriture/effacement mesurées sur 20 cellules simples, la **Fig. III-9** montre que la différence de la fenêtre de programmation entre les deux variantes est presque négligeable. On peut simplement remarquer une dispersion des tensions de seuil plus élevée dans le cas des Si-NCs fabriqués avec le procédé à une étape. Ce résultat sera confirmé par la suite par des mesures effectuées sur des matrices mémoires.

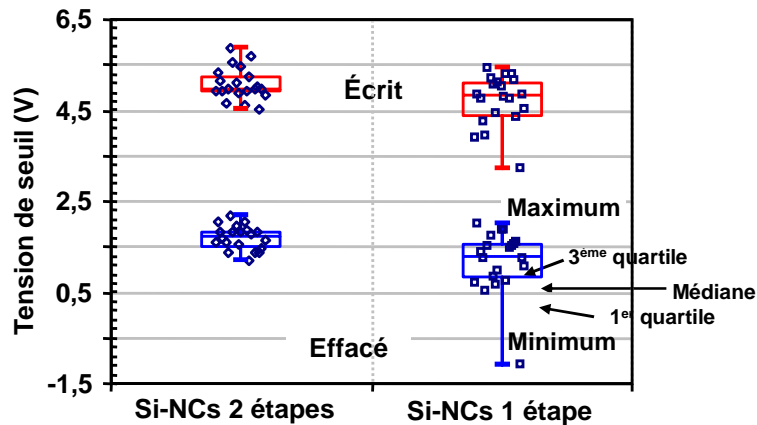


Fig. III-9 : Comparaison des tensions de seuil dans l'état écrit ($V_g=8V$, $V_d=3,75V$, $V_s=0V$, $V_b=-1,5V$, $t=10\mu s$) et effacé ($V_g=-8V$, $V_b=V_d=V_s=-8V$, $t=10ms$) pour les procédés de fabrication des Si-NCs à une étape ou à deux étapes. Les symboles représentent les valeurs des tensions de seuil des 5 cellules testées pour chaque variante. Lecture : $V_d=1V$ et $I_d=1\mu A$. (a) : $t_{tun}=5nm$, $t_{HTO}=8nm$, $\Phi_{dot}=6,5nm$, $R_{dot}\approx 0,2$.

III.2.2.3.B Influence de la taille des nanocristaux

Dans un deuxième temps, nous avons étudié l'influence de la taille des nanocristaux sur les caractéristiques électriques de la cellule. Pour cela, des dispositifs avec des Si-NCs fabriqués avec la première étape de nucléation réalisée dans des conditions identiques (la densité surfacique de sites de nucléation est donc la même) et des diamètres croissants de 4 à 12 nm, ont été fabriqués (**Fig. III-10**). Ceci est équivalent à augmenter le taux de couverture des Si-NCs entre 11% et 43%. La **Fig. III-11** montre la distribution en taille des Si-NCs. On peut voir que plus le diamètre est élevé, plus la dispersion est grande.

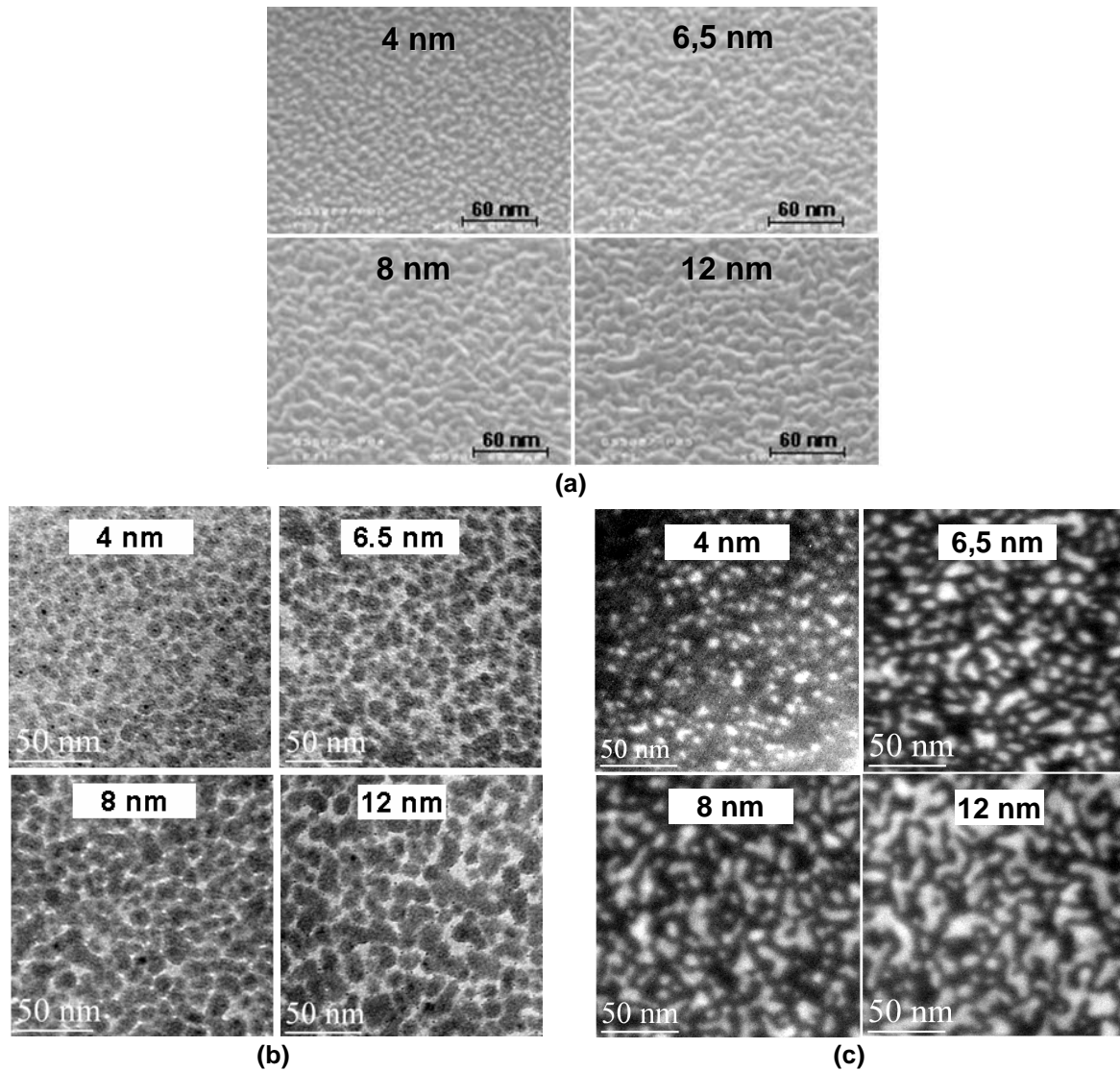


Fig. III-10 : Images MEB (a), TEM en champ clair (b) et EFTEM (Energy-Filtered TEM) (c) de Si-NCs fabriqués avec le procédé LPCVD à 2 étapes avec la première étape de nucléation réalisée dans des conditions identiques et la deuxième étape de croissance avec des temps différents conduisant à des diamètres moyens croissants (4 nm, 6,5 nm, 8 nm et 12 nm). Les taux de couverture correspondants sont : 0,11 ; 0,25 ; 0,34 ; 0,43. Dans l'image TEM en champ clair, les Si-NCs sont représentés en noir et le SiO₂ en gris clair. Dans l'image EFTEM, les Si-NCs sont représentés en blanc et le SiO₂ en noir.

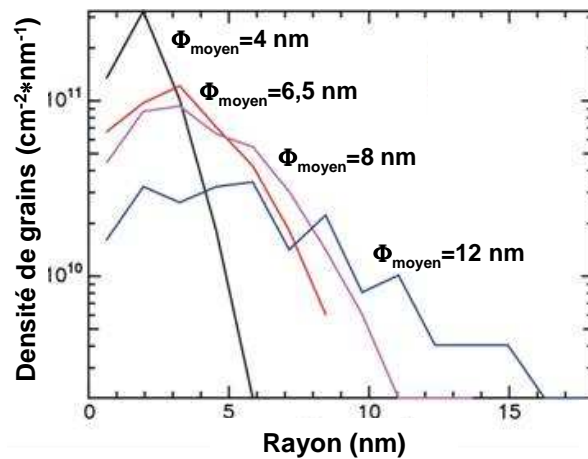


Fig. III-11 : Distributions du rayon des Si-NCs obtenues d'après les images EFTEM pour les quatre échantillons avec des Si-NCs de diamètre moyen croissant.

Les dynamiques cumulées d'écriture en porteurs chauds et d'effacement en Fowler-Nordheim ont d'abord été mesurées sur les différents échantillons. La **Fig. III-12** montre que lorsque la taille des Si-NCs augmente (ce qui correspond ici à une augmentation du taux de couverture), la vitesse d'écriture et le niveau de saturation de la fenêtre de programmation augmentent. Comme nous le montrerons dans le chapitre IV, ceci est en accord avec le modèle théorique [DeSalvo'01] qui affirme que la fenêtre de programmation augmente linéairement avec la surface couverte par les Si-NCs.

Au contraire, les **Fig. III-13(a)-(b)-(c)-(d)** qui présentent les mesures d'effacement pour différentes tensions de grille et pour des dispositifs avec des Si-NCs de 4 nm, 6,5 nm et 8 nm montrent clairement que l'effacement est plus rapide lorsque la taille des Si-NCs diminue. Avec des nanocristaux de taille réduite, les électrons sont stockés dans un plus petit volume. Le champ électrique dans l'oxyde tunnel à proximité des Si-NCs au début de l'effacement est alors plus élevé avec des Si-NCs de plus petit diamètre, ce qui explique une plus grande rapidité d'effacement. Ceci sera démontré par la modélisation dans le chapitre IV.

Finalement, la **Fig. III-14** montre les mesures d'écriture/effacement effectuées sur 20 cellules pour chaque variante de Si-NCs. Premièrement, on voit que la fenêtre de programmation (la différence entre V_{th} écrit et V_{th} effacé) augmente lorsque le diamètre des Si-NCs augmente, conformément aux résultats précédents. La seconde conclusion qui apparaît est que la dispersion de la tension de seuil augmente avec la taille des Si-NCs, ce qui est dû à l'augmentation de la dispersion en taille des plus gros nanocristaux.

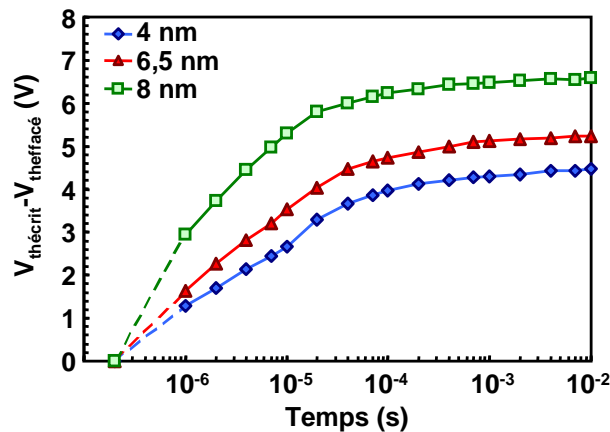


Fig. III-12 : Caractéristiques d'écriture en porteurs chauds ($V_g=8V$, $V_d=3,75V$, $V_b=-1,5V$, $V_s=0 V$) en fonction du temps cumulé de stress pour différentes variantes de taille de Si-NCs (correspondant à différents taux de couverture). Lecture : $V_d=1 V$, $I_d=1 \mu A$. $t_{tun}=5 nm$, $t_{HTO}=10 nm$.

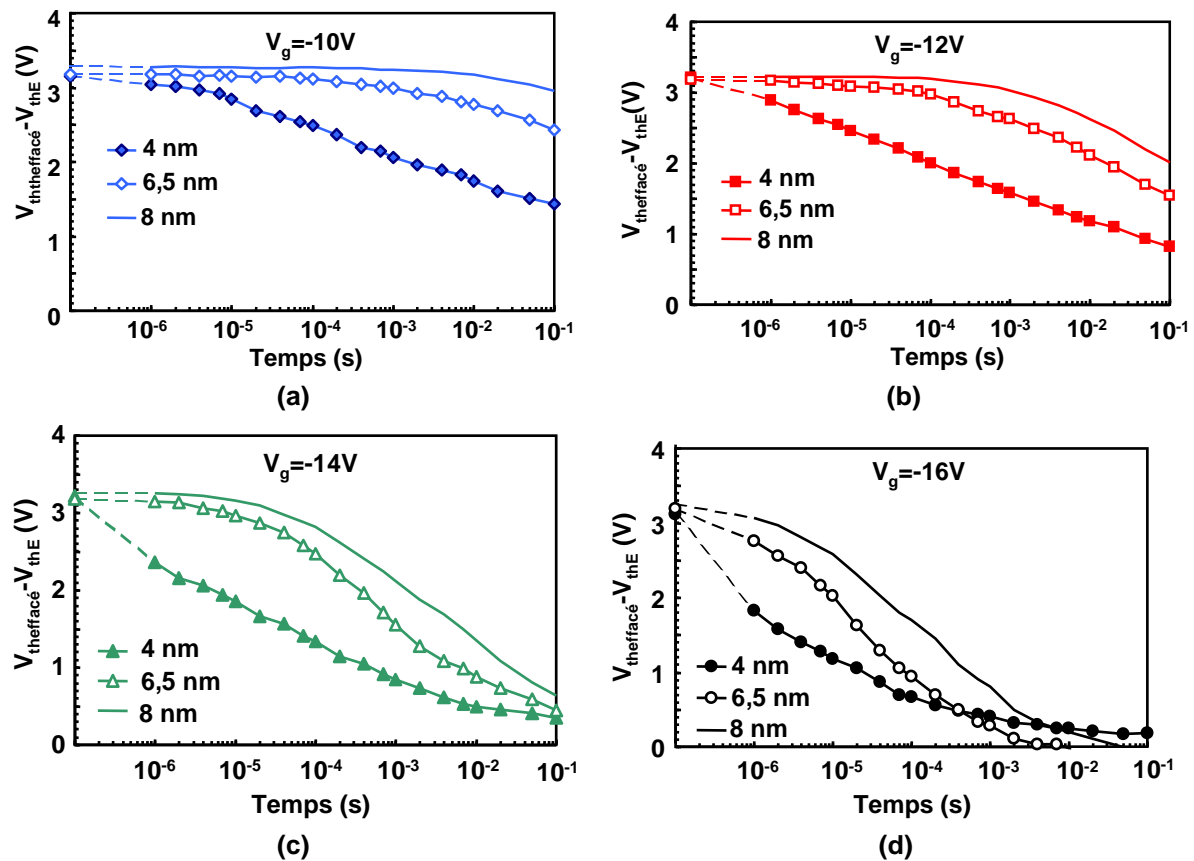


Fig. III-13 : Caractéristiques d’effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour différentes variantes de tailles de Si-NCs (correspondant à différents taux de couverture) et différentes tensions de grille d’effacement : -10 V (a), -12 V (b), -14 V (c), -16 V (d). Lecture : $V_d=1$ V et $I_d=1$ μ A. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm.

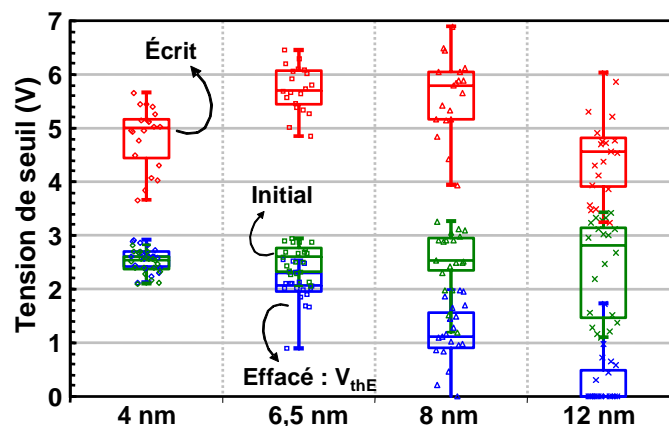


Fig. III-14 : Mesures de la tension de seuil initiale, après effacement ($V_g=-8$ V, $V_b=V_d=V_s=8$ V, $t=10$ ms) et après écriture ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10$ μ s) sur 20 cellules avec les quatre variantes de taille de Si-NCs. Lecture : $V_d=1$ V et $I_d=1$ μ A. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm.

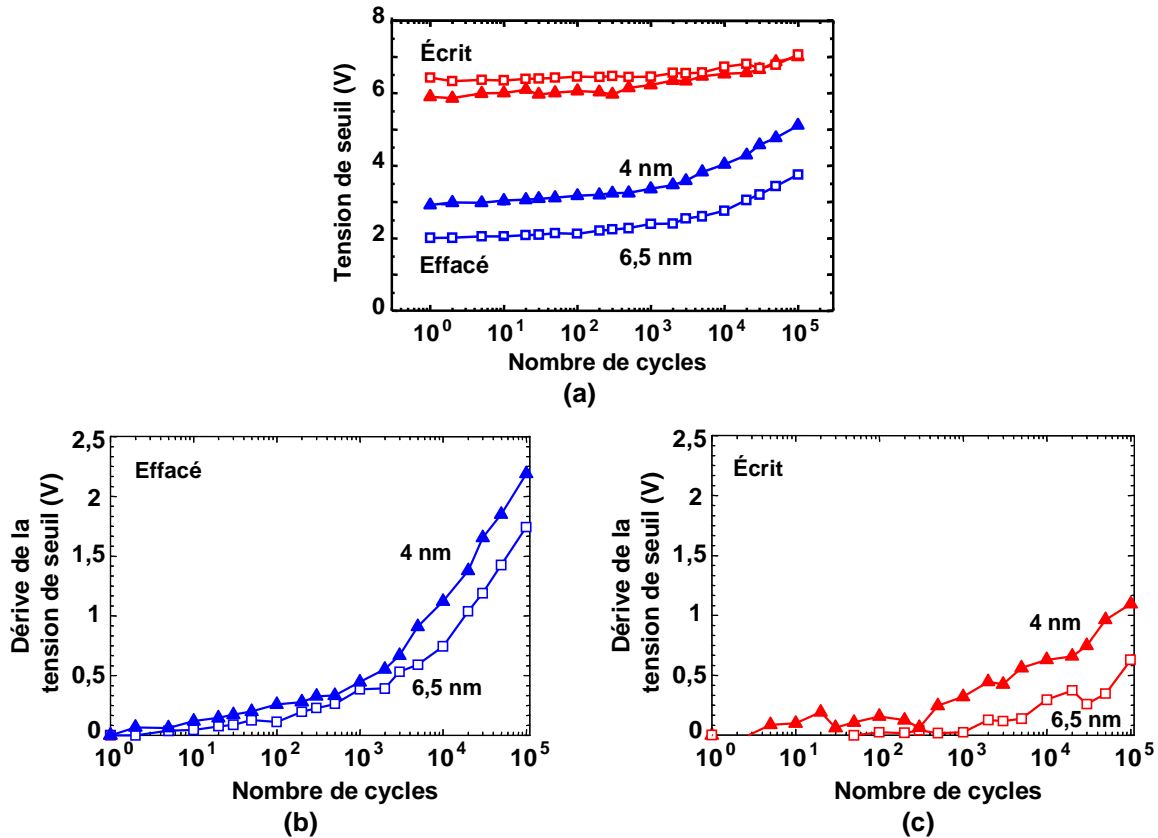


Fig. III-15 : (a) Tensions de seuil au cours de l'endurance et dérive des tensions de seuil effacées (b) et écrites (c) correspondantes mesurées sur des dispositifs avec deux tailles de Si-NCs (4 nm et 6,5 nm). Écriture : $V_g=9$ V, $V_d=4$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$ et effacement : $V_g=-15,5$ V, $V_d=V_s=V_b=0$ V, $t=10ms$. Lecture : $V_d=1V$ et $I_d=100$ nA. $t_{tun}=5$ nm, $t_{HTO}=10$ nm.

L'endurance a été réalisée sur des cellules avec des nanocristaux de 4 et 6,5 nm (ce qui correspond à un taux de couverture de 0,11 et 0,25) (**Fig. III-15(a)**). On observe que la dérive de la tension de seuil au cours de l'endurance est plus élevée pour les nanocristaux les plus petits et donc le taux de couverture le plus faible (**Fig. III-15(b)-(c)**). Cela peut s'expliquer en première approximation par le fait que lorsque le taux de couverture diminue, les électrons peuvent plus facilement atteindre le HTO pour y être piégés de façon irréversible.

III.2.2.4 Influence de la passivation des nanocristaux

III.2.2.4.A Passivation avec NH_3 et NO

Nous avons évoqué dans le Chapitre II l'importance de l'étape de passivation des Si-NCs afin de les protéger de l'oxydation. Nous allons donc étudier dans cette section l'influence de différents types de passivations (en particulier, NH_3 et NO) sur les caractéristiques électriques de la cellule à nanocristaux.

Dans un premier lot, les Si-NCs ont été soit passivés sous NO à $950^\circ C$, soit non passivés. A noter que dans ces dispositifs étudiés, les Si-NCs ont été intégrés dans un procédé de type « froid » et n'ont donc pas subi de fortes oxydations. La capacité à protéger de

l'oxydation des Si-NCs sera étudiée dans le paragraphe suivant. Ici, l'objectif est d'examiner l'influence de la nitruration NO sur les dynamiques d'écriture/effacement.

La **Fig. III-16(a)** montre que la nitruration NO ralentit l'écriture par porteurs chauds, le niveau final de saturation de la fenêtre de programmation étant néanmoins identique. Par exemple, à $10\mu\text{s}$, le décalage de la tension de seuil il existe une différence de $1,2\text{V}$ entre les deux variantes de Si-NCs. Par contre, en ce qui concerne l'effacement, la nitruration NO n'a pas d'influence, comme le montre la **Fig. III-16(b)**.

Nous avons ensuite étudié la capacité de protection des nitrurations NH_3 et NO des Si-NCs par rapport à une oxydation parasite. Les dispositifs ont subi différentes densifications sous N_2 (non oxydante) et sous O_2 à 950°C , après l'étape de dépôt du HTO. La densification sous O_2 est équivalente à la croissance d'un oxyde de 140 \AA . Les dynamiques d'écriture par porteurs chauds (**Fig. III-17**) permettent de tirer plusieurs conclusions :

- On constate que, après une densification O_2 à 950°C , les Si-NCs passivés présentent des caractéristiques d'écriture meilleures que les Si-NCs non passivés. La passivation offre donc bien une protection contre l'oxydation des Si-NCs, avec une légère amélioration dans le cas de la passivation sous NO 950°C par rapport à la passivation sous NH_3 750°C .
- L'oxydation des Si-NCs conduit à une écriture plus lente et à un niveau de saturation beaucoup plus faible, ce qui est significatif d'une réduction de la taille des Si-NCs.
- Enfin, la passivation n'empêche pas complètement l'oxydation des Si-NCs. On observe en effet que les tensions de seuil des cellules écrites avec les Si-NCs passivés et oxydés sont inférieures par rapport à celles des dispositifs ayant uniquement été densifiés sous N_2 .

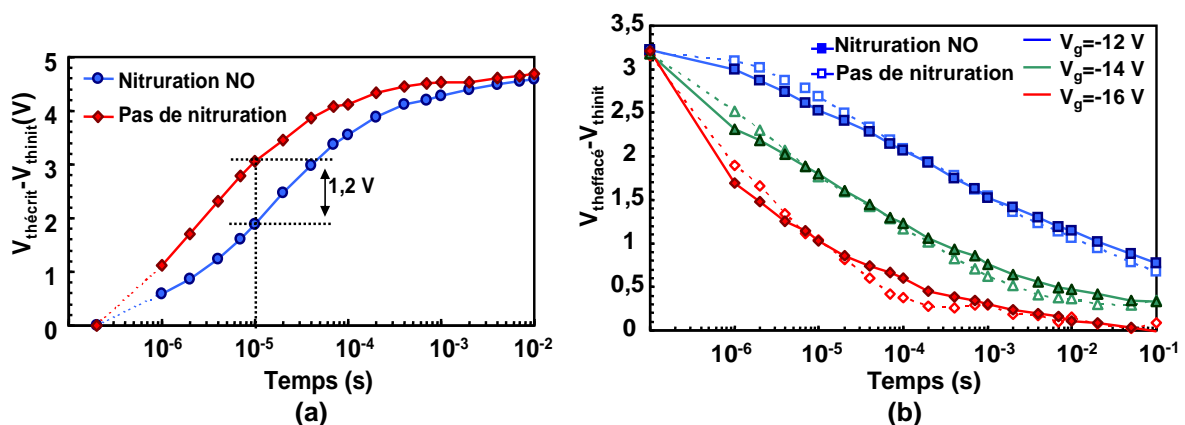


Fig. III-16 : (a) : Caractéristiques d'écriture en porteurs chauds ($V_g=8 \text{ V}$, $V_d=3,75 \text{ V}$, $V_b=-1,5 \text{ V}$, $V_s=0\text{V}$) en fonction du temps cumulé de stress pour des Si-NCs non passivés et des Si-NCs passivés avec du NO. (b) : Caractéristiques d'effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour différentes variantes de tailles de Si-NCs et différentes tensions de grille d'effacement ($V_d=V_s=V_b=0 \text{ V}$). Lecture : $V_d=1\text{V}$ et $I_d=1 \mu\text{A}$. $t_{\text{un}}=5 \text{ nm}$. $t_{\text{HTO}}=8 \text{ nm}$, $\Phi_{\text{dot}}=6,5 \text{ nm}$, $R_{\text{dot}}\approx 0,2$.

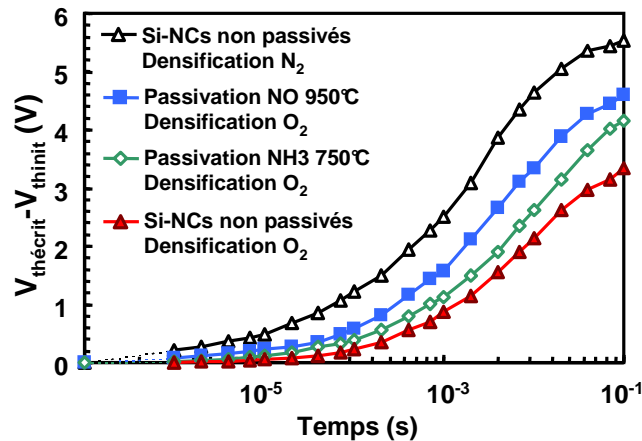


Fig. III-17 : Caractéristiques d'écriture en porteurs chauds ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V) en fonction du temps cumulé de stress pour des Si-NCs non passivés et des Si-NCs passivés avec du NO à 950°C ou sous NH_3 à 750°C suivis par une densification sous N_2 à 950°C ou sous O_2 à 950°C (oxyde équivalent 140 Å). Lecture : $V_d=1$ V et $I_d=1$ μA . $t_{\text{un}}=5$ nm, $t_{\text{HTO}}=10$ nm.

III.2.2.4.B Dépôt d'une couche de Si_3N_4

En plus de la nitruration par NH_3 ou NO , deux autres variantes ont été testées.

La première consiste à fabriquer les Si-NCs dans un four LPVCD et à déposer ensuite dans ce four une couche fine de nitrure sélectivement sur les Si-NCs (**Fig. III-18(a)**). L'épaisseur de la couche de nitrure est estimée à environ 2 nm. L'avantage de cette méthode est d'effectuer le dépôt des nanocristaux et la passivation dans le même équipement, ce qui évite ainsi une oxydation partielle des Si-NCs lorsque les plaques sont sorties du four.

La deuxième variante consiste à déposer les Si-NCs dans un bâti d'épitaxie puis à déposer une couche de nitrure continue d'environ 2 nm dans un four (**Fig. III-18 (b)**).

Tout d'abord, les résultats d'écriture/effacement de la **Fig. III-19(a)-(b)** montrent qu'avec la couche sélective continue de nitrure, il est difficile de revenir à l'état initial après l'effacement, contrairement aux Si-NCs non passivés. Ceci est dû à un piégeage irréversible d'électrons dans le nitrure. Afin de confirmer ce résultat, des mesures d'endurance présentées sur la **Fig. III-19(b)-(c)** ont été réalisées sur les dispositifs avec des Si-NCs non passivés et avec une couche sélective de nitrure. Pour la cellule avec la couche sélective de nitrure on observe une fermeture totale de la fenêtre de programmation après 100K cycles, avec une forte augmentation de la tension de seuil effacée, due à la difficulté à effacer la charge piégée dans le nitrure.



Fig. III-18 : Schémas illustrant la passivation des Si-NCs par une couche sélective (a) ou continue (b) de nitrure.

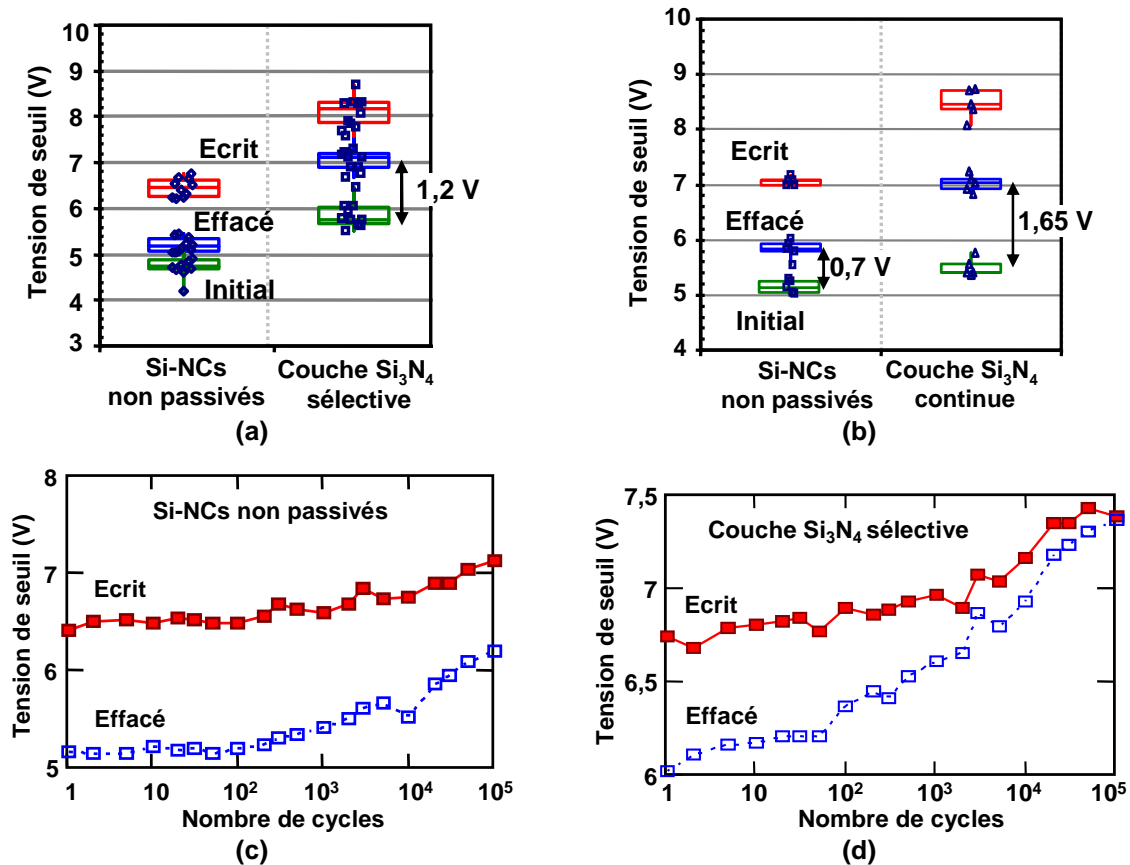


Fig. III-19: Mesures des tensions de seuil des cellules écrites ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$) et effacées ($V_g=-8$ V, $V_b=V_d=V_s=8$ V, $t=10ms$) avec des Si-NCs non passivés et des Si-NCs passivés avec une couche sélective de nitrure (a) ou avec une couche continue (b). Lecture : $V_d=1$ V et $I_d=1$ μA . Caractéristiques d'endurance des dispositifs avec (c) des Si-NCs non passivés et (d) des Si-NCs passivés avec une couche sélective de nitrure. Écriture : $V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$ et effacement : $V_g=-7$ V, $V_s=V_d=V_b=7$ V, $t=10$ ms. Lecture : $V_d=$ et 1 V $I_d=100$ nA.

III.2.2.5 Influence du diélectrique de contrôle

III.2.2.5.A Influence de la nature du diélectrique : HTO ou ONO

Dans cette partie, deux diélectriques de contrôle ont été testés : HTO et ONO (Oxyde-Nitride-Oxyde). L'ONO est le diélectrique qui est actuellement utilisé dans les mémoires Flash standard. L'avantage de la présence d'une couche de nitrure est qu'il constitue une barrière à l'oxydation très efficace et, de plus, les fuites à travers cet empilement sont moins importantes qu'avec un HTO. L'utilisation d'un ONO simplifie donc l'intégration des Si-NCs, car on s'affranchit des problèmes d'oxydation des Si-NCs au cours de la croissance des oxydes de grille de la périphérie CMOS. Toutefois, il avait été démontré dans la littérature [Steimle'04] [Gerardi'07] qu'un piégeage parasite dans le nitrure a lieu lors de l'écriture par porteurs chauds. Afin de vérifier ce phénomène, des cellules avec des Si-NCs et un oxyde de contrôle de type HTO ou ONO, ainsi que des cellules de référence sans Si-NCs ont été fabriquées. La caractérisation électrique de l'écriture par porteurs chauds (Fig. III-20(a)) et de l'effacement par Fowler-Nordheim (Fig. III-20(b)) de ces dispositifs montre un piégeage

parasite et irréversible d'électrons dans la couche de nitrure de l'ONO, qui se superpose au piégeage/dépiégeage d'électrons dans les Si-NCs.

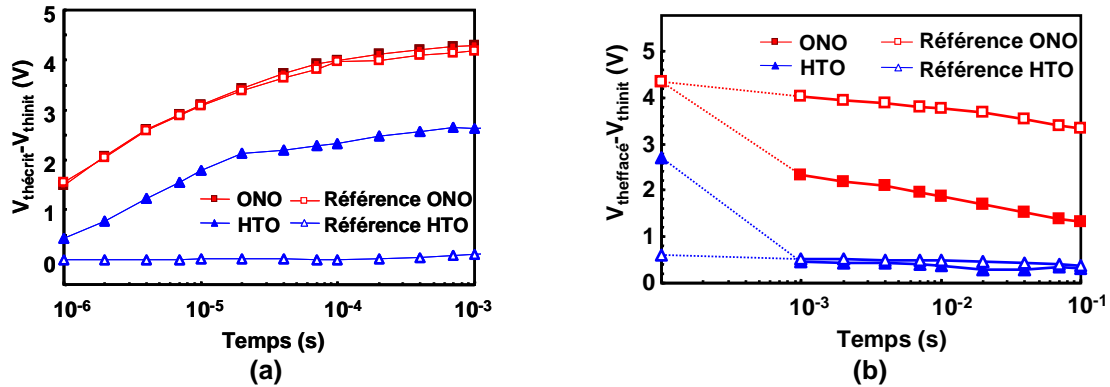


Fig. III-20 : Caractéristiques d'écriture en porteurs chauds (a) ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V) et d'effacement en Fowler-Nordheim (b) ($V_g=-8$ V, $V_s=V_d=V_b=8$ V) en fonction du temps cumulé de stress pour les dispositifs avec des Si-NCs (et les dispositifs de références sans Si-NCs) avec HTO ou ONO en tant que diélectrique de contrôle. Lecture : $V_d=1$ V et $I_d=1$ μ A.

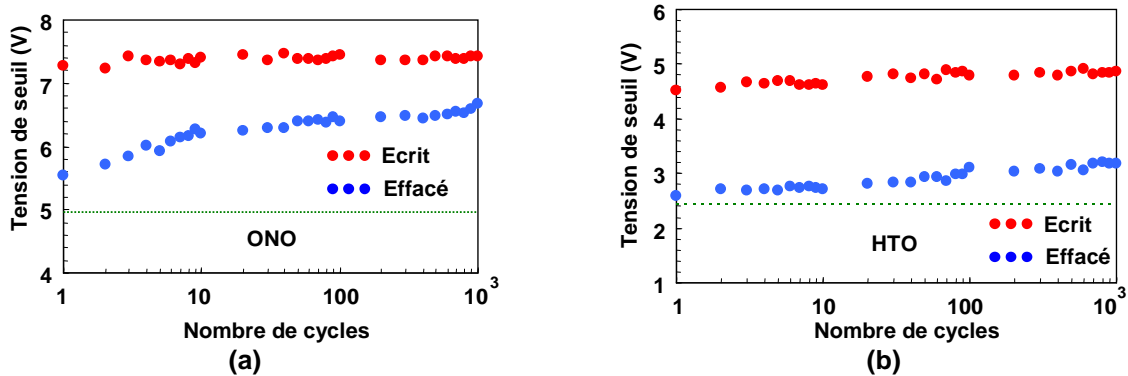


Fig. III-21 : Caractéristiques d'endurance sur les dispositifs avec (a) Si-NCs et ONO et (b) Si-NCs et HTO. Écriture : $V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10$ μ s ; effacement : $V_g=-8$ V, $V_s=V_d=V_b=8$ V, $t=10$ ms. Lecture : $V_d=1$ V et $I_d=0,1$ μ A. $t_{tun}=5$ nm, $t_{HTO}=10$ nm/ $EOT_{ONO}\approx 13$ nm, $\Phi_{dot}\approx 5$ nm, $R_{dot}\approx 0,15$.

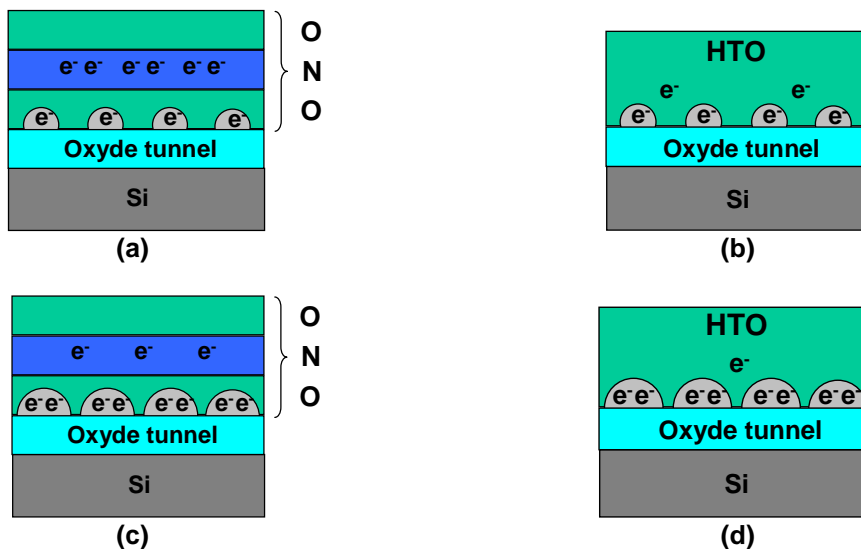


Fig. III-22 : Schémas illustrant le piégeage dans une cellule mémoire avec un diélectrique de grille ONO (a), (c) et HTO (b), (d) pour un taux de couverture de Si-NCs faible (a), (b) et élevé (c), (d).

Nous avons également effectué des tests endurance sur ces dispositifs. On voit clairement que l'endurance la meilleure est obtenue avec le HTO (**Fig. III-21(a)**). Avec la cellule ONO (**Fig. III-21(b)**), la fenêtre de programmation se ferme à cause de la forte augmentation du V_{th} effacé pendant l'endurance, due au piégeage parasite dans le nitrure.

On peut donc conclure que l'utilisation d'un ONO standard comme diélectrique de contrôle n'est pas optimale pour les mémoires à Si-NCs. On peut cependant rappeler que nous avons montré au paragraphe III.2.2.3.B que plus le taux de couverture des Si-NCs était élevé, plus le piégeage parasite dans le HTO était faible. On peut donc penser qu'une façon de diminuer le piégeage parasite dans le nitrure serait d'augmenter le taux de couverture des nanocristaux (**Fig. III-22**). Toutefois, le piégeage dans l'ONO restera toujours supérieur à celui du HTO. C'est pourquoi nous avons choisi d'étudier par la suite uniquement des variantes à base de HTO.

III.2.2.5.B Influence de l'épaisseur, densification du HTO

Le diélectrique de contrôle joue un rôle important dans la fiabilité de la cellule mémoire. La présence de pièges peut donner lieu à une dégradation des caractéristiques d'endurance. Un courant de fuite trop élevé peut donner lieu à une perte de la charge stockée au cours du test de rétention et également pendant la lecture (« gate disturb »). C'est pourquoi nous avons étudié deux paramètres susceptibles d'améliorer la fiabilité des dispositifs : la densification du HTO et son épaisseur.

III.2.2.5.i Influence de la densification du HTO

Le premier paramètre à l'étude est la densification de l'oxyde de contrôle HTO. En effet, un oxyde déposé n'est jamais parfait : il n'est pas toujours stoechiométrique (SiO_x avec $x < 2$) et il contient des défauts (lacunes d'oxygène, liaisons pendantes). Les recuits thermiques à haute température sous O_2 permettent alors de rendre l'oxyde stoechiométrique et de diminuer les défauts en densifiant l'oxyde. L'inconvénient est le risque d'oxydation des Si-NCs si l'oxygène diffuse à travers le HTO. Il est alors possible d'effectuer un recuit sous N_2 qui ne rend pas l'oxyde plus stoechiométrique, mais qui permet quand même de diminuer la quantité de défauts tels que les lacunes d'oxygène et les liaisons pendantes, conduisant à une densification de l'oxyde. Pour cette étude, plusieurs variantes de densifications ont été réalisées : sous N_2 à plusieurs températures : 850°C, 950°C et 1000°C et sous O_2 à 850°C. Il faut noter que pour ces variantes, tous les Si-NCs ont été nitrurés sous NO au préalable pour empêcher une éventuelle oxydation.

Nous avons d'abord effectué des tests d'endurance sur les différentes variantes. Les caractéristiques sont représentées sur les **Fig. III-23(a)-(b)-(c)-(d)-(e)**. On observe une dégradation de l'endurance avec le HTO non densifié, en particulier sur l'état effacé. Afin de quantifier plus précisément les différences entre les différentes options technologiques, nous avons représenté l'augmentation de la tension de seuil sous forme d'histogrammes au cours de l'endurance pour l'état écrit **Fig. III-23(f)** et l'état effacé **Fig. III-23(g)**. Sur l'état écrit,

aucune différence majeure n'apparaît entre les différentes densifications. Par contre, sur l'état effacé, on observe que la dérive de la tension de seuil des dispositifs avec un HTO non densifié est largement supérieure à celle des dispositifs avec un HTO densifié. Par contre, aucune différence majeure n'est observée entre les différentes densifications.

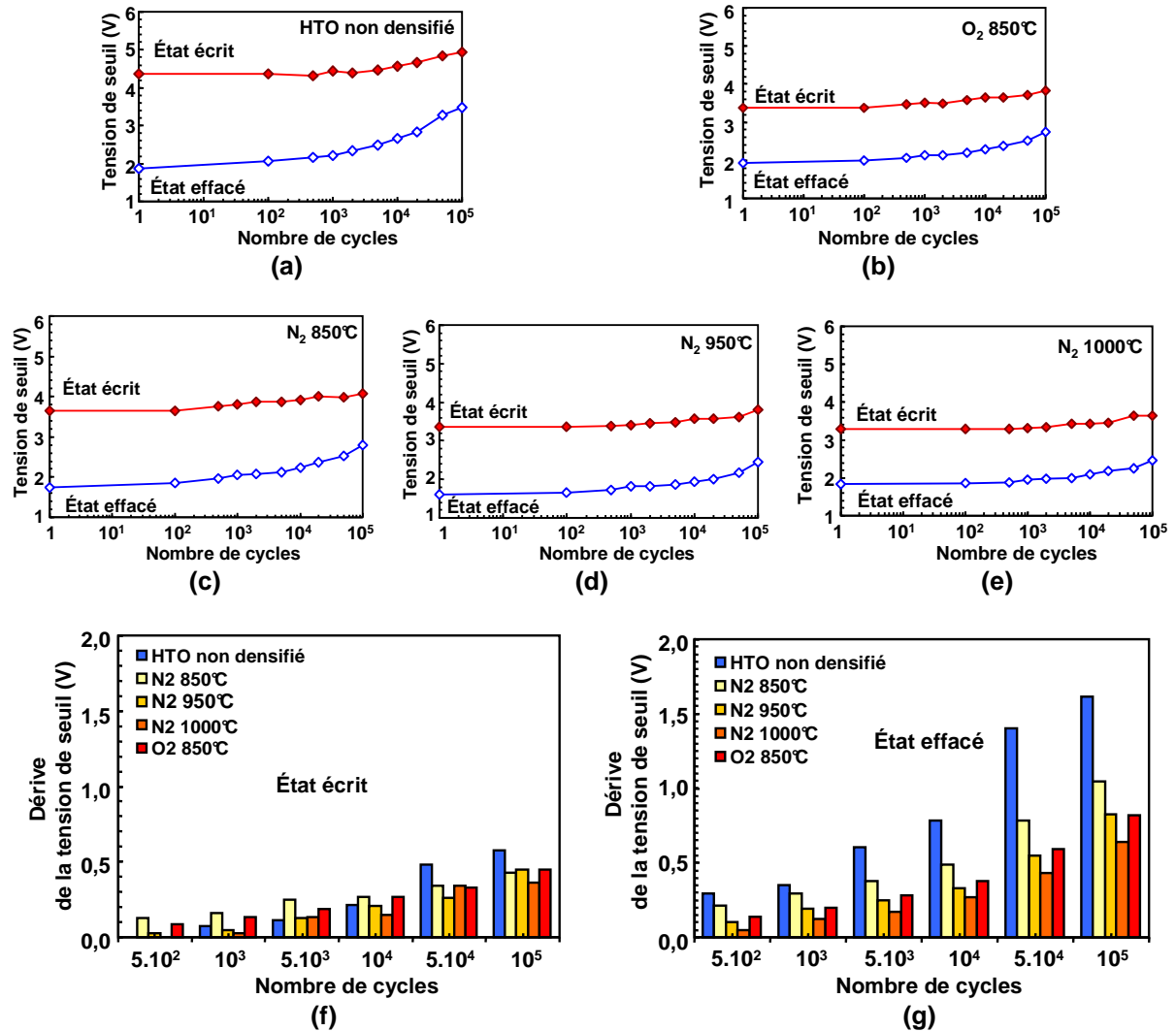


Fig. III-23 : Caractéristiques d'endurance en fonction des différentes densifications du HTO : pas de densification (a), O₂ 850°C (b), N₂ 850°C, N₂ 950°C, N₂ 1000°C. Ecriture : V_g=9 V, V_d=3,75 V, V_b=-1,5 V, V_s=0 V, t=10μs et effacement : V_g=-8 V, V_s=V_d=V_b=8 V, t=10 ms. Histogrammes représentant la dérive de la tension de seuil au cours de l'endurance de l'état écrit (f) et effacé (g) pour les différentes densifications du HTO. Lecture : V_d=1 V et I_d=1 μA. t_{un}=5 nm, t_{HTO}=10 nm, Φ_{dot}≈6,5 nm, R_{dot}≈0,2.

III.2.2.5.ii Influence de l'épaisseur de HTO

L'épaisseur du HTO a une influence sur le couplage entre les Si-NCs et la grille de contrôle ($\alpha_{GC} = C_{CG}/C_T$ avec $C_{GC} = \epsilon_{SiO_2} S/T_{HTO}$, avec S la surface de couplage des Si-NCs et T_{HTO} l'épaisseur du HTO) et également sur la fenêtre de programmation ($\Delta V_{th} = -\Delta Q/C_{CG}$). C'est pourquoi nous avons d'abord étudié l'influence de l'épaisseur du HTO sur les dynamiques d'écriture par porteurs chauds et d'effacement par Fowler-Nordheim.

On peut voir que le décalage de la tension de seuil au cours de l'écriture par porteurs chauds est sensiblement le même quelque soit l'épaisseur du HTO (**Fig. III-24(a)**).

Par contre, concernant l'effacement par Fowler-Nordheim, on observe une augmentation de la vitesse d'effacement lorsque l'épaisseur du HTO diminue (**Fig. III-24(b)-(c)**). Le diagramme de bandes de la **Fig. III-24(d)** montre que la diminution de l'épaisseur du HTO se traduit par une augmentation du champ électrique dans l'oxyde tunnel grâce à un meilleur couplage entre la grille de contrôle et les Si-NCs. Ceci sera expliqué plus en détail dans le chapitre suivant.

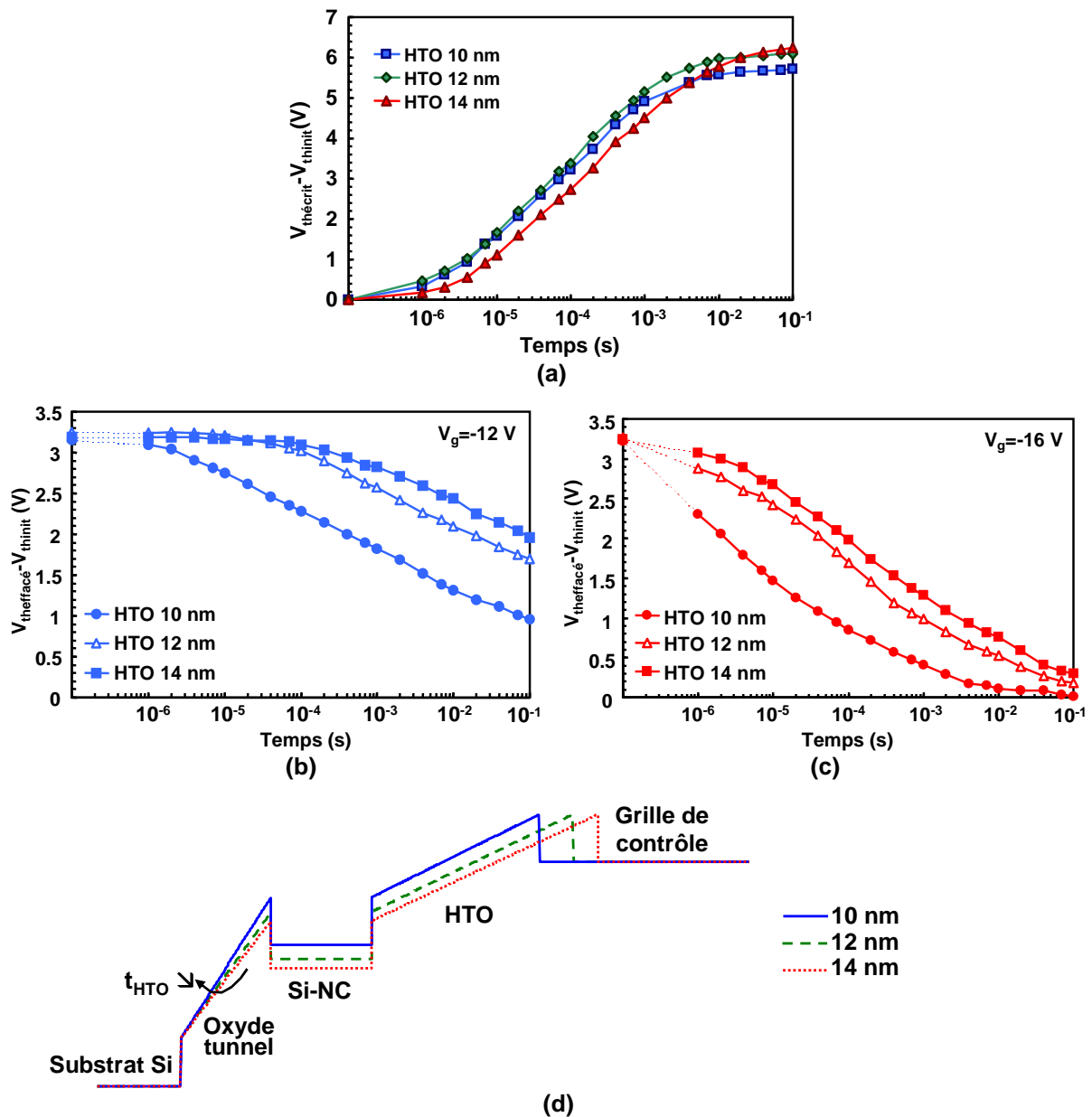


Fig. III-24 : (a) : Caractéristiques d'écriture en porteurs chauds ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V) en fonction du temps cumulé de stress pour différentes épaisseurs de HTO.

(b), (c) : Caractéristiques d'effacement en Fowler-Nordheim en fonction du temps cumulé de stress pour différentes épaisseurs de HTO et différentes tensions de grille d'effacement ($V_d=V_d=V_b=0$ V).

Lecture : $V_d=1$ V et $I_d=1$ μ A. $t_{\text{tun}}=5$ nm, $\Phi_{\text{dot}}\approx 6,5$ nm, $R_{\text{dot}}\approx 0,25$.

(d) : Schéma de la bande de conduction en énergie dans l'empilement de grille de la cellule en fonction de l'épaisseur du HTO.

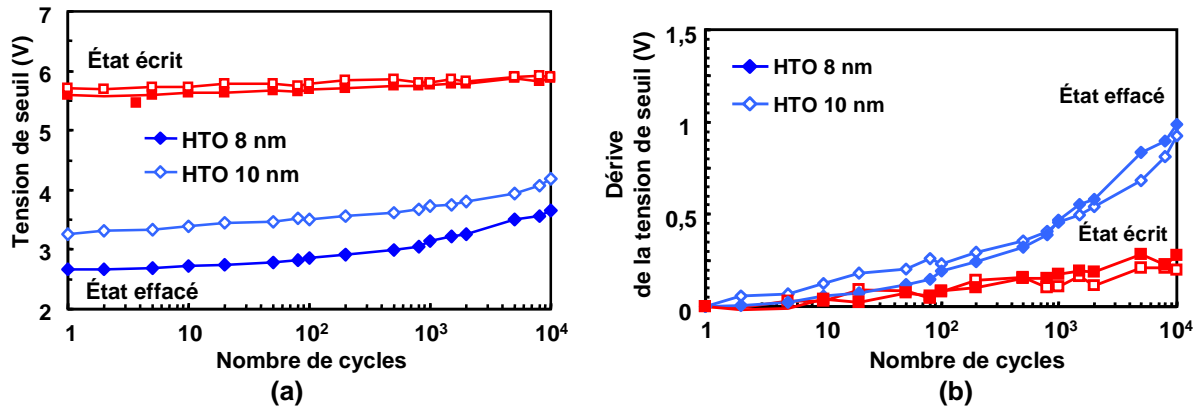


Fig. III-25 : (a) : Caractéristiques d'endurance en fonction de l'épaisseur du HTO (8 et 10 nm). Ecriture : $V_g=6,5$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$ et effacement : $V_g=-8$ V, $V_s=V_d=V_b=8$ V, $t=1$ ms. (b) : Dérive de la tension de seuil (b) au cours de l'endurance en fonction de l'épaisseur du HTO. Lecture : $V_d=1$ V et $I_d=1$ μA . $t_{tun}=4$ nm, $\Phi_{dot}\approx 5$ nm, $R_{dot}\approx 0,15$.

Les caractéristiques d'endurance pour deux épaisseurs de HTO sont comparées dans la **Fig. III-25(a)**. On n'observe pas de différence majeure entre les caractéristiques des deux épaisseurs de HTO, si ce n'est la fenêtre de programmation qui est plus grande pour le HTO le plus fin, conformément à ce qui a été démontré auparavant (la tension de seuil effacée est plus petite pour le HTO 8 nm). Afin d'avoir une comparaison plus précise, la dérive des tensions de seuil écrites et effacées est représentée dans la **Fig. III-25(b)**. On n'observe aucune différence de la dérive de la tension de seuil au cours des cycles d'écriture/effacement entre les deux épaisseurs de HTO.

Des mesures de rétention à 150°C pendant 168 h ont été réalisées pour des HTO de 8, 10 et 14 nm sur une cellule écrite (**Fig. III-26(a)-(b)**). Afin de comparer rigoureusement les échantillons, nous avons représenté la diminution de la tension de seuil en fonction de la charge injectée pendant l'écriture. Pour des conditions d'écriture identiques, la charge injectée dans les nanocristaux augmente lorsque l'épaisseur du HTO diminue **Fig. III-26(c)**. C'est ce qui explique la charge injectée plus élevée pour les HTO les plus fins sur les **Fig. III-26(a)** et **(b)**. On voit sur ces graphiques que plus la charge injectée initialement est grande, plus la perte pendant la rétention est importante, mais l'épaisseur du HTO ne semble pas avoir d'influence sur la rétention de la charge.

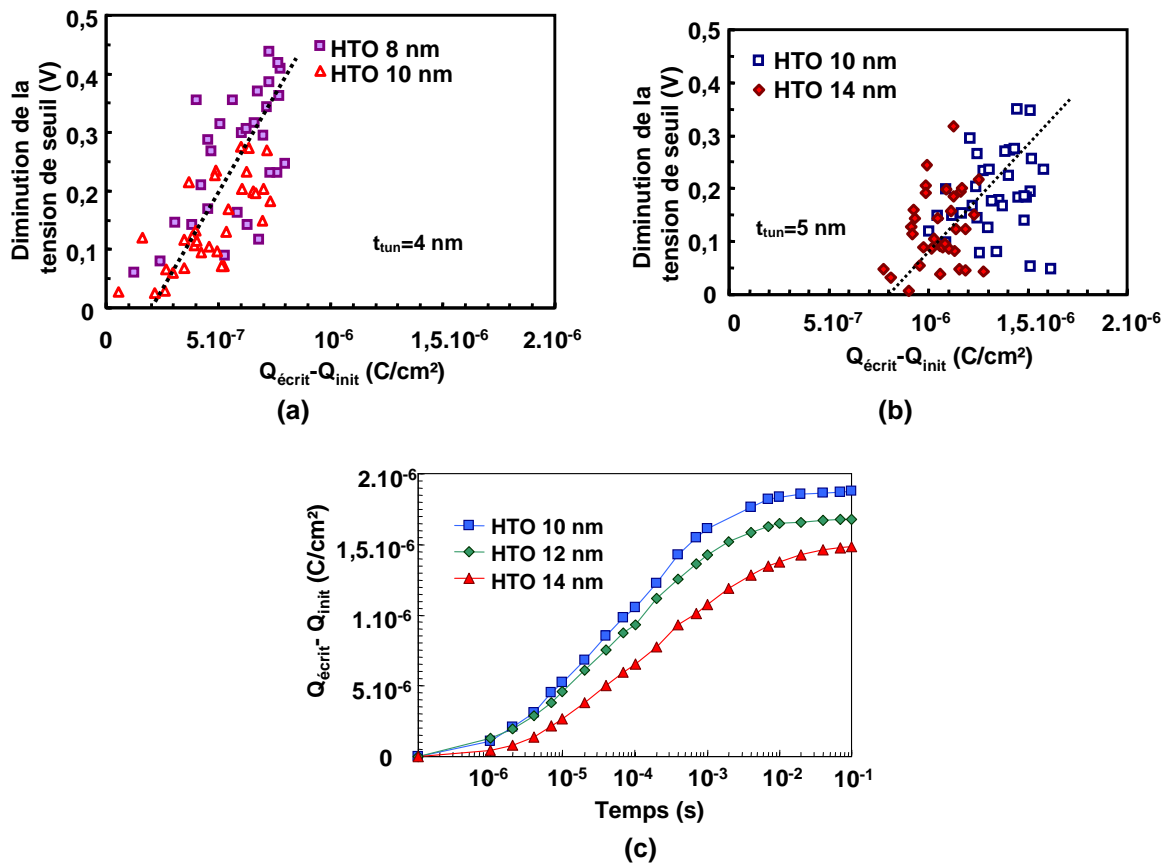


Fig. III-26 : Mesures expérimentales sur 35 cellules de la diminution de la tension de seuil de l'état écrit après 168 h à 150°C en fonction de la charge injectée pendant l'écriture pour des épaisseurs de HTO de (a) 8 et 10 nm et (b) 10 et 14 nm. (c) : Variation de la charge injectée pendant l'écriture ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10\mu s$) pour des épaisseurs de HTO différentes. Lecture : $V_d=1$ V et $I_d=1$ μA . (a) : $t_{tun}=4$ nm, $\Phi_{dot}\approx 5$ nm, $R_{dot}\approx 0,15$. (b), (c) : $t_{tun}=5$ nm, $\Phi_{dot}\approx 6,5$ nm, $R_{dot}\approx 0,25$.

Des mesures de "gate disturb" accéléré ont été aussi réalisées sur des dispositifs ayant différentes épaisseurs de HTO, allant de 8 nm à 14 nm. On parle de "disturb" accéléré car de fortes tensions positives ont été appliquées sur la grille de contrôle, beaucoup plus élevées que les tensions appliquées lors de la lecture sur un produit. Ceci permet de voir plus facilement les effets des tensions et de l'épaisseur du HTO sur la diminution de la tension de seuil.

Nous avons d'abord réalisé sur des dispositifs avec des HTO de 8 et 10 nm des mesures avec des tensions de grille de 0 ; 6 ; 8 et 10 V (**Fig. III-27(a)-(b)**). Lorsque la tension de grille augmente, la variation de la tension de seuil s'accélère. Ensuite, des dispositifs avec des épaisseurs plus importantes de HTO (10, 12 et 14 nm) ont été comparés sur la **Fig. III-27(c)**. Ces mesures montrent clairement l'importance de l'épaisseur du HTO sur le "gate disturb".

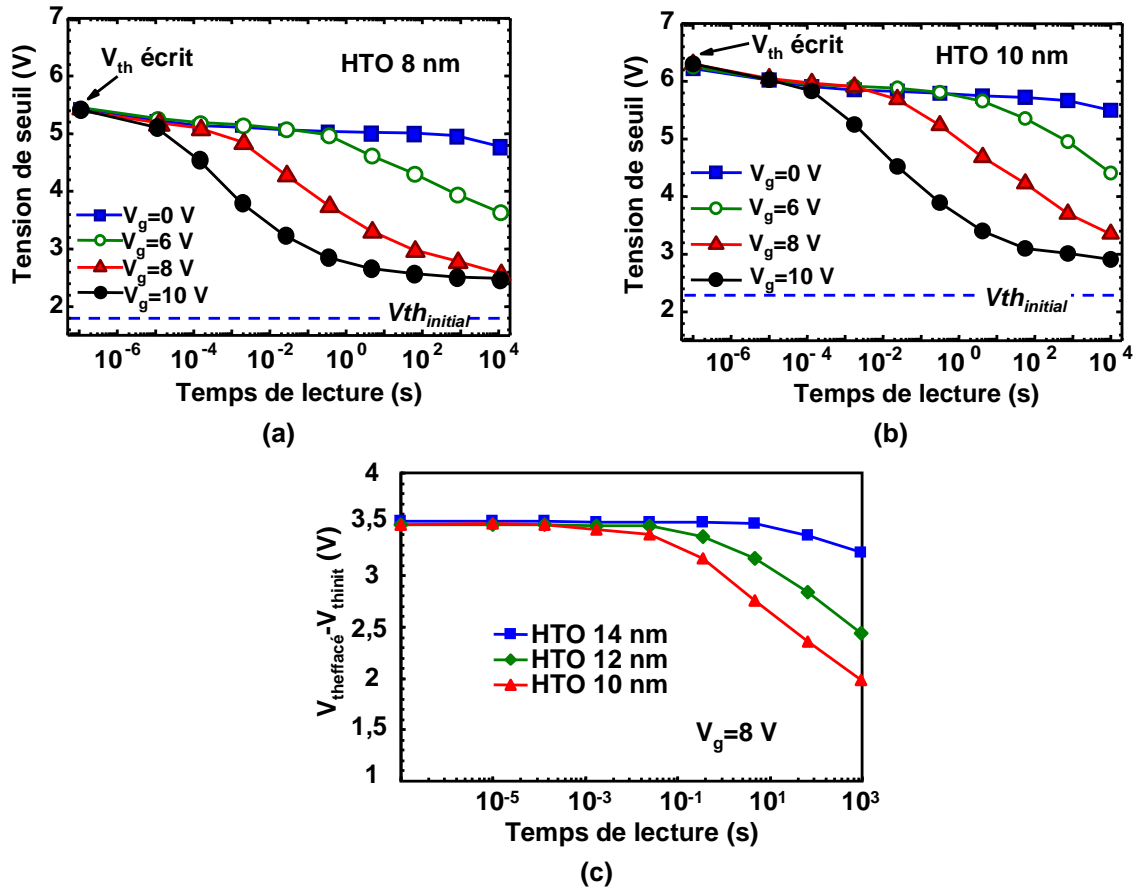


Fig. III-27 : Mesures de la tension de seuil pendant un "gate disturb" accéléré sur un état écrit pour différentes conditions de tension de grille ($V_d=V_d=V_b=0$ V) et une épaisseur de HTO de (a) 8nm et (b) 10 nm. (c) Comparaison de la diminution de la tension de seuil pour trois dispositifs avec des épaisseurs de HTO de 10, 12 et 14nm pour une condition de "gate disturb" fixée ($V_g=8$ V). Lecture : $V_d=1$ V et $I_d=0,3$ nA. (a), (b) : $t_{tun}=4$ nm, $\Phi_{dot}\approx 5$ nm, $R_{dot}\approx 0,15$. (c) : $t_{tun}=5$ nm, $\Phi_{dot}\approx 6,5$ nm, $R_{dot}\approx 0,25$.

III.3 Caractérisation électrique des matrices mémoires à nanocristaux de silicium

III.3.1 Introduction

Un démonstrateur Flash NOR 32 Mb à Si-NCs avec la périphérie CMOS haute et basse tension fonctionnelle a été réalisé sur la base du produit Atmel de technologie 130 nm présenté dans la chapitre II. Les Si-NCs ont été intégrés suivant le deuxième procédé de fabrication décrit dans le chapitre II, paragraphe II.2.3 : la croissance des oxydes de grille des transistors haute et basse tension est faite avant le dépôt des Si-NCs, la périphérie est SASTI (auto-alignée) et la mémoire non-SASTI.

Il est bon de rappeler que les algorithmes standard de programmation des produits Flash couramment utilisés dans l'industrie sont assez complexes. En général, pour l'écriture, on fixe les tensions et le temps ainsi qu'une valeur de la tension de seuil minimum que doivent atteindre tous les bits de la matrice. Un premier pulse est appliqué. Ensuite se déroule l'opération de vérification, pendant laquelle on contrôle si tous les bits ont atteint au moins la valeur minimum de la tension de seuil. Si ce n'est pas le cas, le même pulse est à nouveau appliqué et ainsi de suite jusqu'à ce que tous les bits aient atteint la valeur seuil. L'algorithme d'effacement, quant à lui est plus complexe. L'effacement se déroule en trois étapes : la pré-programmation, l'effacement pur et la programmation « douce » des bits les plus effacés. Dans les mémoires Flash standard, on sait qu'il est possible de sur-effacer les cellules. Avant de procéder à l'effacement, il faut donc s'assurer que tous les bits soient écrits, car sinon, la tension de seuil des cellules déjà effacées serait trop basse à l'issue de l'effacement. Il y a d'abord une étape de vérification, qui permet de repérer les bits effacés. Ensuite, la pré-programmation est effectuée uniquement sur les bits effacés. Cette étape prend entre 200 et 300 ms pour un produit de 32 Mb comme celui que nous avons réalisé ($V_g=8$ V, $V_d=3,75$ V, $V_s=0$ V, $V_b=-1,5$ V). La seconde étape, l'effacement pur, dure entre 50 et 100 ms. La troisième étape, la programmation « douce » consiste à écrire légèrement les bits les plus effacés. Pour cela, on effectue une écriture avec une très faible tension de grille ($V_g=1,75$ V, $V_d=4$ V, $V_s=0$ V, $V_b=-1,5$ V). Cette étape prend environ 50 ms. L'algorithme d'effacement a donc une durée totale comprise entre 300 ms et 450 ms.

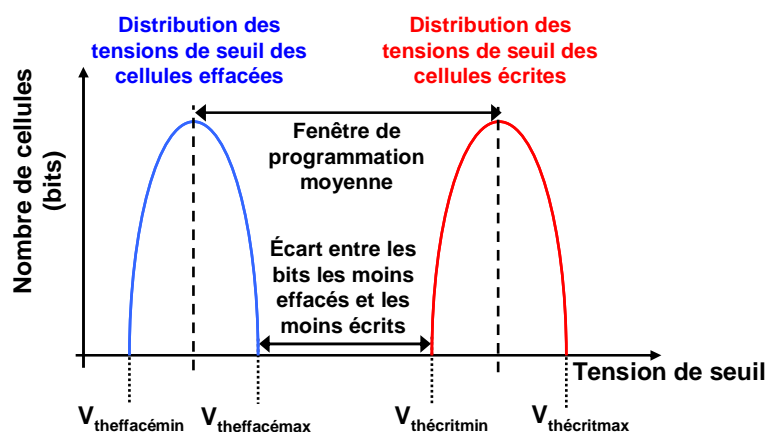


Fig. III-28 : Illustration des distributions des tensions de seuil écrites et effacées.

A noter que les algorithmes ici décrits sont issus de plusieurs années d'amélioration des mémoires Flash et ne peuvent donc pas s'appliquer tels quels à des nouvelles technologies tels que les nanocristaux de Silicium. Pour cela, il faut alors préciser qu'en ce qui concerne l'écriture et l'effacement du démonstrateur Flash NOR 32Mb à nanocristaux de Silicium, nous n'avons pas utilisé les algorithmes complexes du produit. Nous allons donc montrer par la suite des distributions « brutes » (**Fig. III-28**), où nous avons essayé de minimiser les corrections apportées par les algorithmes de programmation. Concernant l'écriture, les conditions sont fixées. Quant à l'effacement, les tensions sont fixées, seul le temps peut varier. En particulier, on fixe une valeur de la tension de seuil, et tant que tous les bits ne sont pas tous en-dessous de cette valeur, un nouveau pulse d'effacement avec un temps plus long est appliqué.

III.3.2 Influence du procédé de fabrication des nanocristaux de silicium

III.3.2.1 Influence de la taille des nanocristaux

Nous avons d'abord étudié l'influence de la taille des Si-NCs sur les distributions des tensions de seuil écrites et effacées. La **Fig. III-29** montre une comparaison des distributions effacées et écrites de quatre secteurs de 512 Kb avec des nanocristaux déposés en LPCVD deux étapes de différents diamètres mais avec la première étape de nucléation identique. Ce sont les mêmes échantillons que ceux présentés dans la Section III.2.2.3.B. On observe, comme on l'avait fait sur les cellules simples, que la fenêtre de programmation augmente lorsque le diamètre augmente. De plus, il apparaît clairement que plus le diamètre moyen des Si-NCs est élevé, plus les distributions écrites et effacées sont larges. Ceci est dû au fait que plus l'étape de croissance des nanocristaux est longue (i.e. le diamètre moyen des Si-NCs est grand), plus les phénomènes de coalescence ont lieu et donc plus la dispersion en taille est importante [**Perniola'03**].

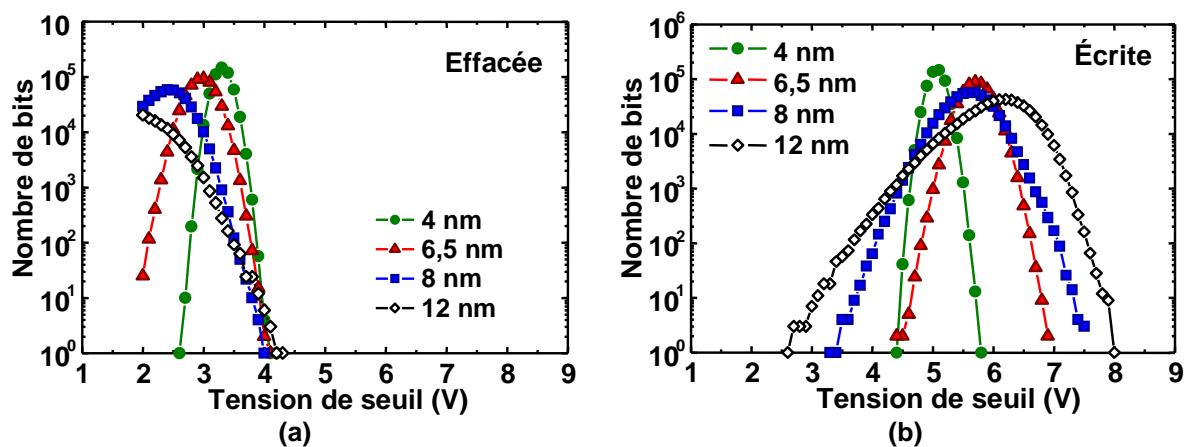


Fig. III-29 : Distributions des tensions de seuil de l'état effacé (a) ($V_g=-9,8$ V, $V_b=7,5$ V, $t=6$ ms) et écrit (b) ($V_g=9$ V, $V_d=4$ V, $V_s=0$ V, $V_b=-1,5$ V, $t=8$ μ s) de secteurs de 512 Kb pour des Si-NCs de diamètre croissant. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=4$ nm ; 6,5 nm ; 8 nm ; 12 nm.

III.3.2.2 Comparaison des procédés une étape/deux étapes

Nous avons également comparé le procédé de fabrication des Si-NCs une étape au procédé deux étapes, afin de vérifier l'influence du procédé de fabrication sur la dispersion des tensions de seuil. Nous avons représenté les distributions effacées et écrites d'un secteur pour les deux procédés de fabrication sur la **Fig. III-30**. On voit que les distributions correspondant aux Si-NCs fabriqués en une étape sont plus larges (environ 500 mV de plus pour la distribution écrite), ce qui confirme que les Si-NCs fabriqués avec le procédé en une étape sont plus dispersés en taille.

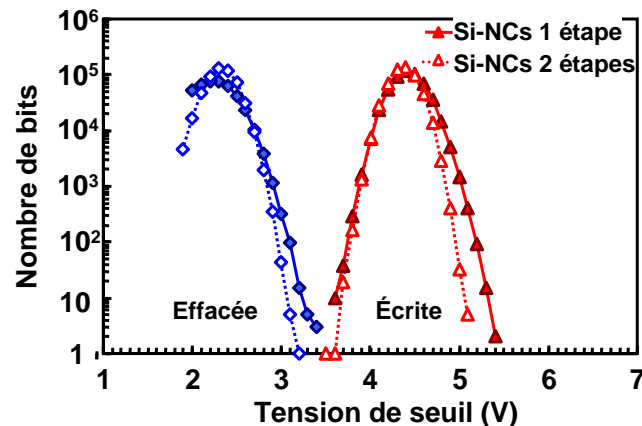


Fig. III-30 : Distributions des tensions de seuil de l'état écrit ($V_g=9$ V, $V_d=4$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=16$ μ s) et effacé ($V_g=-9,8$ V, $V_b=7,5$ V) d'un secteur 512 Kb pour les Si-NCs fabriqués avec le procédé 1 étape et le procédé deux étapes. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=8$ nm, $\Phi_{\text{dot}}=6,5$ nm, $R_{\text{dot}}=0,2$.

III.3.3 Influence des conditions de programmation

La **Fig. III-31** présente les distributions des tensions de seuil de l'état initial, effacé et programmé du produit 32 Mb (64 secteurs de 512 Kb). L'effacement et l'écriture ont été réalisés en utilisant les tensions internes générées par les pompes de charge du produit. On peut voir que la fenêtre de programmation moyenne est supérieure à 3 V. Par contre, l'écart entre les cellules les moins écrites et les cellules les moins effacées est assez faible (environ 500mV).

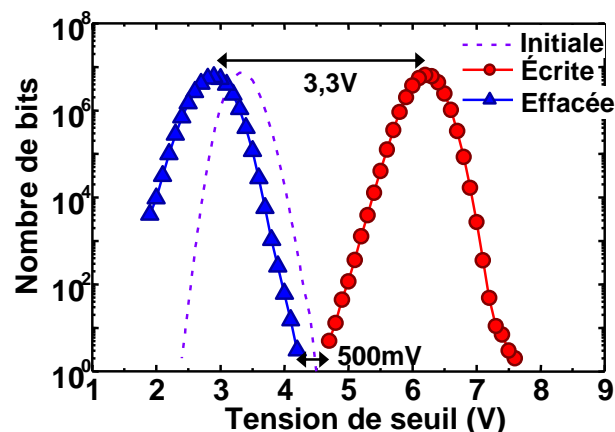


Fig. III-31 : Distributions des tensions de seuil de l'état initial, effacé ($V_g=-8,8$ V, $V_b=7,5$ V, $t=6$ ms) et écrit ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=12$ μ s) du démonstrateur 32 Mb (64 secteurs de 512 Kb). La lecture est effectuée à $V_d=1$ V et $I_d=4$ μ A. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=6,5$ nm, $R_{\text{dot}}=0,25$.

Par la suite, nous allons voir comment les conditions de programmation influencent les distributions des tensions de seuil et plus particulièrement comment on peut améliorer les caractéristiques électriques des matrices (augmenter la fenêtre de programmation, rendre les distributions plus étroites, augmenter l'écart entre les distributions écrites et effacées).

III.3.3.1 Amélioration de la distribution des tensions écrites

Une première solution pour augmenter l'écart entre les distributions effacées et écrites est de décaler la distribution écrite vers des tensions de seuil plus élevées. Nous allons voir plusieurs façons d'y parvenir dans la suite de ce paragraphe.

III.3.3.1.A Influence du temps d'écriture

Des secteurs de 512 Kb ont été écrits par porteurs chauds avec des temps cumulés compris entre 30 et 360 μ s. Les quatre échantillons correspondant aux quatre tailles de nanocristaux décrits dans le paragraphe précédent ont été testés. Les mesures sont représentées sur la **Fig. III-32**.

L'augmentation du temps d'écriture, permet tout d'abord d'augmenter la fenêtre de programmation moyenne (augmentation d'environ 700mV pour les temps considérés) et donc d'espacer les distributions effacées et écrites. On voit de plus que les distributions écrites deviennent plus étroites, surtout pour les Si-NCs de plus grande taille (**Fig. III-32(c)-(d)**).

Enfin, la **Fig. III-32(b)** (correspondant au même échantillon sur lequel les distributions de la 32Mb ont été mesurées, **Fig. III-31**) montre qu'avec un temps de 360 μ s, on peut obtenir une fenêtre de programmation moyenne de 4V, l'écart entre les bits les moins effacés et les moins écrits correspondant étant de 1,8 V.

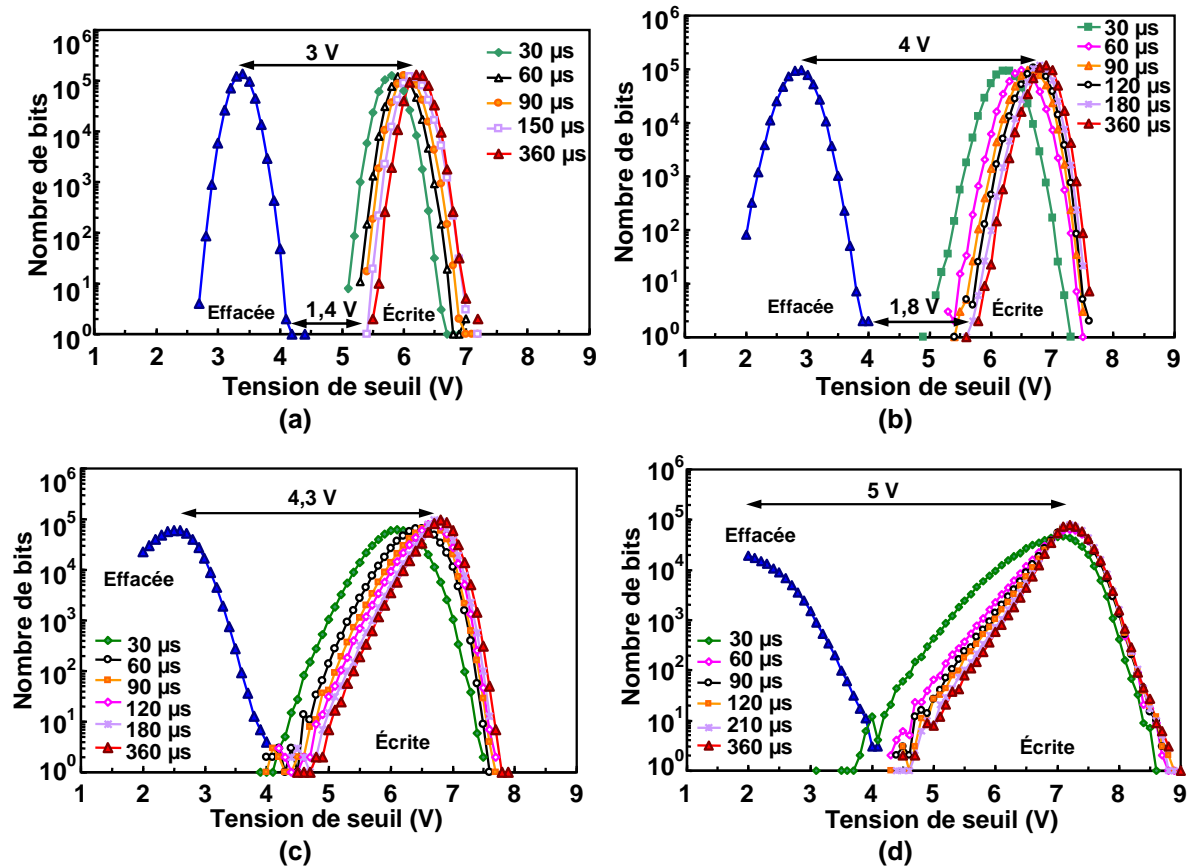


Fig. III-32 : Distributions des tensions de seuil de l'état écrit pour des temps de programmation croissants de $30 \mu\text{s}$ à $360 \mu\text{s}$ ($V_g=8 \text{ V}$, $V_b=3,75 \text{ V}$, $V_s=0 \text{ V}$, $V_b=-1,5 \text{ V}$) de secteurs de 512 Kb pour des Si-NCs de diamètre croissant. Conditions d'effacement : $V_g=-8,8 \text{ V}$, $V_b=7,5 \text{ V}$, $t=6 \text{ ms}$. La lecture de la tension de seuil est effectuée à $V_d=1 \text{ V}$ et $I_d=4 \mu\text{A}$. $t_{\text{un}}=5 \text{ nm}$, $t_{\text{HTO}}=10 \text{ nm}$, $\Phi_{\text{dot}}=4 \text{ nm}$ (a) ; 6,5 nm (b) ; 8 nm (c) ; 12 nm (d).

III.3.3.1.B Influence de la polarisation du substrat

Auparavant, nous avons montré que la polarisation négative du substrat permettait d'améliorer de façon significative la dynamique d'écriture en porteurs chauds des cellules mémoire unitaires. Cette démonstration a également été réalisée sur des secteurs 512 Kb du produit Flash NOR. Les résultats correspondant aux quatre tailles de Si-NCs sont représentés sur la **Fig. III-33**. Lorsque la polarisation du substrat varie de 0 à -1 V, la distribution écrite peut être décalée en moyenne jusqu'à 1,7 V. De plus, on remarque que la distribution écrite devient nettement plus étroite. Ce phénomène s'amplifie lorsque la taille des Si-NCs augmente, c'est-à-dire pour les distributions les plus larges. La polarisation négative du substrat pendant l'écriture par porteurs chauds joue donc un rôle très important dans l'optimisation du fonctionnement des matrices NOR à nanocristaux de silicium.

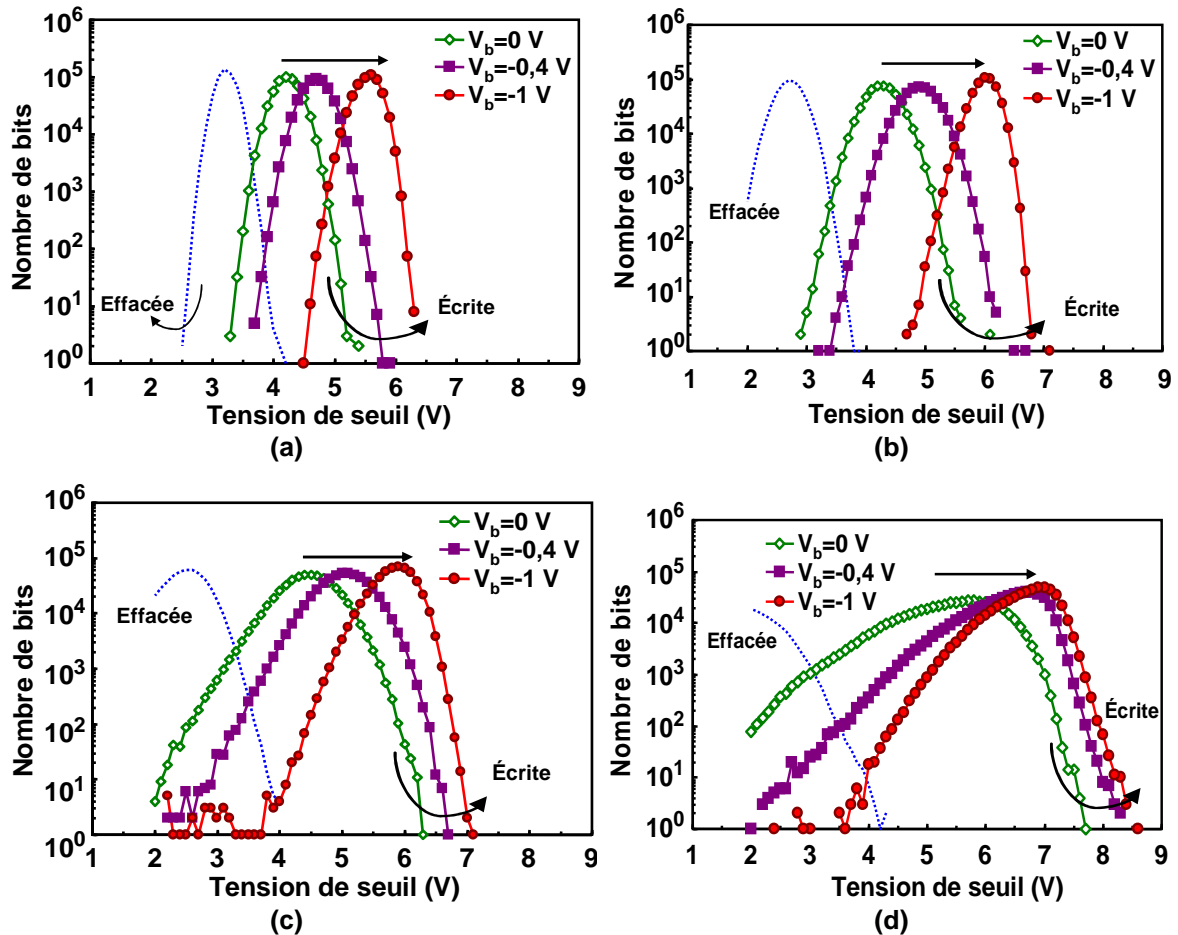


Fig. III-33 : Distributions des tensions de seuil de l'état écrit pour des tensions de substrat décroissantes de 0 V à -1 V ($V_g=8$ V, $V_d=3,75$ V, $V_s=0$ V, $t=12$ μ s) de secteurs de 512 Kb pour des Si-NCs de diamètre croissant. Conditions d'effacement : $V_g=-8,8$ V, $V_b=7,5$ V, $t=6$ ms. La lecture : $V_d=1$ V et $I_d=4$ μ A.

$t_{tun}=5$ nm, $t_{HTO}=10$ nm, $\Phi_{dot}=4$ nm (a) ; 6,5 nm (b) ; 8 nm (c) ; 12 nm (d).

III.3.3.2 Amélioration de la distribution des tensions effacées

Nous avons montré comment décaler la distribution écrite vers les tensions de seuil plus élevées. Il est également possible de décaler la distribution effacée vers des tensions de seuil plus faibles. Une possibilité est de jouer sur la valeur des tensions d'effacement.

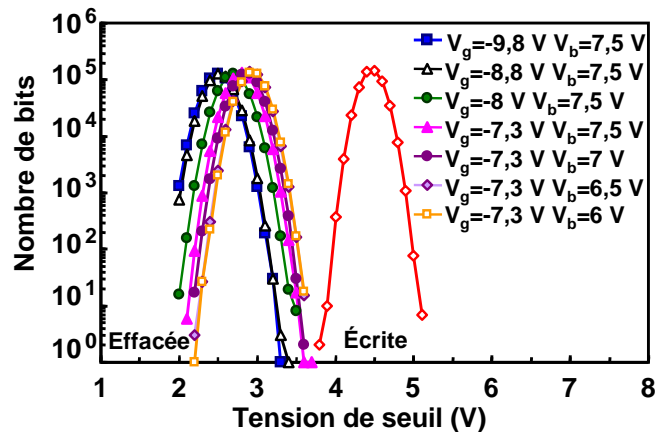


Fig. III-34 : Distributions des tensions de seuil de l'état effacé pour différentes tensions de stress ($t=40$ ms) d'un secteur de 512 Kb. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A. $t_{tun}=5$ nm, $t_{HTO}=8$ nm, $\Phi_{dot}=6,5$ nm, $R_{dot}=0,2$.

La **Fig. III-34** montre l'influence de la différence de potentiel grille-substrat sur la distribution effacée. En variant ($V_g - V_b$) de -13,8 V à -17,3 V, la distribution effacée est décalée d'environ 500mV, ce qui permet d'augmenter d'autant la fenêtre de programmation et en particulier, l'écart entre les bits les moins effacés et les moins écrits.

Toutefois, cette solution n'est pas aussi efficace que l'augmentation du temps d'écriture ou la polarisation négative du substrat pendant l'écriture. En effet, on a montré dans le paragraphe III.2.1.2 que la dynamique d'effacement de la cellule à Si-NCs saturait au niveau initial (vierge), les Si-NCs non dopés ne pouvant être sur-effacés. Si l'on se trouve près de la région de saturation, la tension de seuil ne peut donc être diminuée que faiblement. On peut toutefois souligner que cette saturation de la tension de seuil effacée à un avantage par rapport à la cellule à grille flottante continue. Nous avons expliqué précédemment que l'algorithme d'effacement d'une mémoire Flash standard comportait trois étapes : la vérification des bits effacés, leur pré-programmation et l'effacement pur. Or, les deux premières étapes sont rendues nécessaires par le sur-effacement de la flash standard qui amènerait les bits initialement effacés à avoir une tension de seuil beaucoup trop basse. Dans le cas des Si-NCs, peu importe l'état des bits avant l'effacement, leur tension de seuil ne peut pas descendre sous le niveau initial. L'algorithme d'effacement peut donc être réduit à la seule étape de l'effacement pur, permettant ainsi de réduire considérablement le temps total de l'effacement de la matrice mémoire.

III.4 Fiabilité

Des mesures d'endurance, de rétention en température et de gate disturb ont également été réalisées sur le démonstrateur à Si-NCs. En effet, les mesures statistiques sur matrices mémoires de grande capacité permettent de mieux rendre compte de la fiabilité des mémoires, car elles illustrent le comportement d'un grand nombre de cellules.

III.4.1 Endurance

Le test standard d'endurance a été effectué sur des secteurs avec des Si-NCs de 4 nm (**Fig. III-35(a)**) et des Si-NCs de 6,5 nm (**Fig. III-35(b)-(c)**). Le test d'endurance utilisé est celui du produit Flash standard. La phase d'écriture est réalisée dans des conditions (tensions et temps) fixes. L'algorithme d'effacement est dit « intelligent », car le temps d'effacement varie au cours de l'endurance afin d'obtenir toujours la même tension de seuil effacée. Cet algorithme « intelligent » est évidemment adapté aux mémoires Flash standard qui peuvent être sur-effacées. Cela permet ainsi de compenser le piégeage parasite dans les diélectriques de la cellule au cours de l'endurance. Avec les nanocristaux de silicium, comme nous avons pu le voir dans les sections concernant les cellules unitaires, les dynamiques d'effacement montrent une saturation de la tension de seuil à un niveau « neutre » (les Si-NCs n'étant pas dopés, ne peuvent pas être sur-effacés). Dans l'algorithme utilisé, il y a une valeur maximale du temps d'effacement (1500 s), au bout duquel le programme d'endurance s'arrête si les tensions de seuil des bits effacés ne sont pas toutes inférieures à la limite fixée. C'est ce qui

explique que le test de la **Fig. III-35(a)** ne dépasse pas les 700 cycles. Cette figure montre que la mesure d'endurance est plus critique avec des Si-NCs de plus petite taille, le taux de couverture étant plus faible. Au contraire, avec les Si-NCs de 6,5 nm, on parvient sans problème jusqu'au 10K cycles. Après 5K cycles, l'état effacé n'a pas dérivé (**Fig. III-35(b)**). Après 10K cycles d'écriture/effacement, on observe une légère augmentation de la tension de seuil des distributions écrites et effacées (**Fig. III-35(c)**), ce qui est cohérent avec les résultats obtenus sur cellules. Une solution pour minimiser le piégeage parasite dans les diélectriques serait de modifier les conditions de programmation afin de ne pas trop stresser les diélectriques tout en effaçant de manière suffisante.

A noter que, après 10K cycles, aucun bit erratique n'apparaît et les distributions ne s'élargissent pas, ce qui confirme la robustesse des Si-NCs aux défauts induits dans l'oxyde.

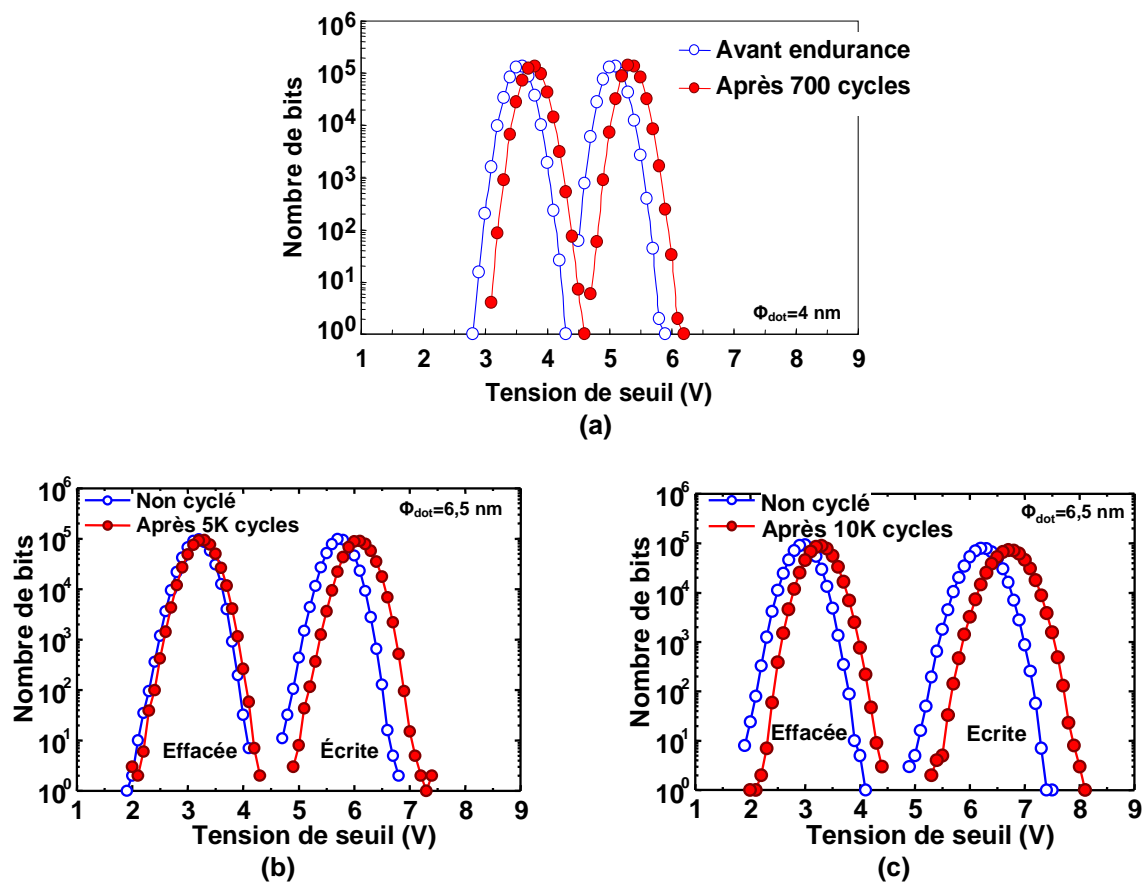


Fig. III-35 : Distributions des tensions de seuil de l'état écrit ((a), (b) : $V_g=9$ V, $V_d=4$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=8$ μ s, (c) : $V_g=8$ V, $V_d=3,75$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=8$ μ s) et effacé ((a), (b) : $V_g=-8$ V, $V_b=7,5$ V, (c) : $V_g=-8,8$ V, $V_b=7,5$ V) sur des secteurs de 512 Kb, mesurées avant et après endurance. (a) : $\Phi_{dot}=4$ nm, 700 cycles, (b) : $\Phi_{dot}=6,5$ nm, 5K cycles, (c) : $\Phi_{dot}=6,5$ nm, 10K cycles. $t_{tun}=5$ nm, $t_{HTO}=10$ nm. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A.

III.4.2 Réétention

Nous nous sommes ensuite intéressés à la réétention des mémoires à Si-NCs. Des mesures à 150°C ont d'abord été réalisées sur un secteur 512 Kb effacé (**Fig. III-36(a)**) et écrit (**Fig. III-36(b)**). Après 140 h, la distribution effacée reste inchangée et on observe une faible diminution de la tension de seuil écrite. La **Fig. III-36(c)** montre des mesures de réétention à 250°C sur l'état écrit d'une matrice 7Mb (14 secteurs de 512 Kb). Après 140 h, la diminution de la tension de seuil est identique à celle observée à 150°C. Entre 140 h et 300 h, la diminution est quasiment nulle. A noter aussi que dans toutes les mesures de réétention, aucun bit marginal ayant une perte de charge plus importante que les autres n'est observé.

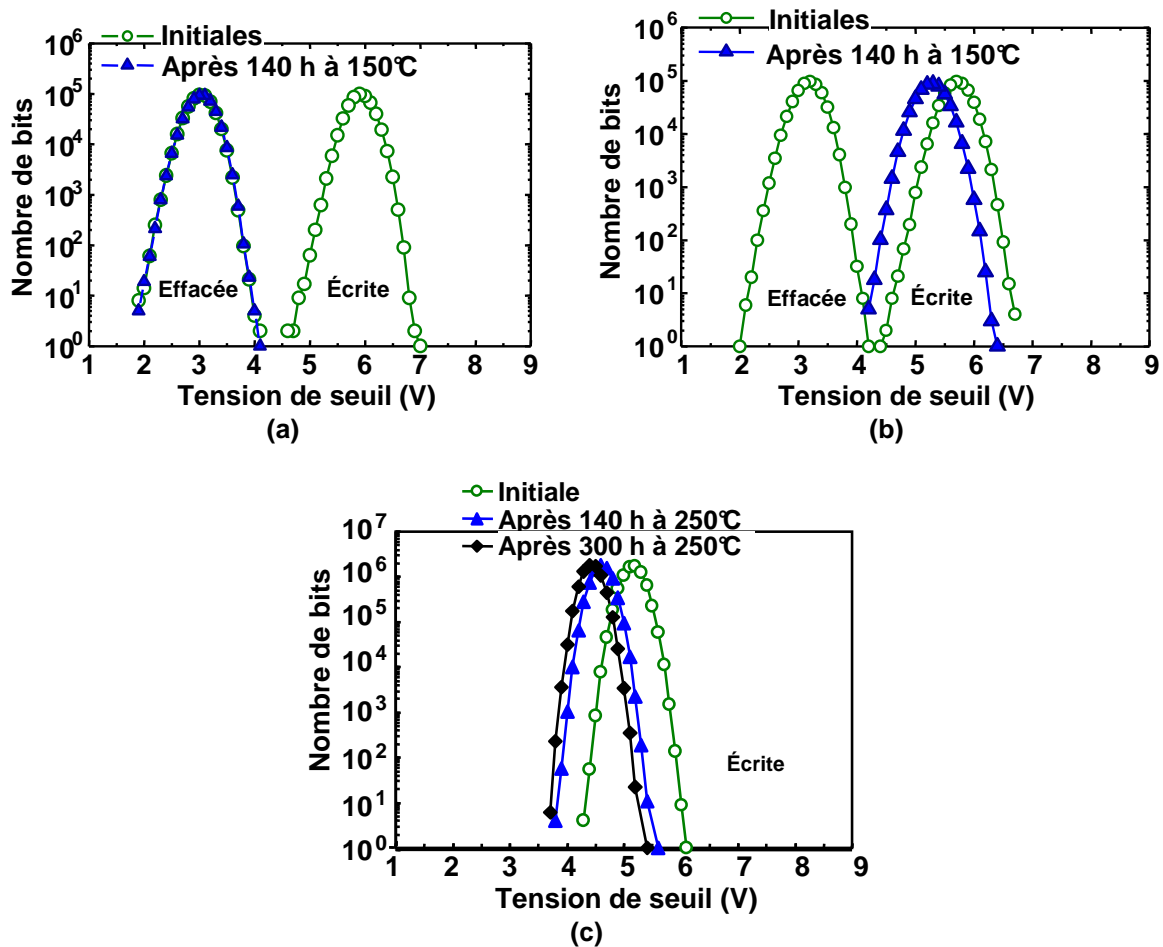


Fig. III-36 : Mesures de réétention à 150°C sur un état effacé (a) ($V_g=-8V$, $V_b=7,5 V$) et écrit (b) ($V_g=9 V$, $V_d=4 V$, $V_s=0 V$, $V_b=-1,5 V$, $t=8 \mu s$) et d'un secteur de 512 Kb et à 250°C sur un état écrit de 7 Mb (14 secteurs de 512 Kb) (c). (a), (b) : $t_{tun}=5 nm$, $t_{HTO}=10 nm$, $\Phi_{dot}=6,5 nm$, $R_{dot}=0,25$. (c) : $t_{tun}=5 nm$, $t_{HTO}=8 nm$, $\Phi_{dot}=6,5 nm$, $R_{dot}\approx 0,2$. La lecture de la tension de seuil est effectuée à $V_d=1 V$ et $I_d=4 \mu A$.

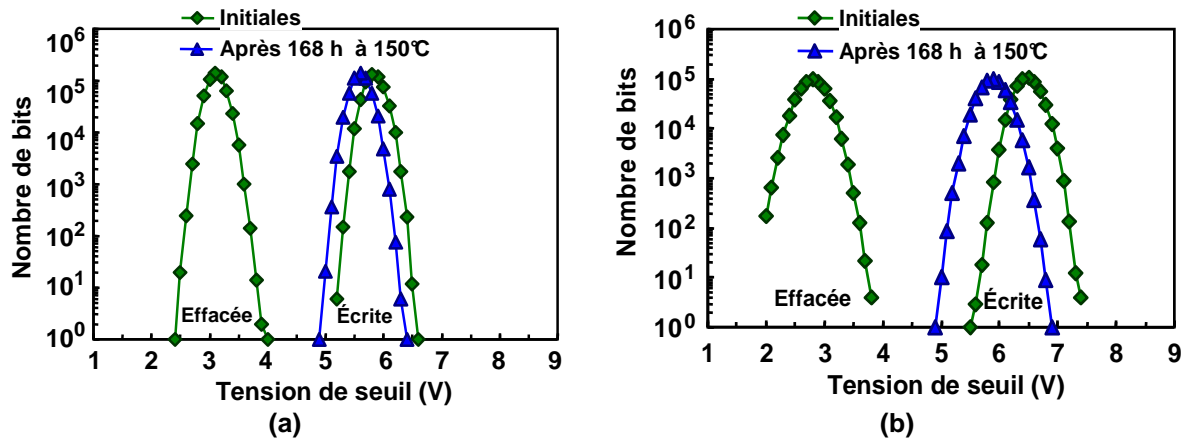


Fig. III-37 : Mesures de rétention à 150°C sur un état écrit élevé ($V_g=8$ V, $V_d=3,75$ V, $V_s=0$ V, $V_b=-1,5$ V, $t=150$ μ s) de secteurs de 512 Kb. Conditions d'effacement : $V_g=-8,8$ V, $V_b=7,5$ V, $t=6$ ms. $t_{\text{tun}}=5$ nm, $t_{\text{HTO}}=10$ nm, $\Phi_{\text{dot}}=4$ nm (a) et 6,5 nm (b). La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A.

Nous avons ensuite étudié l'influence de la valeur de la fenêtre de programmation sur la perte de charge. Des mesures de rétention à 150°C ont été effectuées sur des secteurs écrits avec un temps de 150 μ s (et correspondant à des Si-NCs de 4 et 6,5 nm de diamètre). Après 168 h à 150°C, on peut voir que la perte de charge sur les Si-NCs de 4 nm est très faible (~ 200 mV) (**Fig. III-37(a)**). L'écart entre les distributions écrites et effacées est encore de 1V et la fenêtre de programmation est presque de 3V. Dans le cas des Si-NCs de 6,5 nm, la fenêtre de programmation étant plus élevée, la perte après 168 h à 150°C est d'environ 500 mV (**Fig. III-37(b)**). Cependant, la fenêtre de programmation reste élevée (3,5 V) et l'écart entre les bits les moins effacés et les bits les moins écrits est encore d'environ 1V.

III.4.3 Rétention après endurance

Finalement, des tests de rétention après endurance ont été effectués sur des secteurs avec des Si-NCs de 4 et 6,5 nm de diamètre. En général, ce type de test permet de faire apparaître de problèmes de queues de distributions correspondant à des bits erratiques et donc de déceler les bits ayant un fonctionnement anormal.

Pour les plus petits Si-NCs, la rétention à 150°C après 140 h a été effectuée sur un secteur effacé **Fig. III-38(a)** et écrit **Fig. III-38(b)** après 700 cycles (l'algorithme d'endurance de la Flash standard ne permettant pas d'aller plus loin, comme on l'a expliqué dans le paragraphe précédent). On observe que la distribution effacée ne varie pratiquement pas. Quant à la distribution écrite, après l'augmentation de la tension de seuil au cours des cycles d'écriture/effacement, elle diminue ensuite pour revenir à l'état écrit avant endurance. De plus, aucun bit marginal n'est observé.

Ensuite, un secteur de 512 Kb avec des Si-NCs de plus grande taille a été testé. Dans ce cas, nous avons montré que l'on pouvait atteindre un plus grand nombre de cycles d'écriture/effacement. La rétention après endurance a été mesurée sur l'état écrit. Après 140 h à 150°C sur un secteur cyclé 5K fois, on observe une légère diminution de la tension de seuil écrite (~500 mV) (Fig. III-38(c)), identique à celle mesurée sur un secteur non cyclé (Fig. III-36(b)). On constate, encore une fois, qu'aucun bit marginal et aucun élargissement des distributions n'est observé, même après endurance puis rétention à 150°C. Ceci démontre la robustesse de dispositifs à Si-NCs contre les défauts générés par les stress d'écriture/effacement dans l'oxyde.

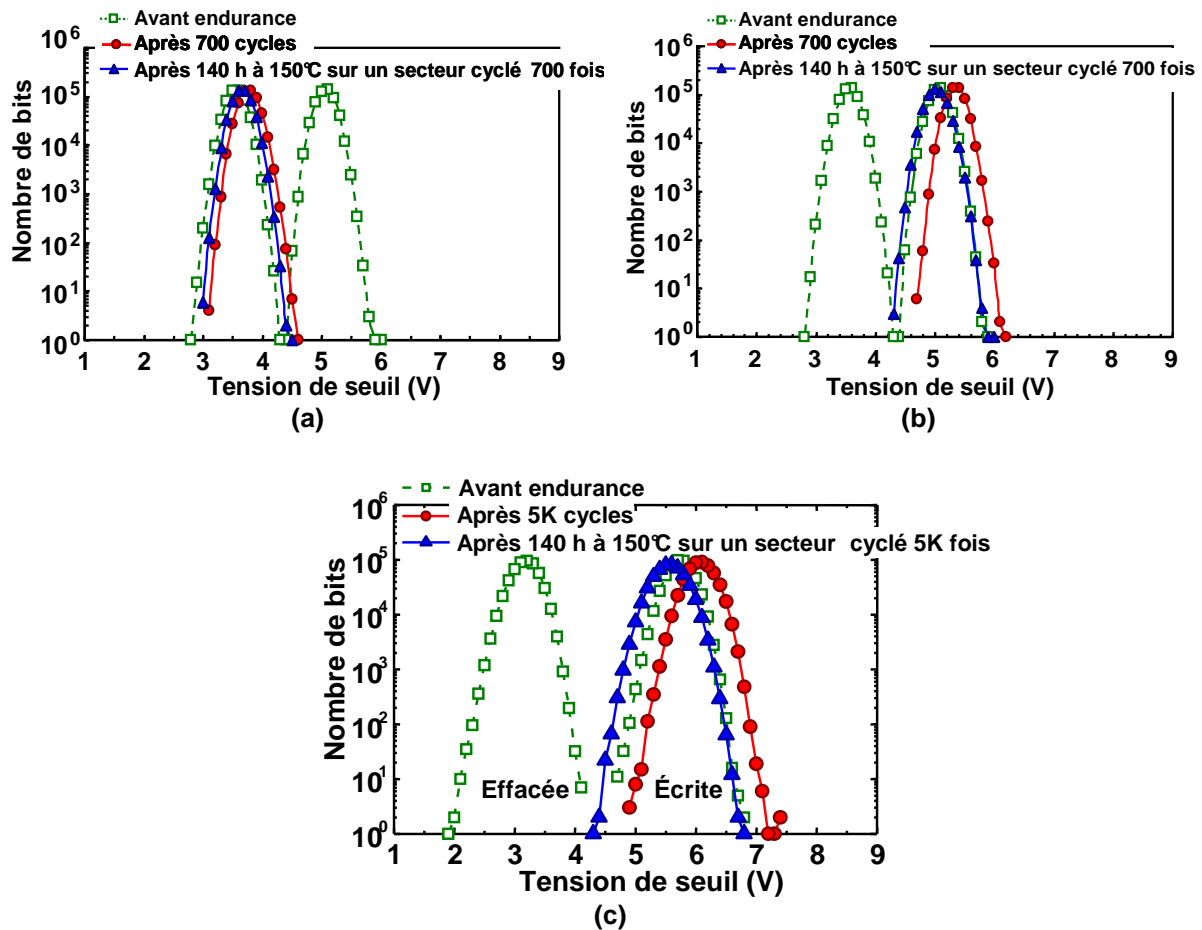


Fig. III-38 : Mesures de rétention à 150°C avant et après endurance sur un état effacé (a) ($V_g=-8V$, $V_b=7,5 V$) et écrit (b) (c) ($V_g=9 V$, $V_d=4 V$, $V_s=0 V$, $V_b=-1,5 V$, $t=8 \mu s$) de secteurs de 512 Kb. (a), (b) : $t_{\text{tun}}=5 \text{ nm}$, $t_{\text{HTO}}=10 \text{ nm}$, $\Phi_{\text{dot}}=4 \text{ nm}$, $R_{\text{dot}}=0,1$. (c) : $t_{\text{tun}}=5 \text{ nm}$, $t_{\text{HTO}}=10 \text{ nm}$, $\Phi_{\text{dot}}=6,5 \text{ nm}$, $R_{\text{dot}}=0,25$. La lecture de la tension de seuil est effectuée à $V_d=1 V$ et $I_d=4 \mu A$.

III.4.4 « Gate disturb »

Dans les sections précédentes, nous avons étudié la perturbation induite lors de la lecture de la cellule écrite avec un oxyde de contrôle de type HTO sur cellules mémoire unitaires. Nous avons mis en évidence le fait qu'il était important d'optimiser l'épaisseur du HTO afin de minimiser le « gate disturb » au cours de la lecture.

Dans cette partie, nous avons souhaité évaluer ce même phénomène sur des matrices NOR à Si-NCs. Pour cela, des secteurs de 512 Kb, avec des épaisseurs de HTO de 8 et 10 nm, ont été programmés dans des conditions standard. Les **Fig. III-39(a)-(b)** montrent que des lectures successives des secteurs écrits entraînent une diminution de la tension de seuil (quelques centaines de millivolts après 10 lectures). On observe de plus que cette diminution est plus faible pour le HTO d'épaisseur égal à 10 nm, conformément à ce qui a été démontré auparavant pour les cellules unitaires. Un secteur de 512 Kb avec un HTO de 10 nm a également été écrit avec un temps de programmation plus élevé de 360 μ s, puis des lectures successives ont été réalisées (**Fig. III-39(c)**). On constate une diminution de la tension de seuil plus importante.

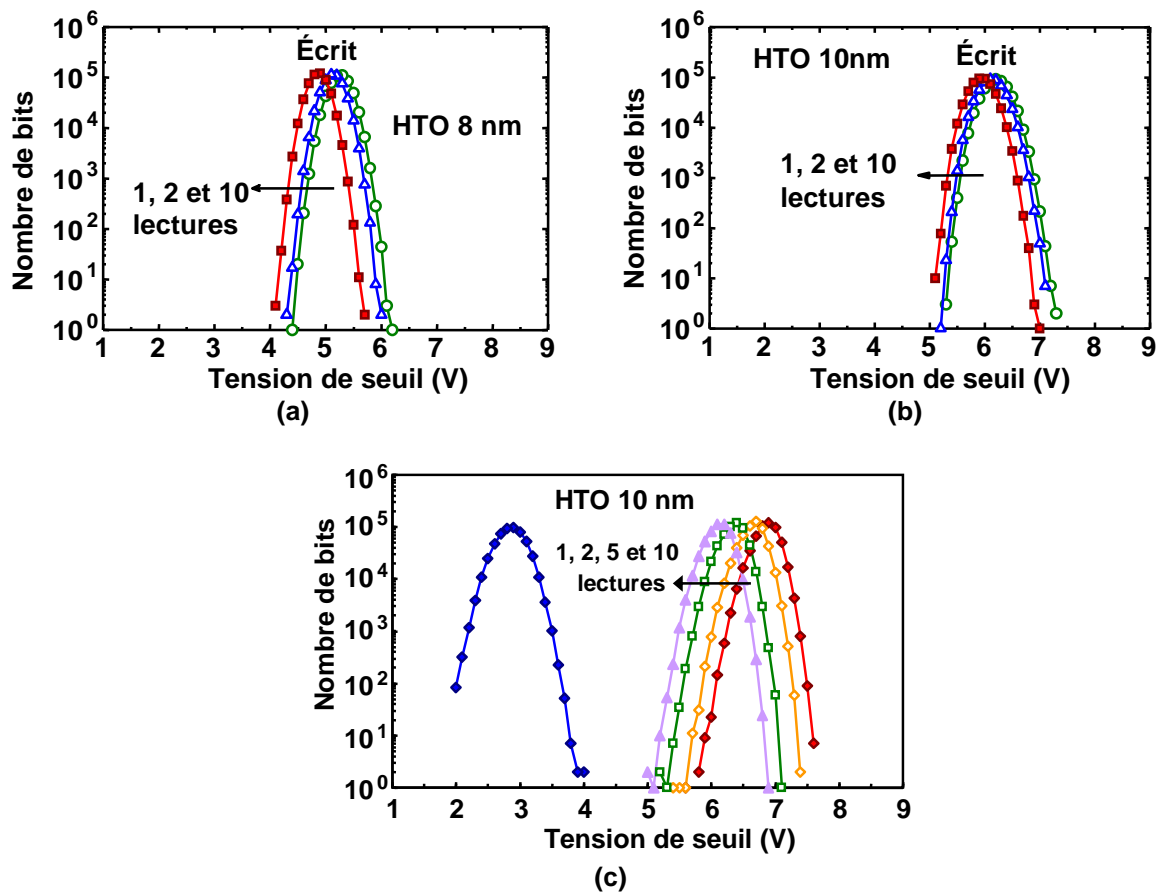


Fig. III-39 : (a), (b) : Distributions des tensions de seuil écrites ($V_g=9$ V, $V_d=4$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=8$ μ s) et effacées ($V_g=-9,8$ V, $V_b=7,5$ V) d'un secteur de 512 Kb après plusieurs lectures. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A. (c) Distributions des tensions de seuil écrites ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1.5$ V, $V_s=0$ V, $t=360$ μ s) et effacées ($V_g=-8,8$ V, $V_b=7,5$ V, 6ms) d'un secteur de 512 Kb après plusieurs lectures. La lecture de la tension de seuil est effectuée à $V_d=1$ V et $I_d=4$ μ A. (a), (b) : $t_{\text{tun}}=4$ nm, $\Phi_{\text{dot}}=5$ nm. (c) : $t_{\text{tun}}=5$ nm, $\Phi_{\text{dot}}=6,5$ nm.

Dans un deuxième temps, des mesures en mode DMA (Direct Access Memory) ont été effectuées sur un mot (16 bits) des deux secteurs précédents. La lecture consiste à mesurer le courant de drain des cellules pour une tension de grille donnée. Chaque lecture dure environ 20 ms. La **Fig. III-40(a)** représente le courant de drain moyen des 16 cellules écrites en fonction du nombre de lectures pour trois tensions de grille différentes 6 V, 7 V et 8 V et pour les deux épaisseurs de HTO. Il faut bien comprendre qu'une augmentation du courant de drain à tension de grille constante est équivalente à une diminution de la tension de seuil de la cellule et donc à une perte de charge (**Fig. III-40(b)**). On voit qu'à $V_g=6$ V, le courant de drain reste constant au cours des lectures. Par contre, à partir de $V_g=7$ V, on observe une augmentation du courant de drain avec le nombre de lectures, ce qui indique une perte de charge.

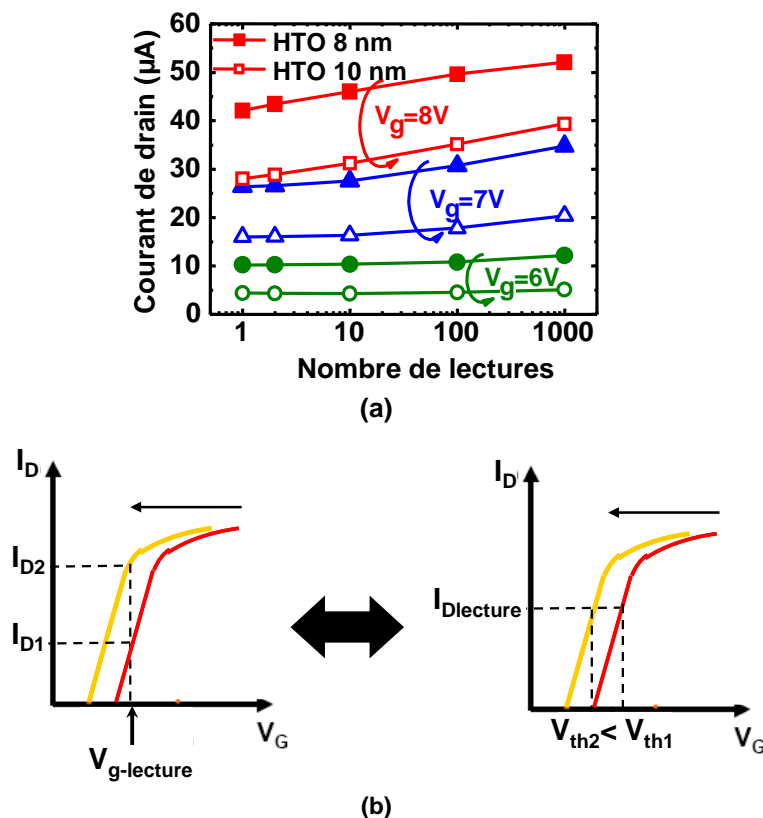


Fig. III-40 : (a) : Lecture en mode DMA (Direct Access Memory) : courant de drain moyen d'une cellule écrite d'un secteur en fonction du nombre de lectures et de la tension de grille. Le temps d'une lecture est d'environ 20 ms. $t_{\text{tun}}=5$ nm, $\Phi_{\text{dot}}\approx 5$ nm, $R_{\text{do}}\approx 0,15$. En mode DMA (Direct Access Memory), l'augmentation du courant de drain est équivalente à une diminution de la tension de seuil en mode standard (mesure de la tension de seuil à un courant de drain donné) (b).

III.5 Conclusion

Dans ce chapitre, nous nous sommes intéressés à la caractérisation électrique des mémoires à nanocristaux de silicium. Nous avons d'abord présenté les résultats électriques obtenus sur des cellules mémoires simples. Une étude exhaustive de l'influence des méthodes de programmation ainsi que des paramètres technologiques sur les caractéristiques d'écriture par électrons chauds et d'effacement par Fowler-Nordheim a été réalisée. Concernant l'effacement, on a montré que l'on observait une saturation de la tension de seuil (au niveau de l'état vierge), contrairement à la cellule Flash classique à grille flottante continue qui peut être sur-effacée. De plus, on a constaté que l'effacement des dispositifs à Si-NCs pouvait être beaucoup plus rapide que celui de la Flash standard.

On a montré que l'on pouvait obtenir une fenêtre de programmation de 4 V avec un temps d'écriture de 10 μ s en utilisant des conditions d'écriture adéquates. Plusieurs paramètres technologiques permettent d'améliorer les performances électriques de la cellule (**Tableau III-2**). L'écriture par porteurs chauds est influencée par l'implant canal et la taille (taux de couverture) des nanocristaux. L'épaisseur de l'oxyde tunnel et du diélectrique de grille HTO ainsi que la taille des Si-NCs jouent sur la vitesse d'effacement. Nous avons également montré que l'utilisation d'un ONO comme diélectrique de grille conduisait à un piégeage parasite important dans le nitrure, rendant plus appropriée l'utilisation d'un HTO en tant que diélectrique couvrant les nanocristaux pour l'application NOR. Par ailleurs, l'épaisseur du HTO doit être suffisante pour éviter les problèmes de perte de charge au cours de la lecture d'une cellule écrite (« gate disturb »).

Dans une deuxième partie, nous nous sommes intéressés aux caractéristiques électriques de matrices mémoires. Un démonstrateur ATMEL Flash NOR 32 Mb avec la logique CMOS de périphérie fonctionnelle a été réalisé. Deux paramètres sont essentiels : la fenêtre moyenne de programmation et l'écart entre les bits les moins effacés et les moins écrits (qui doit être suffisant pour distinguer l'état écrit de l'état effacé). Une fenêtre de programmation de plus de 3 V a été obtenue dans les conditions de programmation standard du produit Flash. Il est possible d'augmenter cette fenêtre jusqu'à 4 V et surtout de séparer suffisamment les distributions effacées et écrites en modifiant les conditions de programmation. Le procédé de fabrication des nanocristaux et surtout leur taille permettent également de rendre les distributions des tensions de seuil plus étroites. On peut également souligner que l'un des avantages des Si-NCs est la réduction de l'algorithme d'effacement (pré-programmation, effacement, programmation douce) à une seule étape (effacement), permettant ainsi de réduire considérablement le temps total d'effacement de la matrice mémoire. Enfin, une étude de fiabilité a été réalisée, démontrant la robustesse des dispositifs à Si-NCs contre les défauts générés par les stress d'écriture/effacement dans l'oxyde.

↗ Fenêtre de programmation		Endurance	Rétention	↘ Gate disturb	↘ Largeur des distributions
Écriture		Effacement			
↗ Niveau de saturation	↗ Vitesse	↗ Vitesse			
•↗ R_{dot}	•↗ Dose implant canal •↗ R_{dot}	•↘ R_{dot} •↘ épaisseur HTO	•↗ R_{dot} •Densification HTO	•↗ t_{tun}	•↘ R_{dot}
R_{dot} min ≈ 10 %			t_{tun} min = 5 nm	t_{HTO} min = 12 nm	R_{dot} max ≈ 25 %

Tableau III-2 : Récapitulatif de l'influence des paramètres technologiques sur les performances électriques de la mémoire à Si-NCs.

Bibliographie

- [DeSalvo'01] B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, B. Guillaumot "Experimental and theoretical investigation of nanocrystals and nitride-trap memory devices", IEEE Trans. on El. Dev., Vol. 48, No. 8, p. 1789, 2001.
- [Esseni'00] D. Esseni, L. Selmi, A. Ghetti, E. Sangiorgi, "Injection Efficiency of CHISEL Gate Currents in Short MOS Devices: Physical Mechanisms, Device Implications, and Sensitivity to Technological Parameters", IEEE Trans. on El. Dev., vol. 47, no. 11, p. 2194, 2000.
- [Gerardi'07] C. Gerardi, V. Ancarani, R. Portoghese, S. Giuffrida, M. Bileci, G. Bimbo, O. Brafa, D. Mello, G. Ammendola, E. Tripiciano, R. Puglisi, S. A. Lombardo, "Nanocrystal Memory Cell Integration in a Stand-Alone 16-Mb NOR Flash Device", IEEE Trans. On Elec. Dev., Vol. 54, No. 6, p. 1376, 2007.
- [Mazen'03] F. Mazen, T. Baron, R. Truche, A. M. Papon and J. M. Hartmann, "A two steps CVD process for the growth of Silicon nano-crystals", Applied Surface Science, vol. 214, pp.359-353, 2003.
- [Perniola'03] L.. Perniola, B. D Salvo, G. Ghibaudo, A. Foglio Para, G. Pananakakis, T. Baron, S. Lombardo, "Influence of dots size and dots number fluctuations on the electrical characteristics of multi-nanocrystal memory devices", Solid-State Electronics, vol. 47, pp. 1637–1640, 2003.
- [Steimle'04] R.F. Steimle, R.Rao, C.T. Swift, K. Harber, S. Straub, R. Muralidhar, B. Hradsky, J.A. Yater, E.J. Prinz, W. Paulson, M.Sadd, C. Parikh, S.G.H. Anderson, T. Huynh, B. Acred, L. Grieve, M. Rossow, R. Mora, B. Darlington, Ko-Min Chang, B.E.White Jr., "Integration of Silicon Nanocrystals into a 6V 4Mb Nonvolatile Memory Array", NVSMW Proc., pp. 73-74, 2004.

Chapitre IV : Modélisation des mémoires Flash à nanocristaux de silicium

Chapitre IV : Modélisation des mémoires Flash à nanocristaux de silicium.....	151
IV.1 Introduction.....	155
IV.2 Le modèle de la quasi-grille flottante	155
<i>IV.2.1 Résultats des simulations avec le modèle de la quasi-grille flottante</i>	158
IV.2.1.1 Effacement Fowler-Nordheim	158
IV.2.1.1.A Influence de l'épaisseur de l'oxyde tunnel.....	158
IV.2.1.1.B Influence de l'épaisseur du HTO de contrôle	161
IV.2.1.1.C Influence du taux de couverture des nanocristaux de silicium	163
IV.2.1.1.D Comparaison avec la Flash standard.....	164
IV.2.1.2 « Gate disturb »	165
IV.3 Ecriture par porteurs chauds	169
<i>IV.3.1 Etude de la localisation de la charge dans une cellule mémoire à nanocristaux de silicium</i>	169
IV.3.1.1 Introduction.....	169
IV.3.1.2 Dispositifs d'étude.....	170
IV.3.1.3 Simulations électrostatiques	171
IV.3.1.4 Simulations dynamiques de l'écriture par électrons chauds et interprétation des données expérimentales.....	172
IV.3.1.4.A Simulations TCAD.....	172
IV.3.1.4.B Modélisation analytique.....	177
IV.4 Conclusion.....	182
ANNEXE 1 : Mécanismes de conduction tunnel Fowler-Nordheim et tunnel direct.....	184
ANNEXE 2 : Quelques éléments de la théorie des porteurs chauds	189
Bibliographie.....	192

IV.1 Introduction

Dans ce chapitre, nous allons décrire la modélisation des dispositifs mémoires à nanocristaux de silicium que nous avons effectuée afin de comprendre et d'interpréter les résultats électriques présentés dans le chapitre III. La première partie traitera du modèle de la quasi-grille flottante [DeSalvo'01], qui permet de simuler l'effacement Fowler-Nordheim et le « gate disturb » des cellules. Dans la deuxième partie, nous nous intéresserons à l'écriture par porteurs chauds et en particulier à la localisation de la charge injectée, qui est une caractéristique des mémoires à sites de piégeage discrets. Pour cela, des simulations TCAD, ainsi qu'un modèle analytique couplé à des mesures électriques seront présentés.

IV.2 Le modèle de la quasi-grille flottante

La dynamique de chargement/déchargement des dispositifs mémoires à nanocristaux de silicium peut être modélisée par la théorie de la quasi-grille flottante [DeSalvo'01]. Cette approche est basée sur la continuité du courant [Frohman'69] et sur les équations de base des mémoires à grille flottante [Groeseneken'98]. Le modèle capacitif traditionnel de la Fig. IV-1 est notamment utilisé.

La charge piégée dans le nanocrystal (Q_{FG}) est déterminée par la balance entre les courants d'entrée et de sortie, respectivement à travers l'oxyde tunnel et l'oxyde de contrôle (Fig. IV-1), grâce à l'équation :

$$\frac{dQ_{FG}}{dt} = J_{out} - J_{in} \quad \text{Équation IV-1}$$

Avec J_{in} : courant de remplissage du nanocrystal depuis le substrat et la grille – J_{out} : courant de vidage du nanocrystal vers le substrat et la grille.

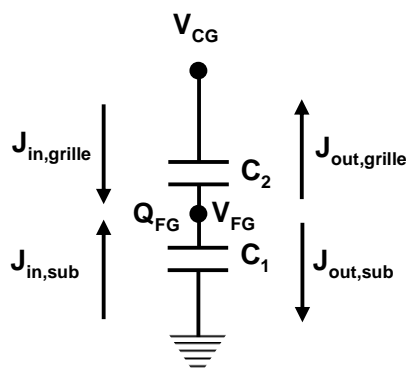


Fig. IV-1 : Modèle capacitif d'une cellule mémoire utilisé dans l'approche de la quasi-grille flottante.

Le courant de remplissage J_{in} peut être décomposé en deux flux : $J_{in,sub}$, le flux d'électrons émis depuis le substrat traversant l'oxyde tunnel et $J_{in,grille}$ le flux d'électrons injectés depuis la grille de contrôle traversant l'oxyde de contrôle. On peut décomposer de la même manière le courant de vidage J_{out} en $J_{out,sub}$ et $J_{out,grille}$, les flux d'électrons partant du nanocrystal et traversant l'oxyde tunnel et l'oxyde de contrôle, respectivement.

Ces courants sont de type tunnel Fowler-Nordheim et tunnel direct et sont calculés à partir de la transparence dont l'expression est la suivante (voir l'Annexe 1 pour le calcul détaillé) :

$$Tr(E, V_{ox}) = \exp\left(-\frac{4\sqrt{2m_{ox}}}{3q\hbar} \frac{t_{ox}}{V_{ox}} \left[(\Phi_{SiO_2} - E)^{3/2} - H(\Phi_{SiO_2} - V_{ox} - E) \cdot (\Phi_{SiO_2} - V_{ox} - E)^{3/2} \right]\right) \quad \text{Équation IV-2}$$

Avec Φ_{SiO_2} : barrière de potentiel Si/SiO₂ (en eV) - E : énergie électronique d'injection (en eV) - V_{ox} : chute de potentiel à travers l'oxyde considéré (en eV), $t_{ox}=t_{tun}$: épaisseur de l'oxyde tunnel - m_{ox} : masse effective de l'électron dans l'oxyde - \hbar : constante de Planck réduite - q : charge de l'électron - H : fonction de Heaviside.

Dans un premier temps, le courant $J_{in,sub}$ peut être évalué en utilisant l'expression calculée dans l'Annexe 1 [O'Dwyer'73] :

$$J_{in,sub} = \frac{4\pi q m_{Si} kT}{h^3} \int_0^{\Phi_{SiO_2}} Tr(E, V_{FG}) \cdot \ln\left[1 + \exp\left(-\frac{E - E_{Fsub}}{kT}\right)\right] dE \quad \text{Équation IV-3}$$

Avec E_{Fsub} : niveau de Fermi du substrat - m_{Si} : masse effective de l'électron dans le silicium - T : température (en K), k : constante de Boltzmann.

Le courant $J_{in,grille}$ peut être exprimé de manière similaire avec les paramètres appropriés.

Tout comme les deux composantes du courant J_{out} :

$$J_{out,sub} = \frac{4\pi q m_{Si} kT}{h^3} \int_0^{\Phi_{SiO_2}} Tr(E, -V_{FG}) \cdot \ln\left[1 + \exp\left(-\frac{E - E_{Fdot}}{kT}\right)\right] dE \quad \text{Équation IV-4}$$

et

$$J_{out,grille} = \frac{4\pi q m_{Si} kT}{h^3} \int_0^{\Phi_{SiO_2}} Tr(E, V_{CG} - V_{FG}) \cdot \ln\left[1 + \exp\left(-\frac{E - E_{Fdot}}{kT}\right)\right] dE \quad \text{Équation IV-5}$$

Avec E_{Fdot} : niveau de Fermi dans le nanocrystal - V_{CG} : potentiel de la grille de contrôle.

Il faut préciser que dans le cas du courant des Si-NCs vers le substrat ou la grille, l'énergie électronique E a pour niveau de référence la bande de conduction du nanocrystal et non pas celle du substrat comme dans le cas du courant entrant.

Les équations précédentes peuvent se simplifier si l'on applique l'approximation du courant à basse température [Schuegraf'92] :

$$J(V_{ox}) = A \left(\frac{V_{ox}}{T}\right)^2 Tr(V_{ox}) \quad \text{Équation IV-6}$$

Avec $A = q^3 m_{Si} / (16\pi^2 \hbar m_{ox} \Phi_{SiO_2})$

Il est possible de prendre en compte la charge d'inversion dans le canal. On obtient ainsi la formule ci-dessous [Weinberg'77] :

$$J_{in,sub} = f_{sub} \cdot Q_{inv} \cdot Tr(0, V_{FG}) \quad \text{Équation IV-7}$$

Q_{inv} : valeur absolue de la charge d'inversion - f_{sub} : fréquence d'échappement ($\approx 10^{13}$ Hz [Weinberg'77]).

Concernant le courant partant des Si-NCs J_{out} , il est également possible de prendre en compte le fait que les Si-NCs ne soient pas dopés. En effet, dans ce cas, la statistique de Boltzmann peut s'appliquer afin de simplifier l'expression du courant J_{out} [Stratton'69] :

$$J_{out} \approx \frac{4\pi q^3 m_{Si} (kT)^2}{h^3} \exp\left(\frac{E_{Fdot}}{kT}\right) \cdot [Tr(kT, -V_{FG}) + Tr(kT, V_{CG} - V_{FG})] \quad \text{Équation IV-8}$$

De plus, en posant dans l'Équation IV-8 $\exp(E_{Fdot} / kT) = n_{dot} / N_C$, $n_{dot} = |Q_{FG}| / (q \cdot \Phi_{dot})$ et ensuite en définissant $f_{dot} = 4\pi q^2 m_{Si} (kT)^2 / h^3 \Phi_{dot} N_C$ qui peut être considérée comme la fréquence d'échappement du Si-NC (N_C est la densité d'état effective dans la bande de conduction du silicium, Q_{FG} la concentration d'électrons dans le nanocrystal par unité de surface, Φ_{dot} le diamètre du Si-NC) on obtient :

$$J_{out} = f_{dot} \cdot |Q_{FG}| [Tr(kT, -V_{FG}) + Tr(kT, V_{CG} - V_{FG})] \quad \text{Équation IV-9}$$

Après le calcul des courants, la deuxième étape consiste à exprimer le potentiel des Si-NCs V_{FG} . Pour cela, on utilise le modèle capacitif classique de la Fig. IV-1 qui donne [DeSalvo'00] [Groeseneken'98] :

$$V_{FG} = \frac{C_2}{C_1 + C_2} \cdot V_{CG} + \frac{Q_{FG}}{C_1 + C_2} \quad \text{Équation IV-10}$$

Avec C_1 et C_2 les capacités de couplage substrat/Si-NC et grille de contrôle/Si-NC, respectivement.

La dernière étape consiste au calcul de la charge dans les Si-NCs. Cette approche est dite « quasi-grille flottante » car elle prend en compte le fait que la grille flottante n'est pas continue, mais discrète, en introduisant le facteur R_{dot} , qui représente la portion de surface couverte par les Si-NCs. Ce facteur introduit une pondération dans l'influence des charges piégées sur la tension de seuil de la cellule :

$$V_{th} = V_{th_{init}} - \frac{Q_{FG} R_{dot}}{C_2} \quad \text{Équation IV-11}$$

Le produit $Q_{FG} R_{dot}$ tient compte de la densité de charge effective. Les charges ne sont pas considérées comme donnant lieu à une densité discrètement répartie au-dessus du canal du transistor, mais comme une densité de charge uniformément répartie au-dessus du canal. Cette expression trouve sa justification théorique à partir de l'analyse des fluctuations du courant de drain au cours du chargement dans des pièges, comme cela a été fait précédemment pour la modélisation du signal RTS (Random Telegraph Signal) [Roux'92].

IV.2.1 Résultats des simulations avec le modèle de la quasi-grille flottante

IV.2.1.1 Effacement Fowler-Nordheim

Nous avons d'abord réalisé des simulations des dynamiques d'effacement de la cellule mémoire à nanocristaux de silicium afin de mieux comprendre les mesures présentées dans le chapitre III. Ces modélisations ont permis de cerner l'influence des épaisseurs de l'oxyde tunnel et de contrôle, ainsi que l'influence du taux de couverture des Si-NCs. Enfin, nous avons comparé les dynamiques d'effacement d'une Flash standard et d'une mémoire à Si-NCs.

IV.2.1.1.A Influence de l'épaisseur de l'oxyde tunnel

Nous allons ici étudier l'influence de l'épaisseur de l'oxyde tunnel sur l'effacement. Pour cela, nous avons représenté les diagrammes de bande obtenus avec le modèle de la quasi-grille flottante pour trois épaisseurs d'oxydes tunnel (4 nm, 5 nm et 6 nm) à $t=0$ (**Fig. IV-2(a)**) et à $t=10\mu\text{s}$ (**Fig. IV-2(b)**). Les courants dans l'oxyde tunnel (**Fig. IV-2(c)**) et dans le HTO (**Fig. IV-2(d)**) au cours de l'effacement sont également représentés.

Le fait de diminuer l'épaisseur de l'oxyde tunnel se traduit par une augmentation du champ électrique initial dans celui-ci (**Fig. IV-2(a)**), et donc une augmentation du courant à travers l'oxyde tunnel au début de l'effacement (**Fig. IV-2(c)**), d'où un effacement plus rapide. Les Si-NCs se vident de leur électrons, le courant J_{tun} diminue et à partir d'un certain temps ($\sim 10\mu\text{s}$ ici), on observe que le courant dans l'oxyde tunnel est identique quelque soit l'épaisseur de l'oxyde tunnel (le champ électrique dans l'oxyde tunnel étant le même **Fig. IV-2(b)**). Si on observe les diagrammes de bande après $10\mu\text{s}$ d'effacement, on remarque que la diminution de l'épaisseur de l'oxyde tunnel entraîne également une augmentation du champ électrique dans le HTO. Ceci induit une augmentation du courant dans le HTO qui permet d'atteindre l'équilibre entre les courants J_{tun} et J_{HTO} plus rapidement. Il faut noter que le courant J_{HTO} reste toutefois très inférieur au courant J_{tun} jusqu'aux temps les plus élevés et que l'on a donc bien un effacement des Si-NCs.

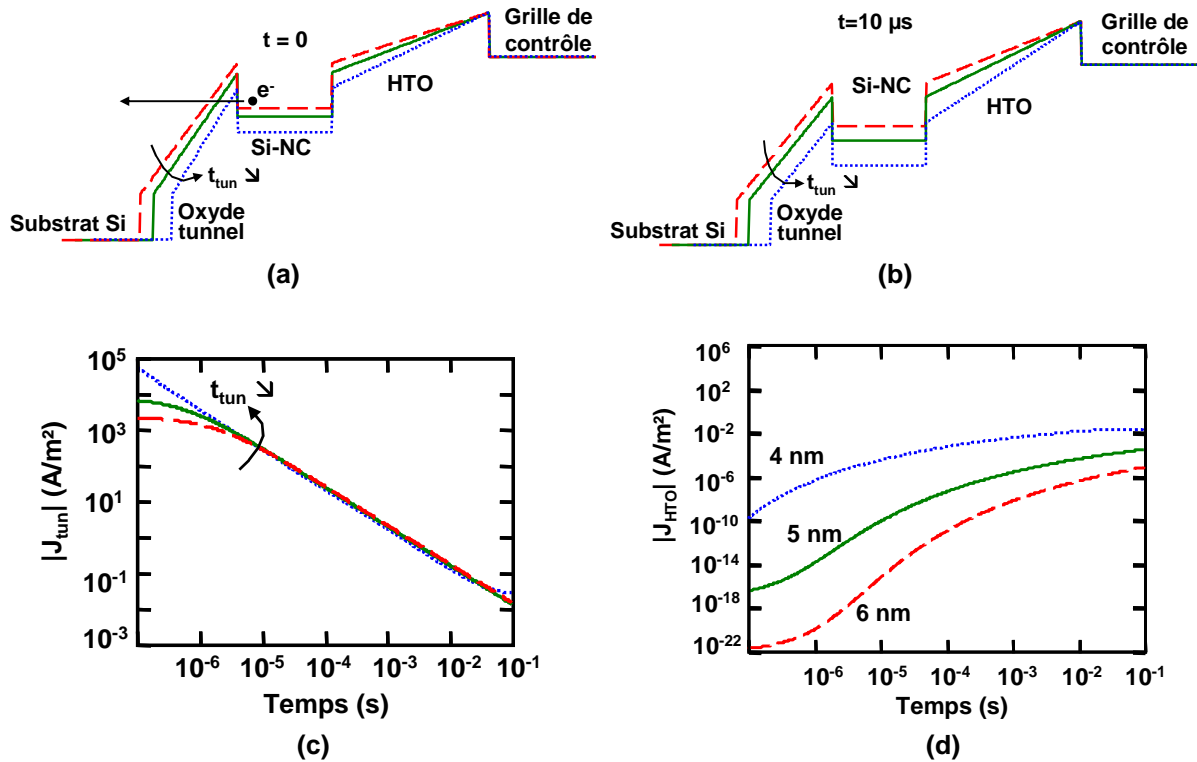


Fig. IV-2 : Diagramme des bandes de conduction en énergie de l'empilement de grille avec $V_g = -14$ V, à $t=0$ (a) et $t=10 \mu s$ (b), pour trois épaisseurs d'oxyde tunnel (4 nm, 5 nm et 6 nm). Pour des raisons de clarté, nous avons fait l'approximation d'un potentiel constant dans les nanocristaux dans les diagrammes de bande. Courants au cours de l'effacement dans l'oxyde tunnel (c) et dans le HTO de contrôle (d) pour les trois épaisseurs d'oxyde tunnel.

Dans le chapitre III, nous avons montré que diminuer l'épaisseur de l'oxyde tunnel permettait d'augmenter la vitesse d'effacement en mode Fowler-Nordheim. Nous avons effectuées les simulations correspondant aux mesures montrées dans le chapitre III. Les **Fig. IV-3(a)-(b)** et **Fig. IV-3(c)-(d)** montrent la comparaison mesures-simulations pour différentes épaisseurs nominales d'oxyde tunnel (4 nm/5 nm et 5 nm/6 nm). On peut voir qu'un bon accord est obtenu entre les simulations et les mesures. On observe bien un effacement plus rapide pour les oxydes tunnel plus fins.

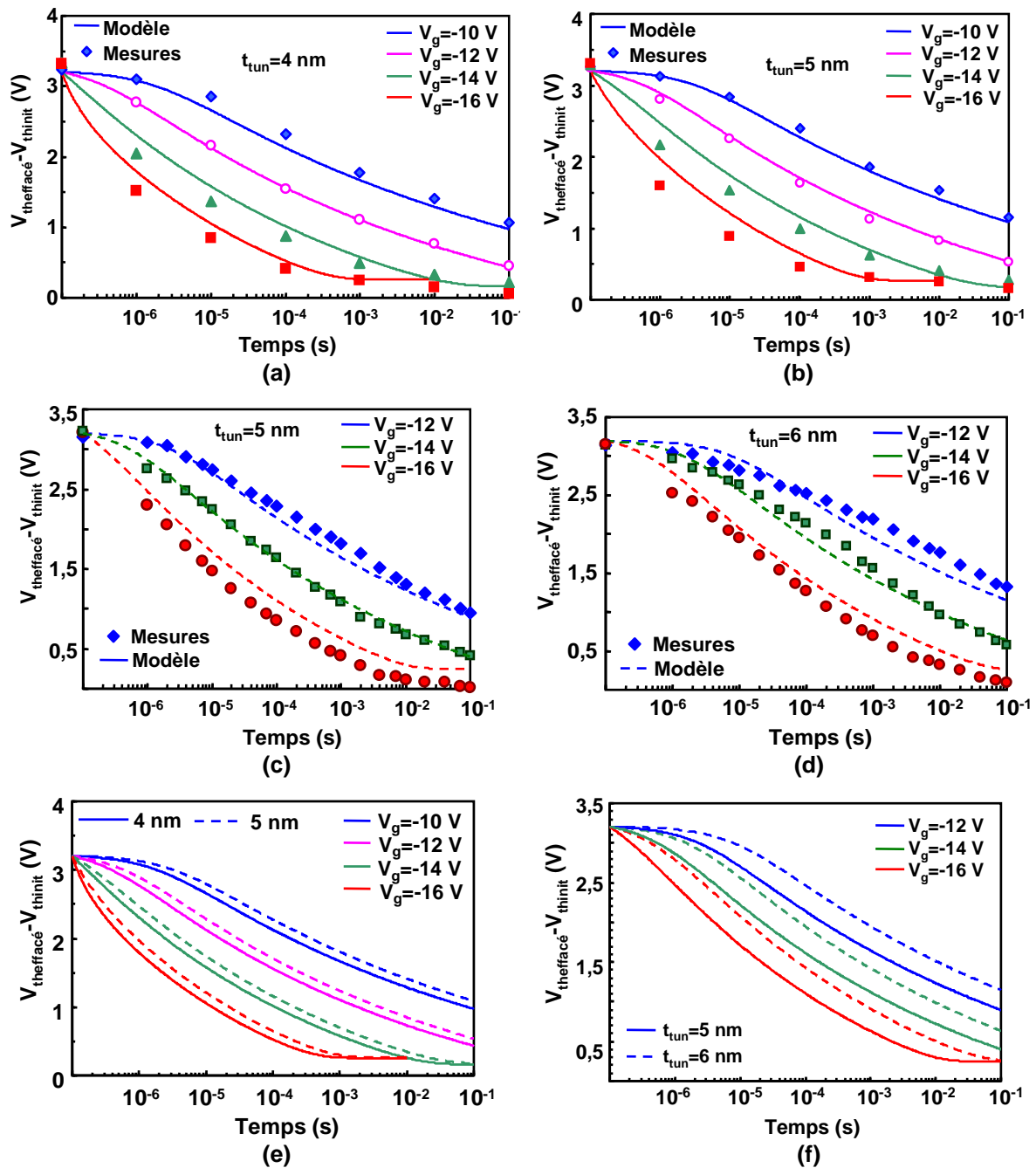


Fig. IV-3 : Comparaison des dynamiques d’effacement en Fowler-Nordheim mesurées et simulées avec le modèle de la quasi-grille flottante pour différentes épaisseurs d’oxyde tunnel (épaisseurs nominales des dispositifs : 4 nm (a), 5 nm (b)-(c) et 6 nm (d) ((a)/(b) et (c)/(d) correspondent à deux lots différents) ; épaisseurs choisies pour les simulations : (a) 4,2 nm, (b) 4,5 nm ; (c) 5,4 nm, (d) 6,2 nm –et différentes tensions de grille d’effacement. Dynamiques d’effacement simulées pour différentes épaisseurs nominales d’oxyde tunnel (4 nm et 5 nm (e), 5 nm et 6 nm (f)).

(a), (b), (e) : $t_{HTO}=10$ nm, $\Phi_{dot}=5$ nm, $R_{dot}\approx 0,15$. (c), (d), (f) : $t_{HTO}=10$ nm, $\Phi_{dot}=6,5$ nm, $R_{dot}\approx 0,25$.

IV.2.1.1.B Influence de l'épaisseur du HTO de contrôle

Comme précédemment, nous avons représenté les diagrammes des bandes de conduction de la cellule mémoire obtenus avec le modèle de la quasi-grille flottante au cours de l'effacement, pour trois épaisseurs de HTO (10 nm, 12 nm et 14 nm) à $t=0$ (**Fig. IV-5(a)**) et à $t=100\mu\text{s}$ (**Fig. IV-5 (b)**), ainsi que les courants dans l'oxyde tunnel (**Fig. IV-5 (c)**) et dans le HTO (**Fig. IV-5 (d)**).

On sait que la diminution de l'épaisseur du diélectrique de contrôle permet d'augmenter le couplage entre les Si-NCs et la grille de contrôle. A $t=0$, l'augmentation du couplage permet d'augmenter le champ électrique dans l'oxyde tunnel (**Fig. IV-5(a)**). Le courant dans l'oxyde tunnel est donc plus élevé au début de l'effacement (**Fig. IV-5(c)**), le champ dans le HTO étant le même pour les différentes épaisseurs de HTO (**Fig. IV-5(a)**). Les Si-NCs intégrés avec le HTO le plus fin vont donc se décharger plus rapidement. A partir d'un certain temps ($\sim 100\mu\text{s}$ ici), on constate que le potentiel dans les Si-NCs est identique et que le courant J_{tun} atteint une valeur identique quelque soit l'épaisseur de l'oxyde HTO (**Fig. IV-5(b)**). La diminution de l'épaisseur du HTO se traduit alors par une augmentation du champ électrique dans le HTO (**Fig. IV-5 (b)**). Ceci induit une augmentation du courant dans le HTO qui permet d'atteindre l'équilibre entre les courants J_{tun} et J_{HTO} plus rapidement (**Fig. IV-5 (d)**). Il faut noter encore une fois, que le courant J_{HTO} reste toutefois négligeable par rapport au courant J_{tun} jusqu'aux temps les plus élevés et que l'on a donc bien un effacement des Si-NCs.

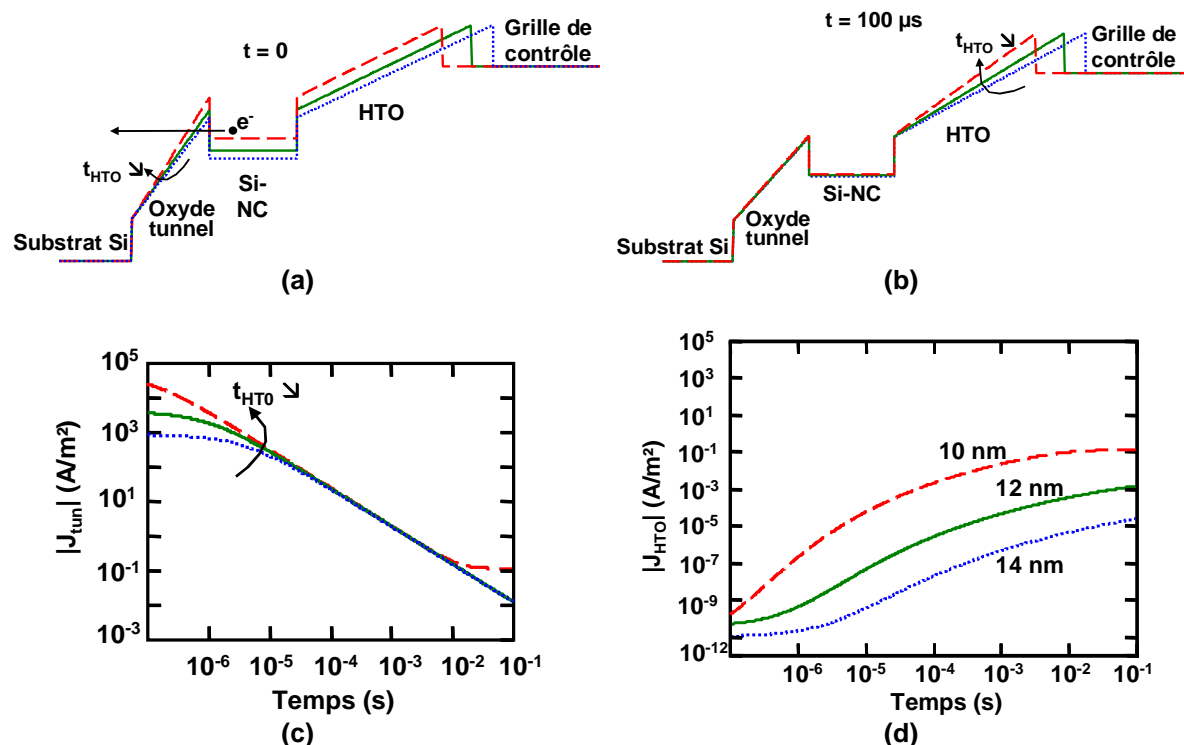


Fig. IV-4 : Diagramme des bandes de conduction en énergie des l'empilement de grille de la cellule mémoire correspondant à $V_g=-16\text{ V}$, $t=0$ (a) et $t=100\mu\text{s}$ (b), pour trois épaisseurs de HTO (10 nm, 12 nm et 14 nm). Courants au cours de l'effacement dans l'oxyde tunnel (c) et dans le HTO de contrôle (d) pour les trois épaisseurs de HTO.

Nous avons également simulé l'effacement Fowler-Nordheim pour trois épaisseurs de HTO différentes (10 nm, 12 nm et 14 nm). Les **Fig. IV-5(a)-(b)-(c)** montrent une comparaison entre simulations et mesures. Encore une fois le modèle de la quasi-grille flottante permet de reproduire correctement les dynamiques d'effacement pour les trois épaisseurs de HTO et pour des tensions de grille différentes. Les caractéristiques simulées **Fig. IV-5(d)-(e)** montrent bien que plus l'épaisseur de HTO est faible, plus l'effacement est rapide.

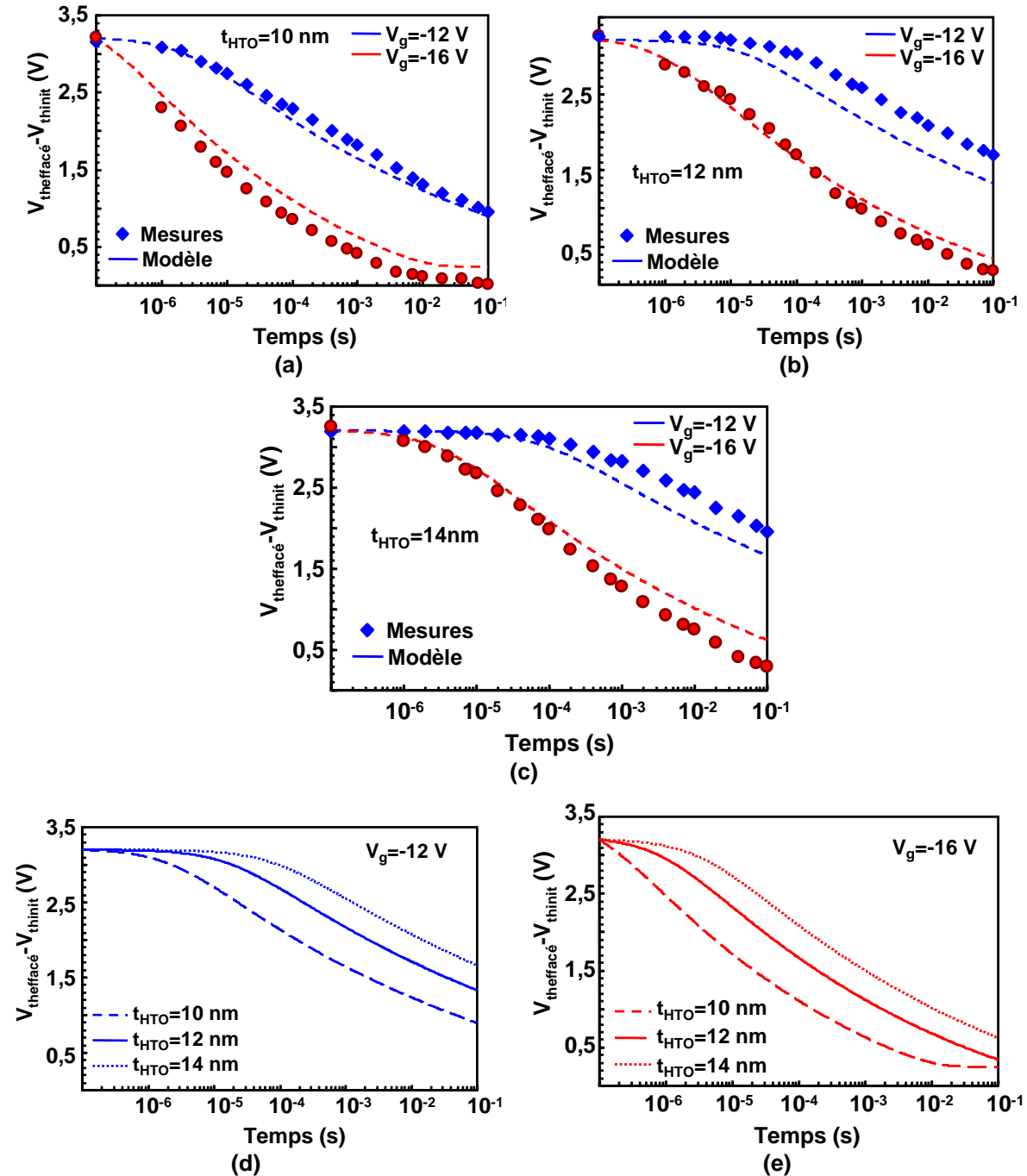


Fig. IV-5 : Comparaison des dynamiques d'effacement en Fowler-Nordheim mesurées et simulées (avec le modèle de la quasi-grille flottante) pour différentes épaisseurs de HTO (10 nm (a), 12 nm (b) et 14 nm (c)) et différentes tensions de grille d'effacement. Dynamiques d'effacement simulées pour différentes épaisseurs de HTO (10 nm, 12 nm et 14 nm) et différentes tensions de grille d'effacement (-12 V (d) et -16 V (e)). $t_{\text{un}} \approx 5 \text{ ns}$, $\Phi_{\text{dot}}=6.5 \text{ nm}$, $R_{\text{dot}} \approx 0.25$.

IV.2.1.1.C Influence du taux de couverture des nanocristaux de silicium

Dans le Chapitre III, nous avons montré que plus les Si-NCs étaient petits, plus l'effacement était rapide. Comme le montre la **Fig. IV-6**, ce phénomène peut être correctement simulé grâce au modèle de la quasi-grille flottante.

Afin de comprendre l'influence de la taille des Si-NCs, nous avons représenté dans un premier temps sur la **Fig. IV-7** le diagramme des bandes (à $t=0$) correspondant à la **Fig. IV-6(c)**. Il faut préciser encore une fois que nous avons pris en compte l'augmentation de la taille des Si-NCs par une augmentation du taux de couverture R_{dot} . On voit clairement que pour les Si-NCs de plus petit diamètre, le champ électrique dans l'oxyde tunnel est plus élevé, ce qui permet un effacement plus rapide.

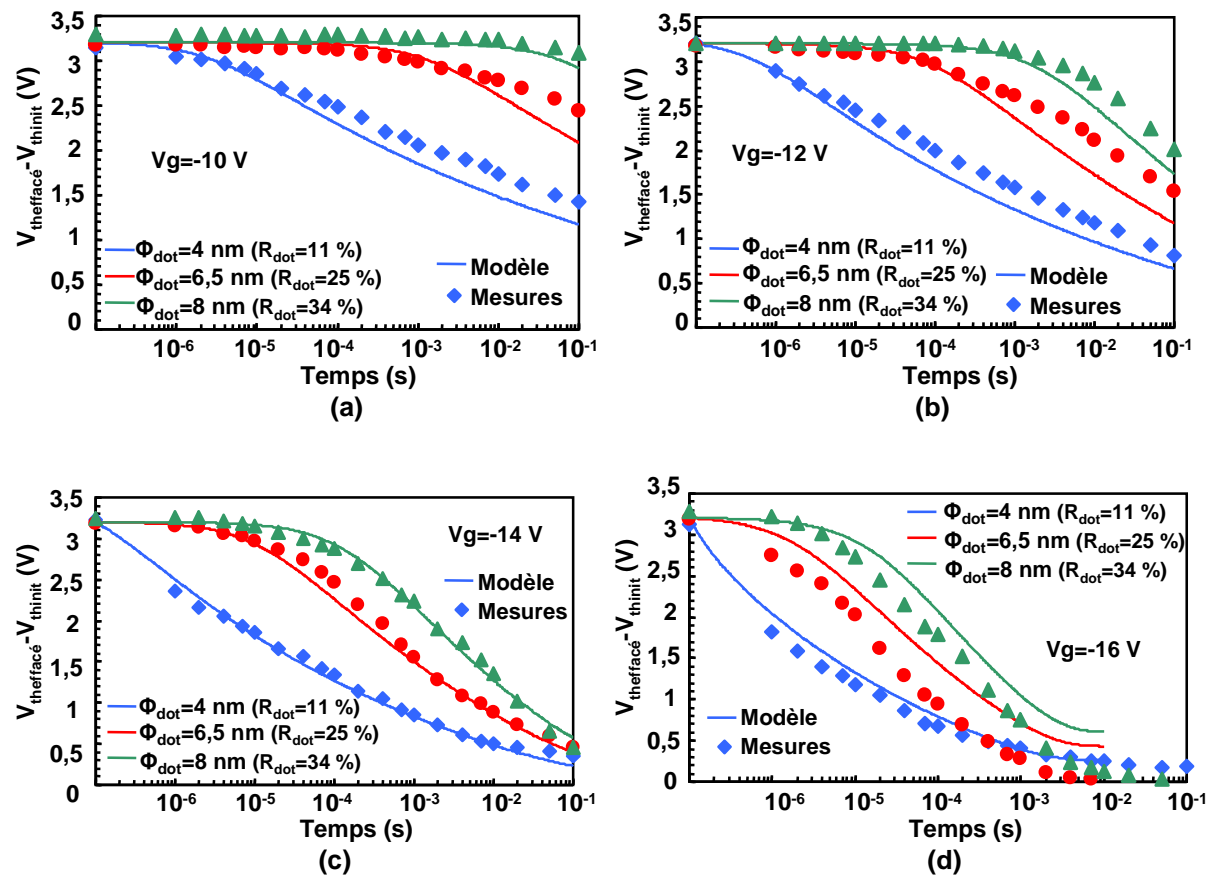


Fig. IV-6 : Comparaison des dynamiques d'effacement en Fowler-Nordheim mesurées et simulées (avec le modèle de la quasi-grille flottante) pour différentes tailles de Si-NCs (4 nm ($R_{\text{dot}}=11\%$), 6,5 nm ($R_{\text{dot}}=25\%$) et 8 nm ($R_{\text{dot}}=34\%$)) et différentes tensions de grille d'effacement (-10 V (a), -12 V (b), -14 V (c) et -16 V (d)). $t_{\text{tun}} \approx 5$ nm, $t_{\text{HTO}} \approx 10$ nm.

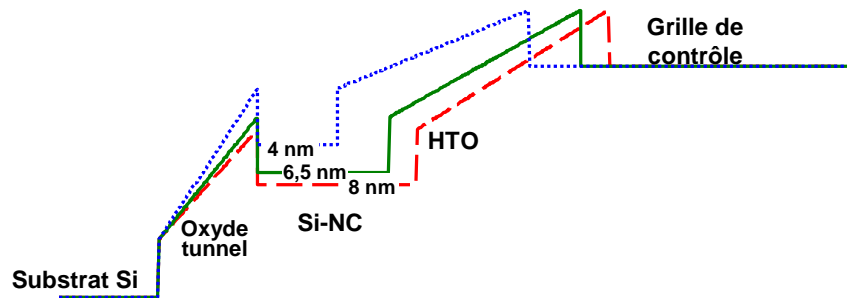


Fig. IV-7 : Diagramme des bandes de conduction de l'empilement de grille à $t=0$ et $V_g=-14$ V pour trois tailles de Si-NCs (4 nm, 6,5 nm et 8 nm), correspondant à des taux de couverture R_{dot} différents (respectivement 11% ; 25% et 34%).

Le fait de diminuer le taux de couverture R_{dot} permet d'augmenter la vitesse d'effacement (**Fig. IV-8(a)**) en augmentant le champ dans l'oxyde tunnel (**Fig. IV-8(b)**). En effet, la même quantité de charge est stockée sur une plus petite surface ; le champ électrique est donc localement plus fort.

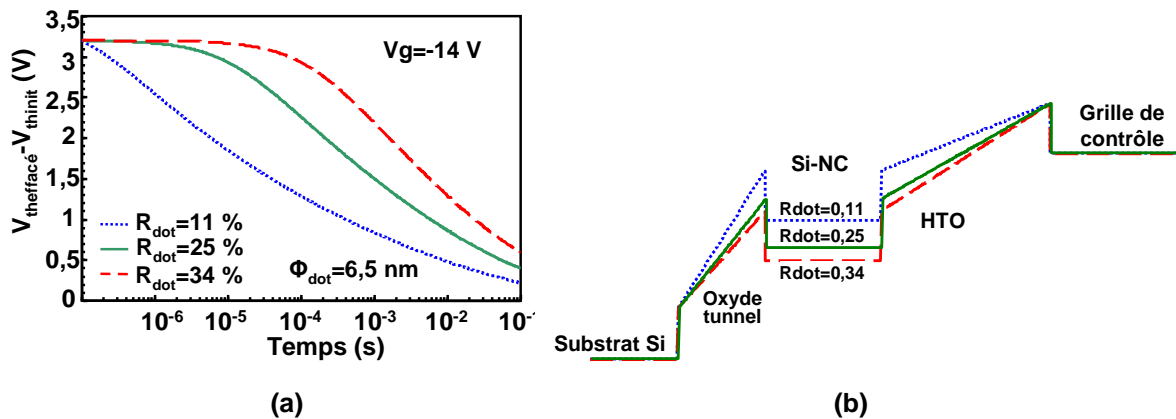


Fig. IV-8 : (a) Dynamiques d'effacement simulées pour des Si-NCs avec différents taux de couverture (11% ; 25% et 34%) et un diamètre identique (6,5 nm) et diagramme des bandes de conduction de l'empilement de grille correspondant à $t=0$ (b). $t_{tun} \approx 5$ nm, $t_{HTO} = 10$ nm.

IV.2.1.1.D Comparaison avec la Flash standard

Le modèle de la quasi-grille flottante permet également de simuler l'effacement de la cellule à grille flottante continue (**Fig. IV-9(a)**) en prenant un taux de couverture de 100 % et en tenant compte du couplage grille flottante/grille de contrôle qui est 2 à 2,5 fois plus élevé par rapport aux Si-NCs.

Nous avons précédemment comparé les dynamiques d'effacement de la cellule à Si-NCs et à grille flottante continue écrites dans les conditions standard ($V_g=8$ V, $V_d=3,75$ V, $V_b=-1,5$ V, $V_s=0$ V, $t=10$ μ s) (Figure III-3 du paragraphe III.2.1.2, chapitre III). Or, à conditions de programmation identiques, la fenêtre de programmation de la Flash est plus élevée que celle des Si-NCs. Nous avons donc effectué une simulation de la Flash standard avec une fenêtre de programmation identique à celle du dispositif à Si-NCs (**Fig. IV-9(b)**).

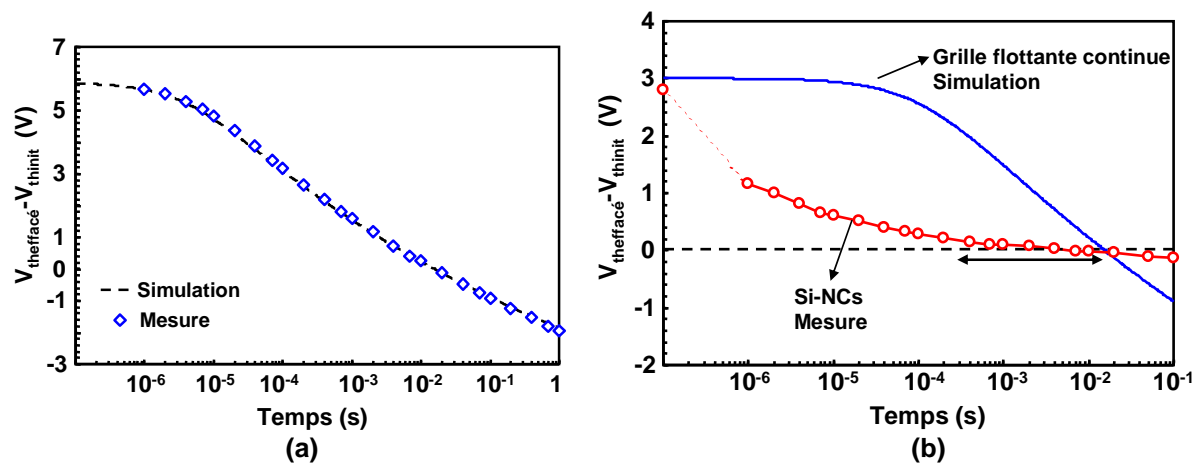


Fig. IV-9 : (a) : Comparaison des dynamiques d'effacement en Fowler-Nordheim ($V_g = -16$ V) mesurées et simulées (avec le modèle de la quasi-grille flottante) de la cellule à grille flottante continue. (b) : Comparaison des dynamiques d'effacement ($V_g = -16$ V) d'une cellule à Si-NCs ($t_{\text{tun}} = 5$ nm, $t_{\text{HTO}} = 10$ nm, $\Phi_{\text{dot}} = 5$ nm, $R_{\text{dot}} \approx 0,15$) et d'une grille flottante continue à fenêtre de programmation identique.

Pour atteindre la tension de seuil initiale, on constate qu'il faut presque deux décades de temps supplémentaires à la Flash standard par rapport à la cellule à Si-NCs. Il faut bien sûr souligner que la vitesse d'effacement du dispositif à Si-NCs dépend, comme on l'a montré précédemment, de plusieurs paramètres technologiques (épaisseur de l'oxyde tunnel, du HTO, taux de couverture des Si-NCs). La différence de temps d'effacement entre les deux dispositifs dépend donc des paramètres de la cellule à Si-NCs.

IV.2.1.2 « Gate disturb »

Comme nous l'avons évoqué dans le chapitre précédent, l'utilisation d'un HTO relativement fin peut conduire à des problèmes de « gate disturb ». Pendant la lecture d'une cellule écrite, une tension positive est appliquée sur la grille. De plus, le fait que les Si-NCs soient chargés conduit à une augmentation du champ électrique dans le HTO qui peut conduire à une fuite des électrons des Si-NCs vers la grille de contrôle par effet tunnel Fowler-Nordheim (**Fig. IV-10(a)**) et donc à l'effacement de la cellule. Nous avons donc simulé le phénomène de « gate disturb » grâce au modèle de la quasi-grille flottante. Comme le montrent les **Fig. IV-10(b)-(c)**, le modèle permet de prédire correctement la variation de la tension de seuil au cours du « disturb » pour différentes tensions de grille et différentes épaisseurs de HTO.

L'influence de la tension de grille est illustrée sur le diagramme de la **Fig. IV-10(a)**. On peut voir que plus la tension de grille est élevée, plus le champ électrique dans le HTO est fort, ce qui conduit à une fuite plus rapide des électrons vers la grille de contrôle.

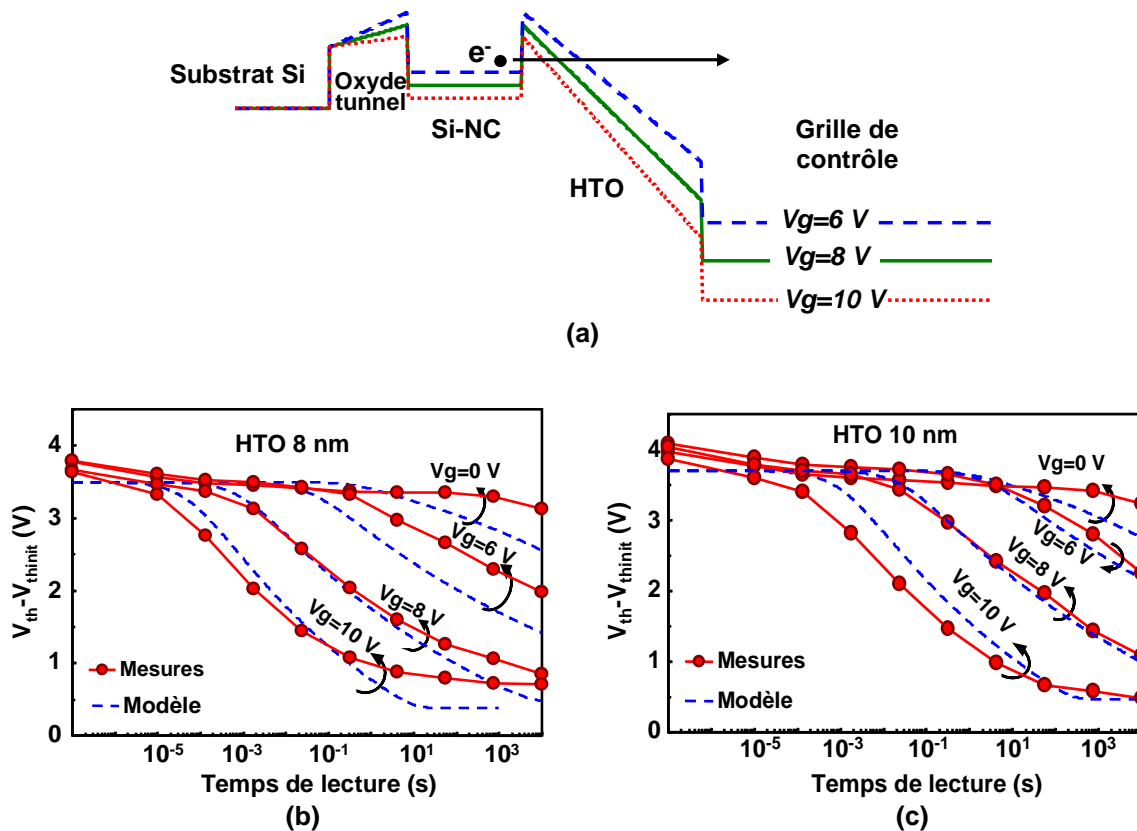


Fig. IV-10 : (a) : Diagramme des bandes de conduction de l'empilement de grille pour différentes tensions de lectures à $t=0$. Comparaison de la variation de la tension de seuil au cours du « gate disturb » mesurée et simulée (avec le modèle de la quasi-grille flottante) pour différentes tensions de grille et deux épaisseurs de HTO (8 nm (a) et 10 nm (b)). $t_{tun} \approx 4\text{ nm}$, $\Phi_{dot} = 5\text{ nm}$, $R_{dot} \approx 0,15$.

Ensuite nous nous sommes intéressés à l'influence de l'épaisseur du HTO. La **Fig. IV-11(a)** montre la comparaison entre les simulations et les mesures de « gate disturb accéléré » pour trois épaisseurs de HTO. On obtient un bon accord entre les caractéristiques simulées et les mesures. Plus le HTO est fin, plus la perte de charge est rapide. En effet, le diagramme des bandes de la **Fig. IV-11(b)** montre que le champ électrique initial dans le HTO est d'autant plus élevé que le HTO est fin.

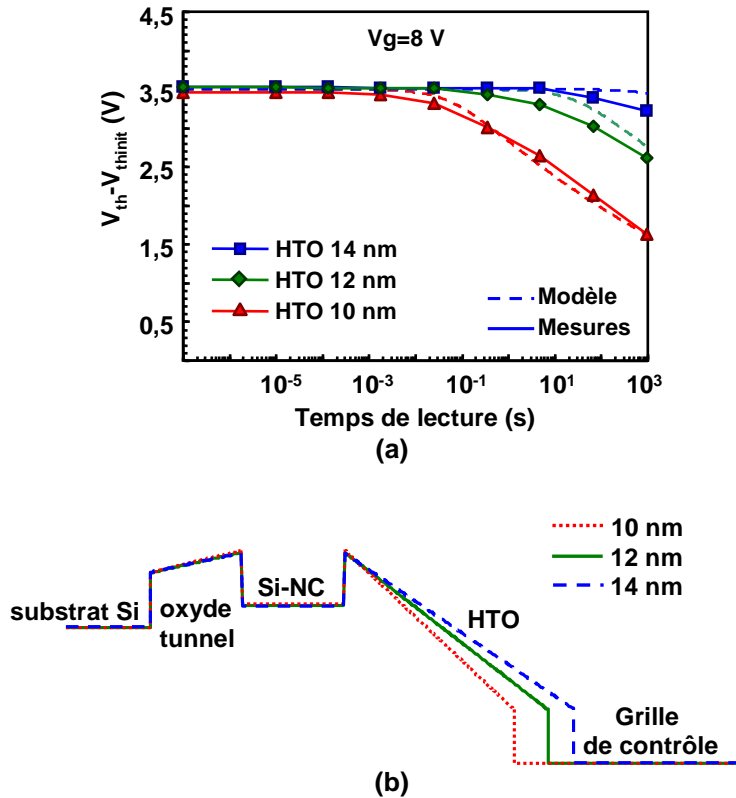


Fig. IV-11 : (a) : Comparaison de la variation de la tension de seuil au cours d'un « gate disturb accéléré » ($V_g=8$ V) mesurée et simulée (avec le modèle de la quasi-grille flottante) pour différentes épaisseurs de HTO (10 nm, 12 nm et 14 nm). $t_{tun} \approx 5$ nm, $\Phi_{dot} = 6,5$ nm, $R_{dot} \approx 0,25$. (b) Diagramme des bandes de conduction de l'empilement de grille à $t=0$.

Finalement, nous avons utilisé le modèle de la quasi-grille flottante, calibré sur les mesures de la **Fig. IV-11(a)** pour prédire la diminution de la tension de seuil après 10 ans dans des conditions de « gate disturb » réalistes. Ces conditions correspondent aux distributions standards présentées dans le chapitre III (**Fig. IV-12(a)** : $V_{th} - V_{thinitial} = 3$ V et $V_{glecture} = 4,5$ V ; **Fig. IV-12(b)** : $V_{th} - V_{thinitial} = 4$ V et $V_{glecture} = 5$ V). Dans le premier cas, le modèle prédit qu'un HTO de 12 nm au minimum est nécessaire pour minimiser l'effet du « gate disturb » au bout de 10 ans. Dans le second cas, la fenêtre de programmation plus élevée implique une tension de lecture plus grande, et par conséquent une perte de charge plus rapide. C'est pourquoi dans ce cas il est nécessaire d'utiliser un HTO plus épais.

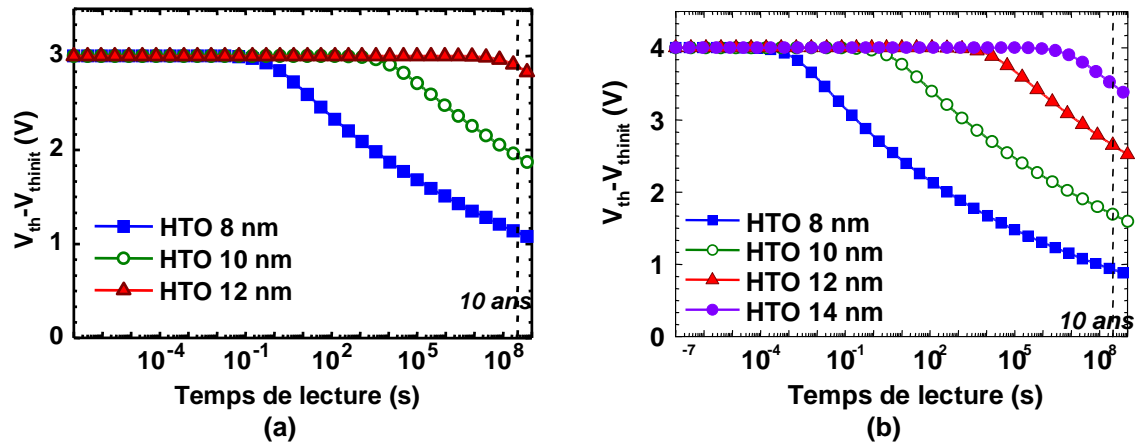


Fig. IV-12 : Simulations du « gate disturb » pour différentes épaisseurs de HTO. (a) $V_{th} - V_{thinitial} = 3$ V et $V_{glecture} = 4,5$ V, ce qui correspond aux conditions de la Figure III-31 paragraphe III.3.3 du chapitre III. (b) : $V_{th} - V_{thinitial} = 4$ V et $V_{glecture} = 5$ V, ce qui correspond aux conditions de la figure Figure III-32(b) paragraphe III.3.3 du chapitre III obtenues avec un temps d'écriture plus long.

IV.3 Ecriture par porteurs chauds

La théorie des porteurs chauds a commencé à être développée à la fin des années '70 pour les questions de fiabilité du transistor MOS, puis les modèles ont été appliqués aux mémoires non volatiles, notamment aux EEPROM, vers 1990. Aujourd'hui, avec l'apparition des mémoires à sites de piégeage discrets, la modélisation de l'injection des électrons chauds demande des révisions ultérieures. En effet, la localisation de la charge injectée devient un nouveau paramètre à prendre en compte.

IV.3.1 Etude de la localisation de la charge dans une cellule mémoire à nanocristaux de silicium

IV.3.1.1 Introduction

Dans le cas d'une mémoire à grille flottante continue, la charge injectée au niveau du drain lors d'une écriture par porteurs chauds se redistribue presque instantanément et uniformément dans la grille flottante. Au contraire, dans le cas d'une mémoire à pièges discrets, la charge injectée près du drain reste localisée au-dessus de la jonction canal/drain (**Fig. IV-13**). Ce phénomène requiert une étude approfondie en vue de l'optimisation des performances de la cellule mémoire. Récemment, l'injection des porteurs chauds a été étudiée dans le cas des mémoires ayant du nitrure comme couche de stockage [**Ingresso'02**] [**Hagenbeck'04**] [**Sun'04**] [**Hagenbeck'06**] [**Tempel'06**]. Des simulations de l'injection en fonction des conditions de programmation ont été réalisées, en particulier pour les dispositifs de type NROM. Il a également été montré que l'utilisation d'un modèle de transport prenant en compte l'énergie des porteurs était nécessaire pour simuler correctement le courant d'injection (**Fig. IV-14**).

Nous allons dans cette partie s'attacher à la compréhension de la localisation de la charge dans les dispositifs à Si-NCs, aussi bien d'un point de vue expérimental que théorique.

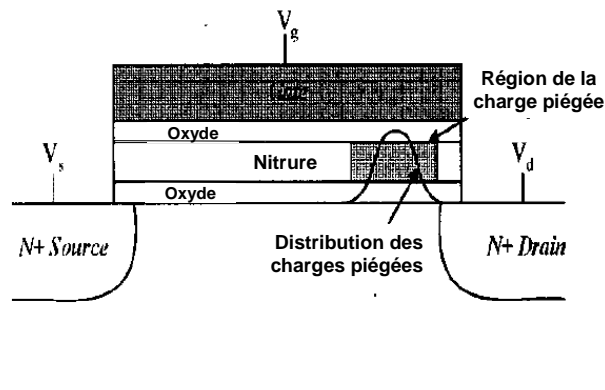


Fig. IV-13 : Représentation de la charge piégée dans une cellule de type SONOS après une écriture par porteurs chauds [**Sun'04**] [**Hagenbeck'04**].

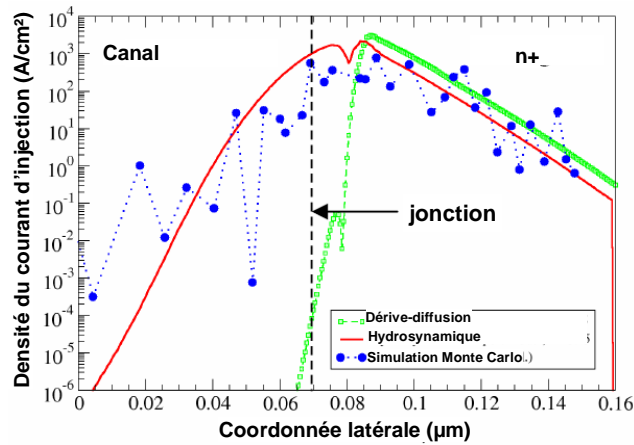


Fig. IV-14 : Simulation de la programmation avec le modèle de l'électron chanceux et différents modèles de transport et simulations Monte Carlo. $V_d = 4$ V, $V_g = 9$ V, $V_s = V_b = 0$ V [Hagenbeck'04].

IV.3.1.2 Dispositifs d'étude

Des tests électriques ont été réalisés sur des cellules mémoires à Si-NCs LPCVD de diamètre moyen 5 nm et de densité environ 10^{12}cm^{-2} (valeurs nominales). Les paramètres de la cellule sont détaillés dans la **Fig. IV-15**.

Des simulations TCAD de ce dispositif ont été réalisées avec les outils commerciaux Silvaco [Silvaco]. Le logiciel Athena permet de simuler la fabrication de la cellule (dépôts, gravure, implantations...) et le logiciel Atlas permet de simuler les caractéristiques électriques de la cellule. Le dispositif 2D simulé est représenté dans la **Fig. IV-16**. La simulation de la fabrication de la cellule a été réalisée en utilisant des paramètres (épaisseurs, implantations, dimensions) les plus proches de la réalité. Les nanocristaux ont été supposés uniformément répartis le long du canal. On fait de plus l'approximation que les nanocristaux soient des carrés métalliques de 5 nm de diamètre, avec une densité surfacique de 10^{12}cm^{-2} .

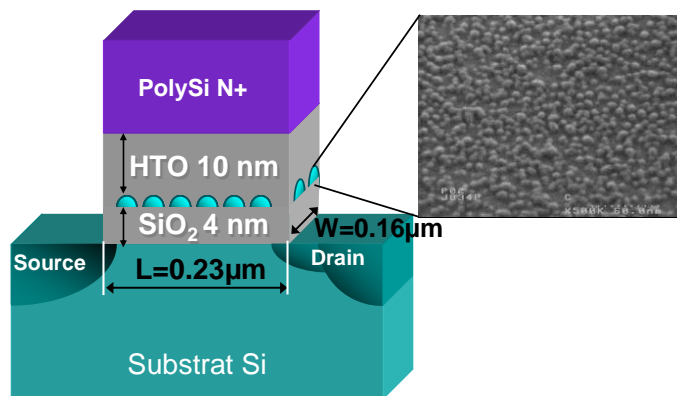


Fig. IV-15 : Schéma du dispositif mémoire à Si-NCs étudié. Insert : Image MEB des Si-NCs.

IV.3.1.3 Simulations électrostatiques

Nous avons d'abord étudié l'influence de la longueur de la zone chargée sur la fenêtre de programmation de la cellule. Nous avons réalisé des simulations électriques de dispositifs mémoires avec différentes longueurs chargées, depuis le drain et s'étendant vers le canal (Fig. IV-17). On voit que la fenêtre de programmation sature avant que toute la longueur du canal ne soit couverte. Pour une cellule de $0,23 \mu\text{m}$ de longueur, la tension de seuil sature à partir d'une longueur de la zone chargée de 150 nm . La valeur de la fenêtre de programmation à saturation dépend linéairement du nombre d'électrons par nanocristal. Dans des mémoires à Si-NCs, il n'est donc pas nécessaire de charger les nanocristaux sur toute la longueur du canal pour obtenir la fenêtre de programmation maximale. On remarque cependant qu'entre 50 nm et 150 nm de longueur chargée, la fenêtre de programmation peut augmenter de plusieurs volts. Il est donc important d'optimiser les conditions de programmation afin d'atteindre une fenêtre la plus grande possible.

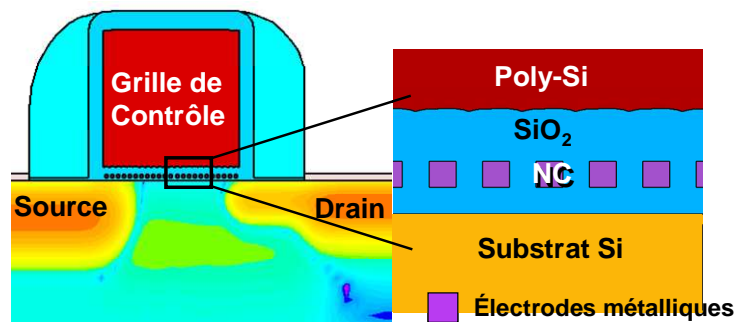


Fig. IV-16 : Structure 2D simulée.

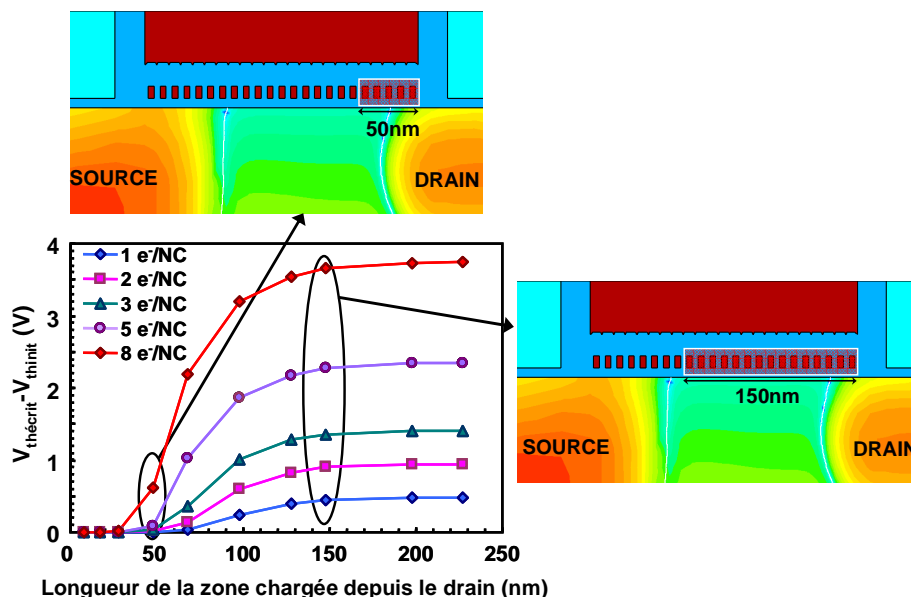


Fig. IV-17 : Fenêtre de programmation simulée en fonction de la longueur de la zone chargée et pour différents nombres d'électrons par nanocristal chargé. Lecture : $V_{ds}=1 \text{ V}$

IV.3.1.4 Simulations dynamiques de l'écriture par électrons chauds et interprétation des données expérimentales

IV.3.1.4.A Simulations TCAD

Dans un premier temps, nous avons comparé les deux modèles de transport dans le canal « dérive-diffusion » [Selberherr'84] et « energy-balance » [Stratton'62] [Stratton'72]. Le modèle energy-balance, contrairement à celui de dérive-diffusion, permet d'accéder à l'énergie des porteurs. Pour cela, nous avons comparé l'écriture par porteurs chauds sur la structure à NCs (**Fig. IV-16**) et sur une structure à grille flottante continue, ayant les mêmes caractéristiques technologiques, en termes d'épaisseurs des oxydes et d'implants (**Fig. IV-18**). Le modèle choisi pour le calcul du courant d'injection est le modèle de l'électron chanceux (**Annexe 2**) [Tam'84], car il permet d'obtenir des résultats qualitatifs avec des temps de calcul raisonnables. Les outils proposent également un deuxième modèle, appelé « Concannon » [Concannon'95]. Ce modèle permet une meilleure prise en compte des effets non locaux, mais les temps de calcul requis augmentent de manière significative.

La **Fig. IV-19** représente une coupe du potentiel le long des nanocristaux. Chaque palier correspond au potentiel d'un nanocrystal (supposé métallique). Plus le palier est bas, plus le nanocrystal est chargé.

La simulation de l'écriture par électrons chauds a été réalisée en utilisant les deux modèles de transport « dérive-diffusion » et « energy balance » pour les structures à grille flottante continue et à NCs. Le champ électrique dans les structures a été représenté après écriture sur les **Fig. IV-20(a)-(b)**. On remarque qu'avec la grille flottante continue, le champ électrique dans l'oxyde de contrôle est uniforme, ce qui n'est pas le cas avec les nanocristaux, à cause de la localisation de la charge au niveau de la jonction drain. Ensuite, nous avons réalisé une coupe du potentiel le long de la grille flottante (**Fig. IV-20(c)**) et des NCs (**Fig. IV-20(d)**) après écriture. On montre clairement la localisation des électrons dans les NCs situés au niveau de la jonction drain, contrairement à la structure à grille flottante continue qui présente un potentiel constant le long du canal. De plus, on constate que dans le cas de la grille flottante continue, le potentiel (et donc la fenêtre de programmation de la cellule) (**Fig. IV-20(e)**) sont identiques quelque soit le modèle de transport utilisé. Au contraire, avec la structure à NCs, on observe que le modèle de transport a une influence sur la répartition de la charge (**Fig. IV-20(d)**) et les fenêtres de programmation résultantes sont très différentes (**Fig. IV-20(f)**). En effet, dans le cas de la cellule standard, une fois injectée, la charge se répartit uniformément dans la grille flottante et la localisation du pic d'injection n'a donc pas d'influence. Dans le cas des cellules à NCs, les électrons restent localisés au niveau de la jonction, la position du pic d'injection et la forme du courant d'injection deviennent des paramètres cruciaux. Il est donc nécessaire d'utiliser un modèle de transport de type « energy balance ».

Ceci démontre bien que la localisation de la charge est un point essentiel pour les mémoires à nanocristaux en vue de leur optimisation.

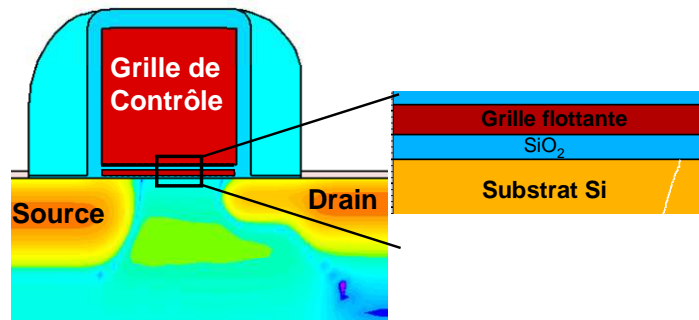


Fig. IV-18 : Structure 2D à grille flottante continue simulée.

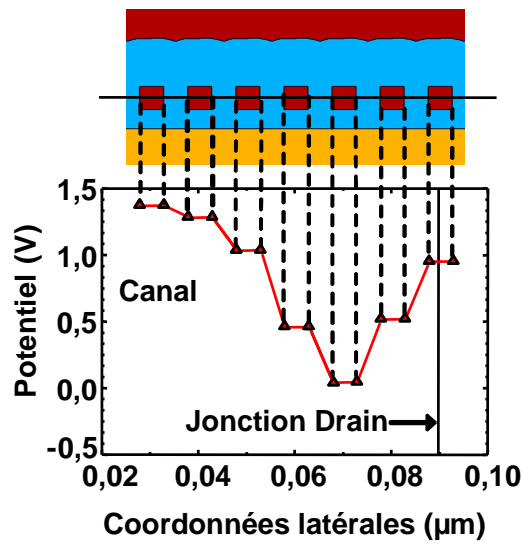


Fig. IV-19 : Coupe du potentiel le long des nanocristaux. Un palier représente le potentiel dans un nanocrystal.

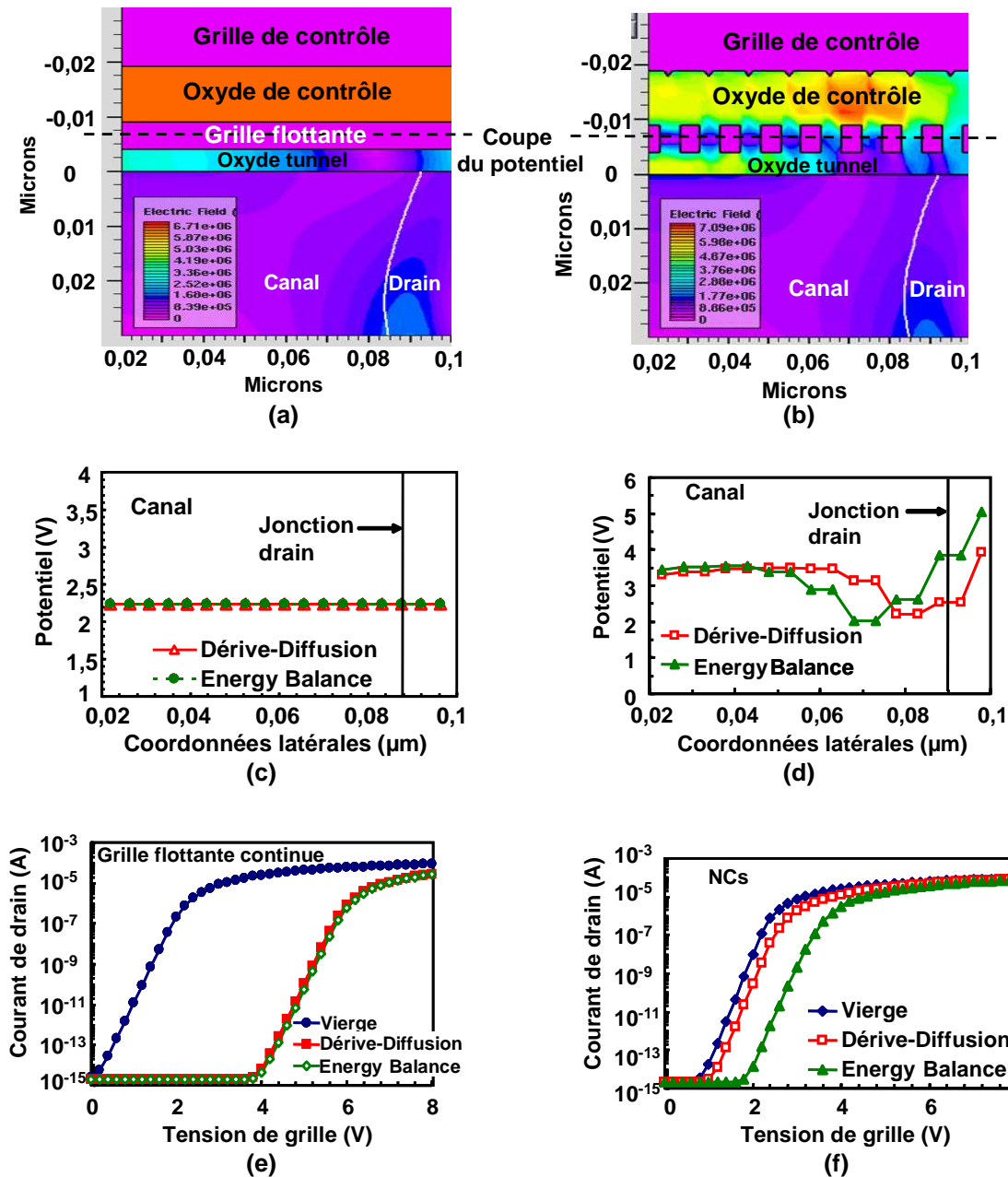


Fig. IV-20 : Simulation de l'écriture par porteurs chauds ($V_g=8$ V, $V_d=4$ V, $V_b=V_s=0$ V, $t=10$ μ s) sur des dispositifs 2D à grille flottante continue et à NCs en fonction du modèle de transport (dérive-diffusion et energy-balance). Champ électrique dans les structures à grille flottante continue (a) et à NCs (b) après écriture. Coupe du potentiel dans la grille flottante (c) et les NCs (d) chargés. I_d - V_g simulées de la structure à grille flottante continue (e) et à NCs (f) avant et après écriture. Lecture : $I_d=10^{-7}$ A et $V_{ds}=0,5$ V.

Nous avons ensuite étudié l'influence des paramètres électriques (tensions et temps d'écriture) sur la localisation de la charge.

Dans un premier temps, des simulations TCAD de l'écriture par électrons ont été effectuées pour différentes tensions de grille et temps d'écriture. On peut voir les caractéristiques I_d - V_g simulées vierges et après écriture sur les **Fig. IV-21(a)-(c)**. L'utilisation du modèle de l'électron chanceux couplé au modèle de transport « energy balance » permet d'obtenir un bon accord entre les mesures et les simulations (**Fig. IV-21(a)-(c)**).

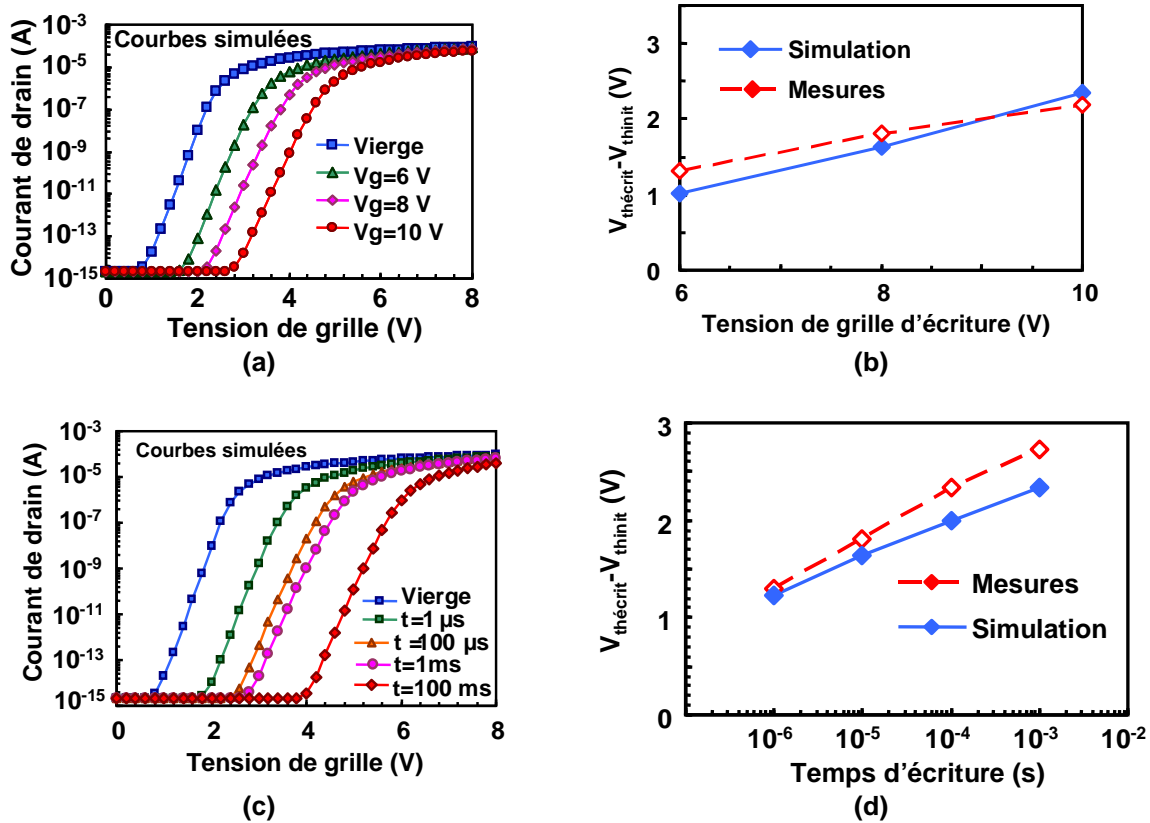


Fig. IV-21 : (a) Id-Vg simulées du dispositif à nanocristaux vierge et écrit par injection d'électrons chauds en fonction de la tension de grille appliquée pendant l'écriture ($V_d=5$ V, $V_b=V_s=0$ V, $t=10$ μ s). (b) : Comparaison entre les fenêtres de programmation simulées correspondantes et les mesures. (c) Id-Vg simulées du dispositif à nanocristaux vierge et écrit par injection d'électrons chauds en fonction du temps d'écriture ($V_g=8$, $V_d=5$ V, $V_b=V_s=0$ V). (d) : Comparaison entre les fenêtres de programmation simulées correspondantes et les mesures. Lecture : $V_{ds}=0,5$ V et $I_d=10^{-7}$ A.

Ensuite, nous avons représenté le champ électrique et le potentiel dans les NCs après écriture pour différentes tensions de grille d'écriture (**Fig. IV-22**). Lorsque la tension de grille appliquée pendant l'écriture augmente, le champ électrique dans l'oxyde de contrôle augmente (**Fig. IV-22(a)-(b)**) et le potentiel dans les NCs diminue (**Fig. IV-22(c)**), ce qui indique que la charge injectée augmente. Mais on voit clairement sur la **Fig. IV-22(c)** que la région de localisation des électrons injectés reste la même quelque soit la tension appliquée sur la grille pendant l'écriture.

Les mêmes figures sont présentées en fonction du temps d'écriture (**Fig. IV-23**). On constate également une augmentation de la quantité de charge injectée lorsque le temps augmente, mais la région de localisation change au cours du temps. En particulier, on observe sur la **Fig. IV-23(c)** que la charge injectée se décale en direction du canal. Ceci peut s'expliquer par le fait que les électrons qui sont progressivement injectés dans les NCs modifient localement le champ électrique dans l'oxyde tunnel et décalent le pic d'injection en direction du canal.

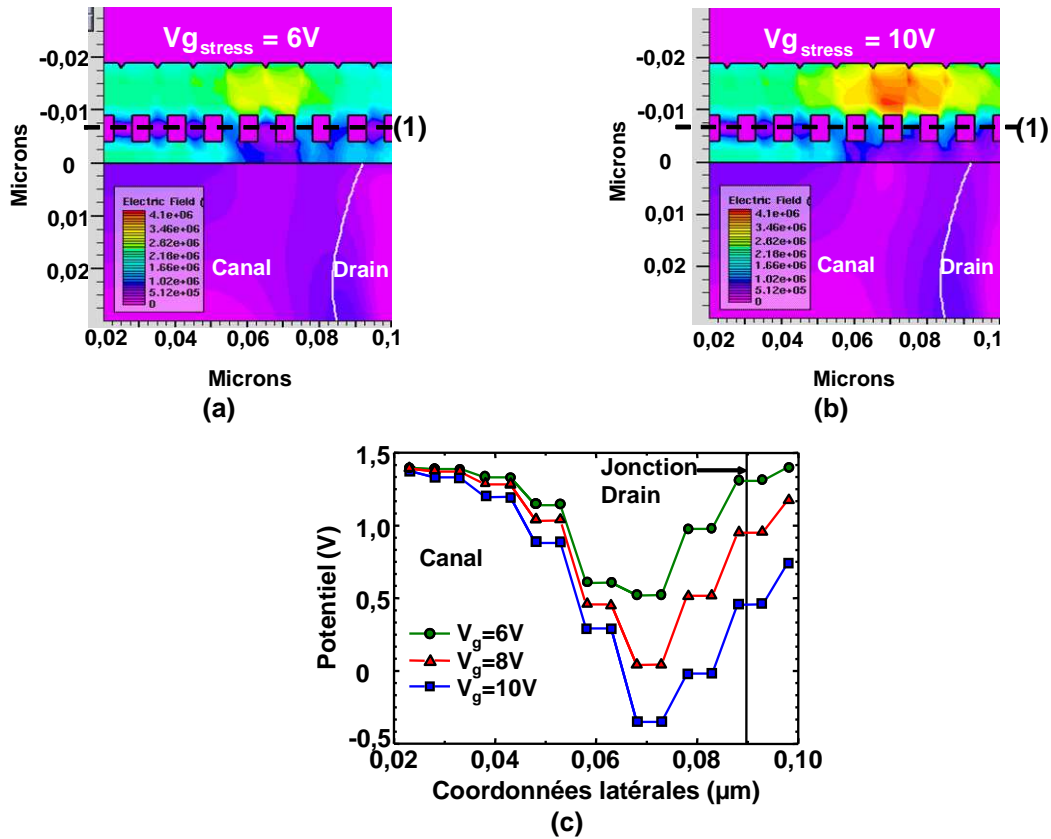


Fig. IV-22 : Champ électrique dans la structure simulée pour différentes tensions de grille d'écriture ($V_g=6 V$ (a) et $V_g=10 V$ (b) tandis que $V_d=5 V$, $V_s=V_b=0 V$ et $t=10 \mu s$. (c) : Potentiel dans les nanocristaux selon la coupe (1) après écriture pour différentes tensions de grille d'écriture.

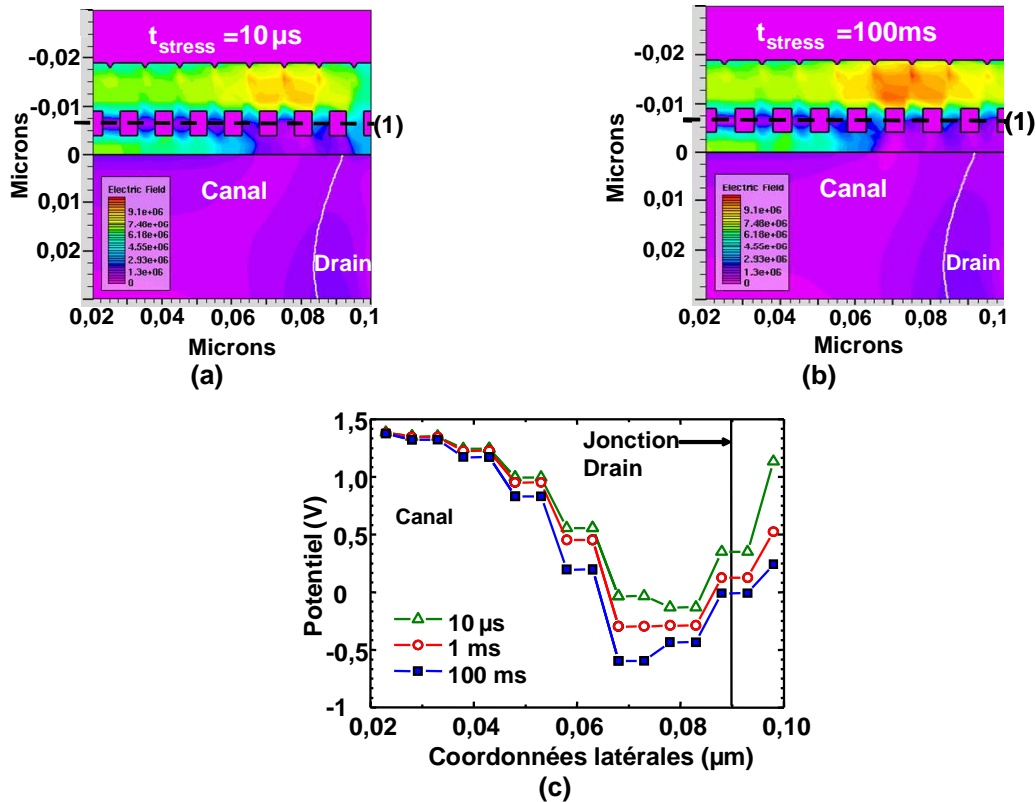


Fig. IV-23 : Champ électrique dans la structure simulée pour différents temps d'écriture ($t=10 \mu s$ (a) et $t=100 ms$ (b) tandis que $V_g=10 V$, $V_d=3,5 V$, $V_s=V_b=0 V$. (c) : Potentiel dans les nanocristaux selon la coupe (1) après écriture pour différents temps d'écriture.

IV.3.1.4.B Modélisation analytique

Dans un deuxième temps, afin de confirmer et d'approfondir les résultats obtenus par les simulations TCAD, nous avons utilisé un modèle analytique décrit en détail dans [Perniola'06]. Ce modèle est basé sur le calcul du potentiel de surface des cellules mémoires à sites de piégeage discrets. Il permet de comprendre le comportement électrostatique des mémoires à sites de piégeage discrets et d'obtenir une formule analytique de la pente sous le seuil et de la tension de seuil des cellules.

La première étape consiste à calculer le potentiel de surface le long du canal d'une cellule vierge (**Fig. IV-24(a)**) :

$$\begin{aligned} \Psi_S(y) = (\Psi_r - \Psi_L) \frac{\sinh(y/\lambda)}{\sinh(L/\lambda)} \\ + (\Psi_l - \Psi_L) \frac{\sinh[(L-y)/\lambda]}{\sinh(L/\lambda)} + \Psi_L \end{aligned} \quad \text{Équation IV-12}$$

Avec $y = 0$ ($y = L$) : contact source (drain), $\Psi_l = V_{bi} + V_s - V_b$, $\Psi_r = V_{bi} + V_d - V_b$, où V_{bi} est le potentiel intrinsèque aux jonctions drain/substrat et source/substrat, V_s : tension de source, V_d : tension de drain, V_b : tension de substrat utilisée pour la lecture des Id - V_g , Ψ_L : Potentiel de surface du canal long selon [Tsividis'99].

Le paramètre λ est défini de la façon suivante :

$$\lambda \equiv \sqrt{\frac{\epsilon_{Si} t_{eox} X_{dep}}{\epsilon_{ox} \eta}} \quad \text{Équation IV-13}$$

Avec t_{eox} : épaisseur équivalente de l'empilement de grille, X_{dep} : épaisseur de la zone de charge d'espace, $\epsilon_{Si(ox)}$: permittivité du silicium (de l'oxyde), η : paramètre de fit (normalement calibré en fittant le potentiel de surface de la cellule mémoire vierge avec des simulations numériques).

Dans la seconde étape, on considère la cellule mémoire écrite par porteurs chauds (**Fig. IV-24(b)**). L'effet électrostatique de la charge est pris en compte grâce au principe de superposition. La région chargée localisée près du drain est supposée uniforme, tandis que le reste de la région au-dessus du canal est non chargée. Le paquet de charge est décrit par le biais de deux paramètres, la longueur effective chargée L_2 et la densité de charge effective Q . Le potentiel perturbateur Ψ_{Sp} (**Fig. IV-24(b)**) est obtenu en intégrant les contributions élémentaires le long des directions verticales x et longitudinales y [Landau'84] :

$$\Psi_{Sp}(y) = \frac{-\rho}{\pi(\epsilon_{Si} + \epsilon_{ox})} \times \int_{L_1}^{L_1+L_2} dy' \int_{t_1}^{t_1+t_m} \ln(\sqrt{(y-y')^2 + x^2}) dx \quad \text{Équation IV-14}$$

Avec ρ : densité de charge injectée par unité de volume - t_1 : épaisseur de l'oxyde tunnel - t_m : hauteur de la région chargée, $L_1 = L - L_2$. (**Fig. IV-25**).

Finalement, le potentiel de surface est : $\Psi_{Stot} = \Psi_S + \Psi_{Sp}$.

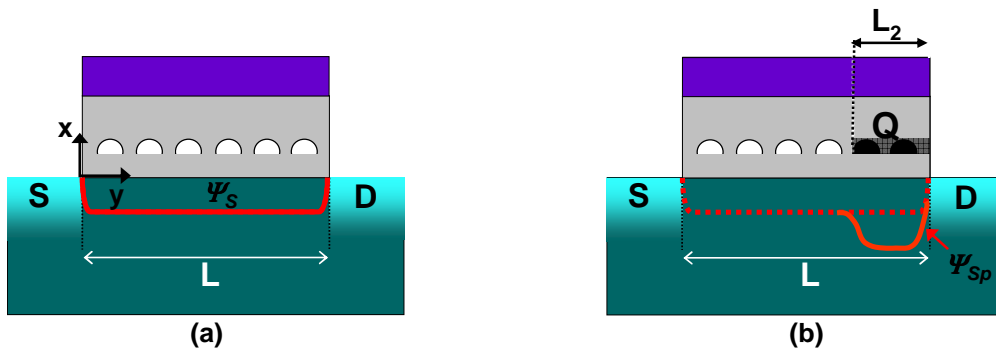


Fig. IV-24 : Potentiel de surface le long du canal d'une cellule mémoire vierge (a) et écrite par porteurs chauds (b).

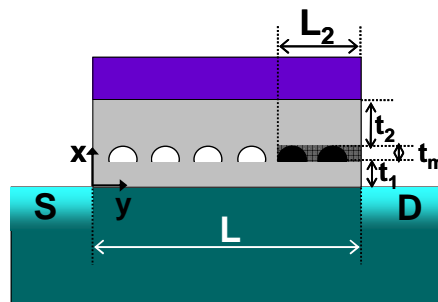


Fig. IV-25 : Paramètres intervenant dans le calcul du potentiel de surface total.

Nous avons utilisé ce modèle afin d'extraire des informations sur la distribution de la charge injectée lors de l'écriture par porteurs chauds dans nos dispositifs à Si-NCs, présentés dans le paragraphe IV.3.1.2. Pour cela, des tests d'écriture par électrons chauds ont été réalisés en fonction des tensions et temps d'écriture. La tension de seuil écrite est lue en mode direct ou « forward » (en appliquant $V_d=1V$ et $V_s=0V$) et en mode inverse ou « reverse » (en appliquant $V_d=0V$ et $V_s=1V$). L'utilisation de ces deux modes de lecture est une particularité des mémoires à deux bits de type NROM, mais peut s'appliquer à toutes mémoires à sites de piégeage discrets. En effet, comme nous l'avons montré précédemment, pour ce type de mémoire, la charge injectée lors de l'écriture par porteurs chauds reste localisée au-dessus de la jonction drain. Dans ce cas, la tension de seuil lue en mode direct (V_{thF}) est inférieure à celle lue en inverse (V_{thR}) (**Fig. IV-26**). Cela peut se comprendre parce qu'au cours de la lecture en mode direct, la charge est écrantée par le potentiel appliqué sur le drain, réduisant son impact sur le potentiel du canal. Au contraire, la lecture en mode inverse est plus sensible aux charges négatives injectées au niveau du drain, qui font décroître plus fortement la conduction dans le canal [Larcher'02]. On comprend alors que lorsque la longueur du paquet de charge augmente, la différence entre les tensions de seuil directe et inverse (ΔV_{RF}) diminue. Les mesures effectuées nous donnent les valeurs de la tension de seuil de la cellule vierge (V_{thinit}) et les valeurs de V_{thF} et V_{thR} pour les différentes conditions d'écriture testées. Ces valeurs nous permettent de définir deux paramètres $\Delta V_{thot}=V_{thR}-V_{thinit}$ et $\Delta V_{RF}=V_{thR}-V_{thF}$ (**Fig. IV-26**).

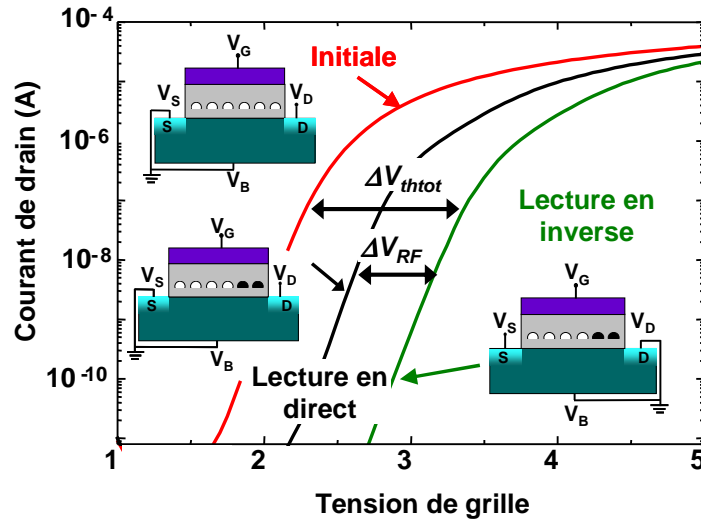


Fig. IV-26 : Illustration des modes de lecture direct et inverse d'une mémoire à sites de piégeage discrets.

De plus, à partir du modèle analytique, il est possible d'obtenir un graphe représentant les isocourbes de ΔV_{thtot} et ΔV_{RF} en fonction de la longueur effective chargée L_2 et de la densité de charge injectée Q . Sur ce graphe, en reportant les valeurs expérimentales de ΔV_{thtot} et ΔV_{RF} , on obtient un point à l'intersection de deux isocourbes ΔV_{thtot} et ΔV_{RF} , à partir duquel on peut alors déduire les valeurs de Q et L_2 correspondantes (Fig. IV-27).

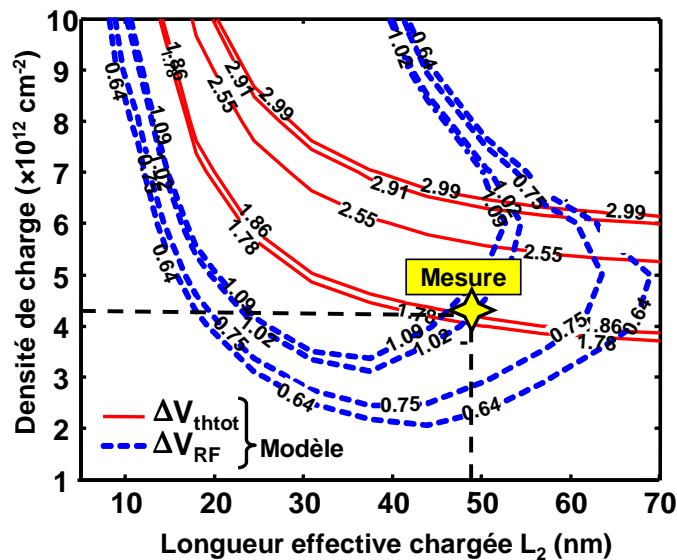


Fig. IV-27 : Exemple d'isocourbes de ΔV_{thtot} (lignes continue) et ΔV_{RF} (lignes en pointillés) obtenues à partir du modèle analytique en fonction de la densité de charge Q et de la longueur effective chargée L_2 . En reportant le point de mesure sur ce graphe, on obtient les paramètres Q et L_2 correspondants.

Nous avons donc effectué cette démarche pour toutes nos mesures afin d'obtenir une description du paquet de charge injecté (Q et L_2) en fonction des conditions d'écriture.

La **Fig. IV-28** montre l'influence de la tension de grille d'écriture. Les données expérimentales (**Fig. IV-28(a)**) ont été reportées sur le graphe analytique (**Fig. IV-28(b)**). On voit que lorsque la tension de grille augmente, la densité de charge Q augmente, tandis que la longueur effective chargée L_2 reste la même. Ceci est en accord avec les résultats de simulation TCAD qui montraient que la région de localisation de la charge restait la même.

Ensuite, l'influence du temps d'écriture a été étudié. Les mesures de la **Fig. IV-29(a)** reportées sur les isocourbes (**Fig. IV-29(b)**) montrent clairement que dans ce cas la longueur de la zone chargée augmente au cours du temps d'écriture, ce qui revient à dire qu'au cours du temps, la charge devient moins localisée et l'injection est décalée en direction du canal. Ceci confirme encore une fois les résultats des simulations TCAD.

Enfin, nous avons étudié l'impact d'une polarisation négative du substrat au cours de l'écriture. Les mesures montrent une large augmentation de la tension de seuil lorsque le substrat est polarisé négativement, puis une saturation à partir de $-1,5$ V (**Fig. IV-30(a)**). Le graphe analytique couplé aux points de mesures (**Fig. IV-30(b)**) montre que la densité de charge Q augmente fortement lorsque V_b diminue de 0 à $-1/-1,5$ V, puis commence ensuite à saturer. On observe le même comportement pour la longueur chargée L_2 . Elle augmente jusqu'à $V_b=-1$ V, puis reste constante. Lorsque le substrat est polarisé négativement, l'injection a lieu sur une zone plus étendue en direction du canal. Ce résultat est cohérent avec la littérature qui décrit le mécanisme d'écriture CHISEL (mécanisme en jeu lorsque $V_b < 0$ V). Ce mécanisme fait intervenir des électrons secondaires, qui sont injectés sur une zone plus étendue en direction du canal [Jungemann'97] [Kencke'98] [Ingrasso'02] [Driussi'04].

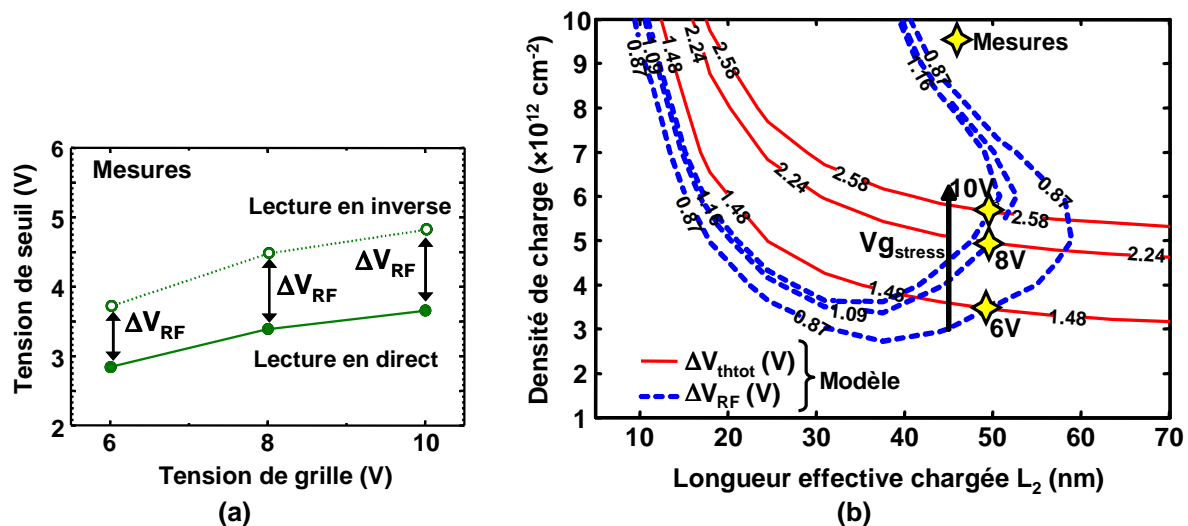


Fig. IV-28 : (a) : Tension de seuil écrite en fonction de la tension de grille appliquée pendant l'écriture ($V_d=5$ V, $V_s=V_b=0$ V, $t=10$ μ s). Lecture : $|V_{ds}|=1$ V et $I_{ds}=100$ nA. (b) : Isocourbes de ΔV_{thtot} (lignes continues) et ΔV_{RF} (lignes en pointillés) obtenues à partir du modèle analytique en fonction de la densité de charge Q et de la longueur effective chargée L_2 . Les points expérimentaux pour différentes tensions de grille d'écriture sont également représentés.

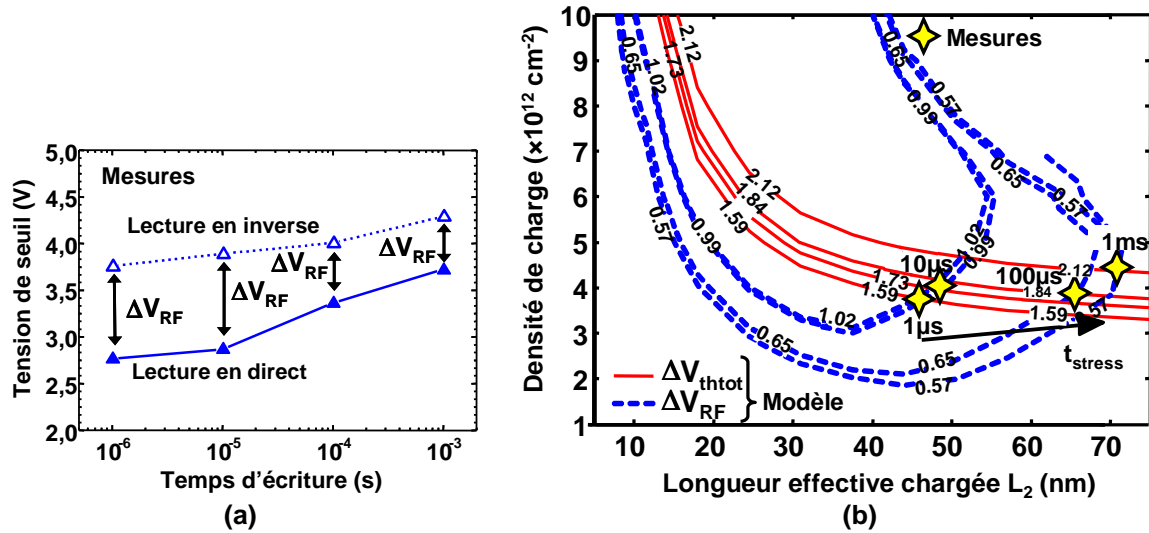


Fig. IV-29 : (a) : Tension de seuil écrite en fonction du temps d'écriture ($V_g=10 \text{ V}$, $V_d=3,5 \text{ V}$, $V_s=V_b=0 \text{ V}$). Lecture : $|V_{ds}|=1 \text{ V}$ et $I_{ds}=100 \text{ nA}$. (b) : Isocourbes de ΔV_{thtot} (lignes continue) et ΔV_{RF} (lignes en pointillés) obtenues à partir du modèle analytique en fonction de la densité de charge Q et de la longueur effective chargée L_2 . Les points expérimentaux pour différents temps d'écriture sont également représentés.

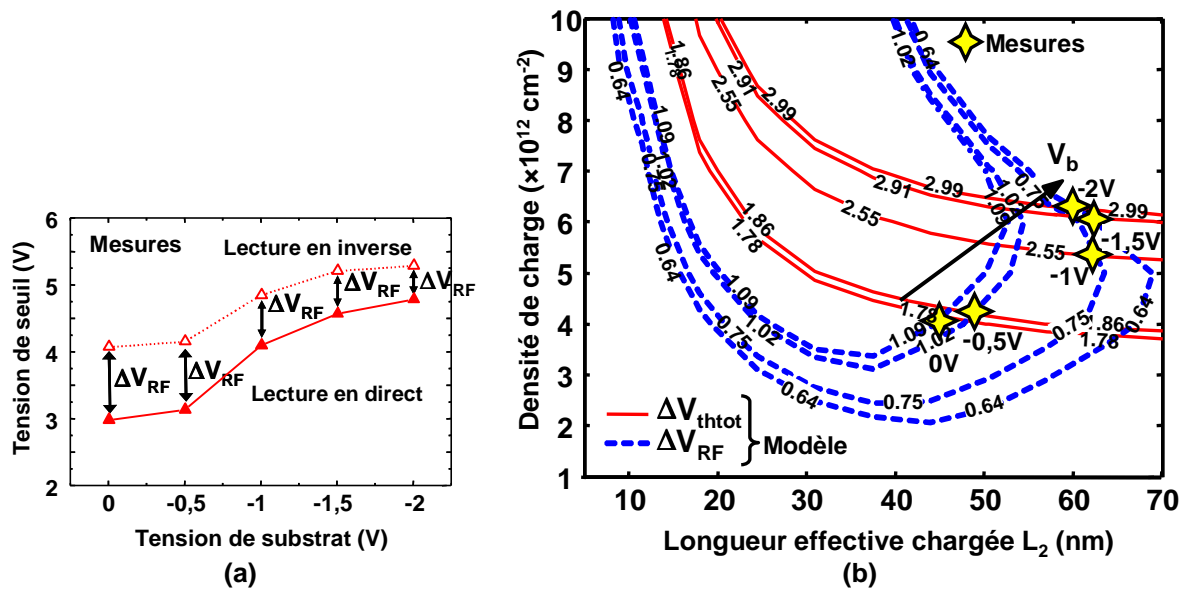


Fig. IV-30 : (a) : Tension de seuil écrite en fonction de la tension de substrat appliquée pendant l'écriture ($V_g=8 \text{ V}$, $V_d=3,5 \text{ V}$, $V_s=0 \text{ V}$, $t=100 \mu\text{s}$). Lecture : $|V_{ds}|=1 \text{ V}$ et $I_{ds}=100 \text{ nA}$. (b) : Isocourbes de ΔV_{thtot} (lignes continue) et ΔV_{RF} (lignes en pointillés) obtenues à partir du modèle analytique en fonction de la densité de charge Q et de la longueur effective chargée L_2 . Les points expérimentaux pour différentes tensions de substrat sont également représentés.

IV.4 Conclusion

Dans ce dernier chapitre, nous nous sommes intéressés à la modélisation des mémoires Flash à nanocristaux de silicium. Dans une première partie, nous avons présenté le modèle dit de la quasi-grille flottante [DeSalvo'01], qui nous a permis de simuler l'effacement Fowler-Nordheim et le « gate disturb », afin de mieux comprendre les résultats électriques présentés dans le chapitre III. Ce modèle est basé sur l'équation de continuité du courant et les équations de base des mémoires Flash. Le fait que la grille flottante ne soit pas continue mais discrète est pris en compte en introduisant le facteur R_{dot} , qui représente la portion de surface couverte par les Si-NCs et qui permet de pondérer l'influence des charges piégées sur la tension de seuil de la cellule. Nous avons montré qu'un bon accord était obtenu entre les mesures et les simulations. Concernant l'effacement, ces modélisations ont permis de cerner l'influence des épaisseurs de l'oxyde tunnel et de contrôle, ainsi que l'influence du taux de couverture des Si-NCs. Nous avons montré que la diminution de l'épaisseur de l'oxyde tunnel se traduisait par une augmentation du champ électrique initial dans celui-ci et donc par un effacement plus rapide. La diminution de l'épaisseur du HTO produit le même effet grâce à l'augmentation du couplage entre les nanocristaux et la grille de contrôle. Lorsque le taux de couverture diminue, la charge est stockée sur une plus petite surface et le champ électrique dans l'oxyde tunnel est donc localement plus fort, ce qui conduit, là encore, à une plus grande vitesse d'effacement.

Comme nous l'avons évoqué dans le chapitre précédent, l'utilisation d'un HTO relativement fin peut conduire à des problèmes de « gate disturb », c'est-à-dire à la fuite des électrons des Si-NCs vers la grille de contrôle par effet tunnel Fowler-Nordheim. Nous avons d'abord mis en évidence qu'une tension de grille de lecture élevée entraînait un champ électrique dans le HTO plus fort et donc une fuite plus rapide des électrons vers la grille de contrôle. Nous avons également montré que le champ électrique initial dans le HTO était d'autant plus élevé que le HTO était fin. Les simulations ont prédit qu'un HTO de 12 nm au minimum était nécessaire pour minimiser l'effet du « gate disturb » après de 10 ans de lecture.

Dans la deuxième partie du chapitre, nous nous sommes intéressés à l'écriture par électrons chauds et en particulier à la localisation de la charge dans les nanocristaux. En effet, dans le cas d'une mémoire à pièges discrets, la charge injectée près du drain reste localisée au-dessus de la jonction canal/drain. La localisation de la charge est un point essentiel pour les mémoires à nanocristaux en vue de leur optimisation. Des simulations TCAD dynamiques de l'écriture par électrons chauds ont d'abord été effectuées. Nous avons montré que pour une cellule à nanocristaux, l'utilisation d'un modèle de transport prenant en compte l'énergie des porteurs était nécessaire pour simuler correctement l'injection des électrons. Ensuite, nous avons étudié l'influence des conditions d'écriture sur la localisation de la charge à l'aide de simulations TCAD et d'un modèle analytique [Perniola'06] couplé à des mesures expérimentales. La région de localisation des électrons injectés reste la même quelque soit la tension de grille d'écriture. Au contraire, lorsque le temps d'écriture augmente ou lorsqu'une

tension négative est appliquée sur le substrat pendant l'écriture, la charge devient moins localisée et l'injection est décalée en direction du canal.

ANNEXE 1 : Mécanismes de conduction tunnel Fowler-Nordheim et tunnel direct

De nombreux mécanismes de conduction des électrons existent dans les diélectriques. Suivant la température, le type de matériau et son épaisseur, les mécanismes diffèrent [DeSalvoTh]. Dans cette partie, nous nous intéresserons plus particulièrement à deux modes de transport limités par l'injection, autrement dit le mécanisme tunnel Fowler-Nordheim et le mécanisme de tunnel direct.

A. Le mécanisme tunnel Fowler-Nordheim

Lorsque un oxyde est soumis à un fort champ électrique, la barrière énergétique vue par les électrons de la cathode est triangulaire. Les électrons traversent alors la barrière de potentiel de l'oxyde par effet tunnel, en transitant de la bande de conduction de la cathode pour arriver dans la bande de conduction de l'oxyde (Fig. IV-31).

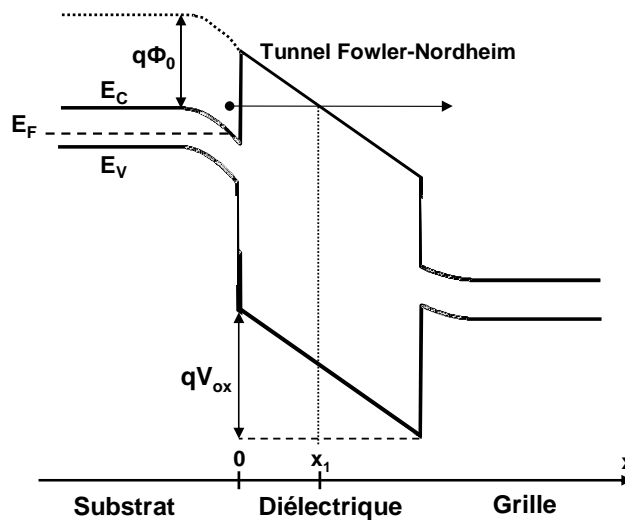


Fig. IV-31 : Représentation du mode de conduction Fowler-Nordheim à travers le diélectrique isolant d'une structure MOS.

La densité de courant à travers une barrière tunnel peut s'exprimer par la formule de Tsu-Esaki [Tsu'73] :

$$J_{\text{FN}} = \frac{4\pi m_{\text{Si}} q}{h^3} \int_{E_C}^{E_C+q\Phi_0} \text{Tr}(E_x) N(E_x) dE_x \quad \text{Équation IV-15}$$

Avec J_{FN} : densité de courant Fowler-Nordheim,

E_C : Energie correspondant au bas de la bande de conduction de l'électrode injectante à l'interface substrat/diélectrique,

Φ_0 : hauteur de la barrière de potentiel en entrée à $x=0$,

Tr : transparence tunnel pour les électrons dont le calcul sera détaillé dans le paragraphe suivant,

N : « supply function », qui sera définie dans le paragraphe suivant,

E_x : énergie électronique transverse - m_{Si} : masse effective de l'électron dans le Silicium, dans le sens de propagation du courant tunnel - h : constante de Planck - q : charge de l'électron.

Nous allons maintenant détailler le calcul des fonctions $N(E)$ et $\text{Tr}(E)$.

La fonction de distribution en énergie des électrons est définie de la façon suivante :

$$N(E_x) = \int_0^{+\infty} f(E) dE_\rho \quad \text{Équation IV-16}$$

Avec E_x et E_ρ les composantes longitudinales et transverses, respectivement telles que $E = E_x + E_\rho$

f : fonction de distribution en énergie des électrons. En considérant que les électrons se comportent comme un gaz de Fermi tridimensionnel [Chang'84], la probabilité d'occupation d'un état d'énergie E est alors donnée par la fonction de Fermi-Dirac :

$$f(E) = \frac{1}{\exp\left(\frac{E - E_F}{kT}\right) + 1} \quad \text{Équation IV-17}$$

Avec E_F : niveau de Fermi dans l'électrode injectante - k : constante de Boltzmann - T : température.

On obtient donc l'Équation IV-18 :

$$N(E_x) = \int_0^{+\infty} \frac{1}{\exp\left(\frac{E_x + E_\rho - E_F}{kT}\right) + 1} dE_\rho \quad \text{Équation IV-18}$$

Cette expression peut être calculée en effectuant le changement de variable $x = \frac{E_x + E_\rho - E_F}{kT}$ et

sachant que $\int \frac{1}{\exp(x)+1} dx = -\ln(e^{-x} + 1)$

On obtient finalement :

$$N(E_x) = kT \ln \left[\exp \left(-\frac{E_x - E_F}{kT} \right) + 1 \right] \quad \text{Équation IV-19}$$

La transparence d'une barrière de potentiel est définie comme la probabilité de passage d'un électron à travers celle-ci. Sa valeur est donc dans l'intervalle [0 ;1]. Le passage d'un électron à travers une barrière tunnel traduit la pénétration de sa fonction d'onde dans celle-ci. Le calcul exact de cette transparence se révèle être un problème mathématique très complexe. Cependant, il est possible d'utiliser l'approximation Wentzel-Kramers-Brillouin (WKB) [Landau'65] [Gasirowicz'95] qui est l'approximation la plus couramment utilisée pour les simulations de courant tunnel [Nagano'94] [Hadjadj'01]. L'approximation WKB s'applique lorsque l'on considère que les longueurs d'onde associées aux fonctions d'onde des porteurs sont petites devant la dimension caractéristique de variation de la barrière de potentiel à traverser.

La transparence peut alors s'écrire :

$$Tr(E) = \exp \left(-\frac{2\sqrt{2}}{\hbar} \int_0^{x_1} [m_{ox}(U(x) - E)]^{1/2} dx \right) \quad \text{Équation IV-20}$$

Avec m_{ox} : masse électronique dans l'oxyde – \hbar : constante de Planck réduite,
 $U(x)-E$: hauteur de la barrière de potentiel à la distance x.

Dans le cas d'une barrière de potentiel triangulaire, l'énergie potentielle peut s'exprimer de la façon suivante :

$$U(x) = q\Phi_0 - qE_{ox}x \quad \text{Équation IV-21}$$

Avec E_{ox} : champ électrique ($= \frac{V_{ox}}{t_{ox}}$)

En remplaçant U(x) par son expression dans l'Équation IV-20, on obtient :

$$Tr(E) = \exp \left(-\frac{2\sqrt{2}}{\hbar} \int_0^{x_1} [m_{ox}(q\Phi_0 - qE_{ox}x - E)]^{1/2} dx \right) \quad \text{Équation IV-22}$$

Sachant que lorsque $x=x_1$, $q\Phi_0 - qE_{ox}x = E$ et en effectuant le changement de variable $v = q\Phi_0 - qE_{ox}x - E$, $dv = -qE_{ox}.dx$ on peut alors facilement calculer la transparence :

$$Tr(E) = \exp \left(-\frac{4\sqrt{2m_{ox}}}{3q\hbar E_{ox}} (q\Phi_0 - E)^{3/2} \right) \quad \text{Équation IV-23}$$

Finalemment, en substituant $\mathbf{N}(E)$ et $\mathbf{Tr}(E)$ par leurs expressions dans l'Équation IV-19, on obtient l'expression de la densité de courant Fowler-Nordheim :

$$J_{\text{FN}} = \frac{4\pi q m_{\text{Si}} kT}{h^3} \int_0^{q\Phi_0} \exp\left(-\frac{4\sqrt{2m_{\text{ox}}}}{3q\hbar E_{\text{ox}}}(q\Phi_0 - E_x)^{3/2}\right) \ln\left[1 + \exp\left(-\frac{E_x - E_F}{kT}\right)\right] dE_x \quad \text{Équation IV-24}$$

B. Le mécanisme tunnel direct

Dans le cas de la conduction par tunnel direct, les électrons transitent par effet tunnel à travers le diélectrique directement de la bande de conduction de la cathode à la bande de conduction de l'anode (**Fig. IV-32**).

La formule de Tsu-Esaki est également valable pour calculer la densité de courant tunnel direct. Seules les bornes de l'intégrale changent, car pour $E > E_C + q\Phi_0 - qV_{\text{ox}}$, le mode de conduction n'est plus direct mais en Fowler-Nordheim.

La densité de courant tunnel direct \mathbf{J}_{TD} est donc la suivante :

$$J_{\text{TD}} = \frac{4\pi m_{\text{Si}} q}{h^3} \int_{E_C}^{E_C + q\Phi_0 - qV_{\text{ox}}} \mathbf{Tr}(E_x) \mathbf{N}(E_x) dE_x \quad \text{Équation IV-25}$$

Le calcul de la fonction $\mathbf{N}(E)$ est le même que celui effectué au paragraphe A.

Concernant le calcul de la transparence, le raisonnement est identique. La seule différence est l'abscisse du point de sortie qui est égal à t_{ox} .

On obtient donc :

$$\mathbf{Tr}(E) = \exp\left(\frac{-2\sqrt{2}}{\hbar} \int_0^{t_{\text{ox}}} [m_{\text{ox}}(U(x) - E)]^{1/2} dx\right) \quad \text{Équation IV-26}$$

L'intégrale se calcule de la même manière avec la condition $q\Phi_0 - qE_{\text{ox}}t_{\text{ox}} = E$, c'est-à-dire $q\Phi_0 - qV_{\text{ox}} = E$. On obtient ainsi l'expression de la transparence :

$$\mathbf{Tr}(E) = \exp\left(-\frac{4\sqrt{2m_{\text{ox}}}}{3q\hbar E_{\text{ox}}} [(q\Phi_0 - E)^{3/2} - (q\Phi_0 - qV_{\text{ox}} - E)^{3/2}]\right) \quad \text{Équation IV-27}$$

Finalemment, la densité de courant tunnel direct est égale à :

$$J_{\text{TD}} = \frac{4\pi q m_{\text{Si}} kT}{h^3} \int_0^{q(\Phi_0 - V_{\text{ox}})} \exp\left(-\frac{4\sqrt{2m_{\text{ox}}}}{3q\hbar E_{\text{ox}}} [(q\Phi_0 - E_x)^{3/2} - (q\Phi_0 - qV_{\text{ox}} - E_x)^{3/2}]\right) \cdot \ln\left[1 + \exp\left(-\frac{E_x - E_F}{kT}\right)\right] dE_x \quad \text{Équation IV-28}$$

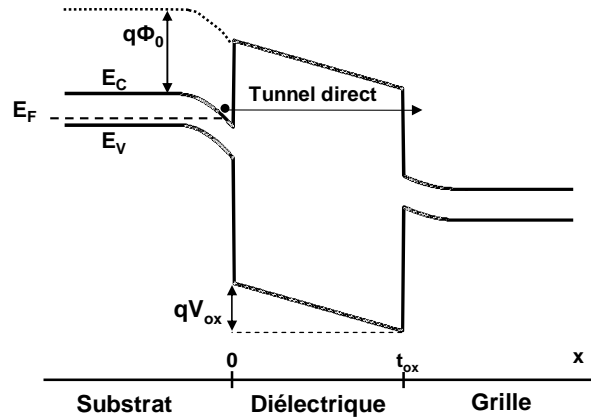


Fig. IV-32 : Représentation du mode de conduction tunnel direct à travers le diélectrique isolant d'une structure MOS.

Les expressions de la transparence tunnel direct et Fowler-Nordheim peuvent être généralisées en une seule expression qui est la suivante :

$$Tr(E) = \exp\left(-\frac{4\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left[(q\Phi_0 - E)^{3/2} - H(q\Phi_0 - qV_{ox} - E) \cdot (q\Phi_0 - qV_{ox} - E)^{3/2} \right]\right) \quad \text{Équation IV-29}$$

Avec **H** : Fonction de Heaviside.

De cette façon, on obtient une expression générale pour les courants tunnel direct et Fowler-Nordheim :

$$J_{\text{tunnel}} = \frac{4\pi q m_{Si} kT}{h^3} \int_0^{q\Phi_0} \exp\left(-\frac{4\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left[(q\Phi_0 - E_X)^{3/2} - H(q\Phi_0 - qV_{ox} - E_X) \cdot (q\Phi_0 - qV_{ox} - E_X)^{3/2} \right]\right) \cdot \ln\left[1 + \exp\left(-\frac{E_X - E_F}{kT}\right)\right] dE_X \quad \text{Équation IV-30}$$

ANNEXE 2 : Quelques éléments de la théorie des porteurs chauds

La configuration de la cellule mémoire lors de l'écriture par électrons est rappelée dans la **Fig. IV-33**. On peut expliquer ce mécanisme de la façon suivante : les électrons sont accélérés le long du canal par le champ électrique latéral appliqué, ce qui leur permet d'acquérir de l'énergie, qu'ils peuvent perdre dans les collisions avec les phonons. A faible champ, c'est-à-dire jusqu'à environ 100kV/cm [Cotrell'79], les électrons sont en équilibre avec le réseau cristallin du silicium. Pour des champs électriques dépassant cette valeur, la condition d'équilibre n'est plus respectée et l'énergie des électrons par rapport au bas de la bande de conduction augmente. Ainsi, une partie des électrons qui arrive au drain a une énergie suffisante pour passer par-dessus la barrière de potentiel Si/SiO₂ (**Fig. IV-34**).

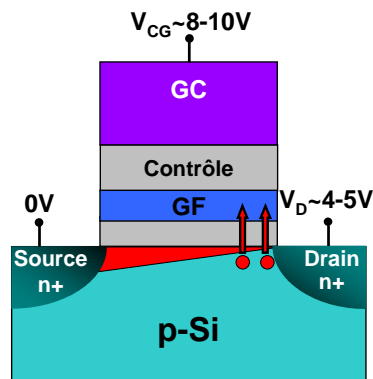


Fig. IV-33 : Conditions de polarisation de la cellule mémoire pendant l'écriture par électrons chauds.

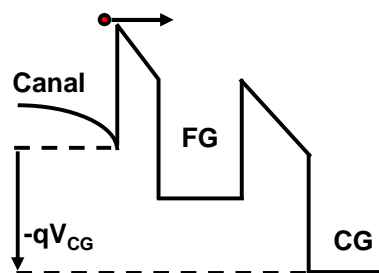


Fig. IV-34 : Diagramme des bandes de conduction au cours de l'écriture par porteurs chauds.

Malgré la complexité de l'injection, on peut trouver des modèles décrivant les conditions de l'injection d'électrons chauds dans la littérature. Celle qui est la plus utilisée est le modèle empirique dit de « l'électron chanceux » (ou « lucky electron »). Ce modèle, introduit par Shockley pour modéliser la ionisation par impact dans les jonctions p-n en 1961 [Shockley'61], a été repris en 1979 pour calculer le courant de grille dans un transistor MOS [Hu'79].

Modèle de « l'électron chanceux »

Ce modèle est basé sur la probabilité qu'un électron, sous l'effet du champ électrique latéral, soit assez « chanceux » pour voyager de la source vers le drain sans subir de collision sur une distance égale à plusieurs fois le libre parcours moyen de l'électron. En arrivant au drain, l'énergie qu'il a acquise doit être suffisante pour qu'après qu'une dernière collision dans le canal l'ait redirigé en direction de l'interface semiconducteur/oxyde il puisse encore surmonter la barrière Si/SiO₂.

La probabilité d'injection est donc le produit des probabilités des événements suivants décrits sur la **Fig. IV-35** [Tam'84] :

- L'électron doit d'abord acquérir une énergie suffisante grâce au champ électrique latéral pour surmonter la barrière Si/SiO₂ après avoir subi une collision le redirigeant vers l'interface (probabilité P_{Φ_B}).
- L'électron suit un chemin sans collision du point de redirection jusqu'à l'interface (probabilité P_1).
- Après avoir surmonté la barrière Si/SiO₂, l'électron ne doit subir aucune collision dans l'oxyde afin de rejoindre la grille flottante sous l'effet du champ électrique vertical. (probabilité P_2).

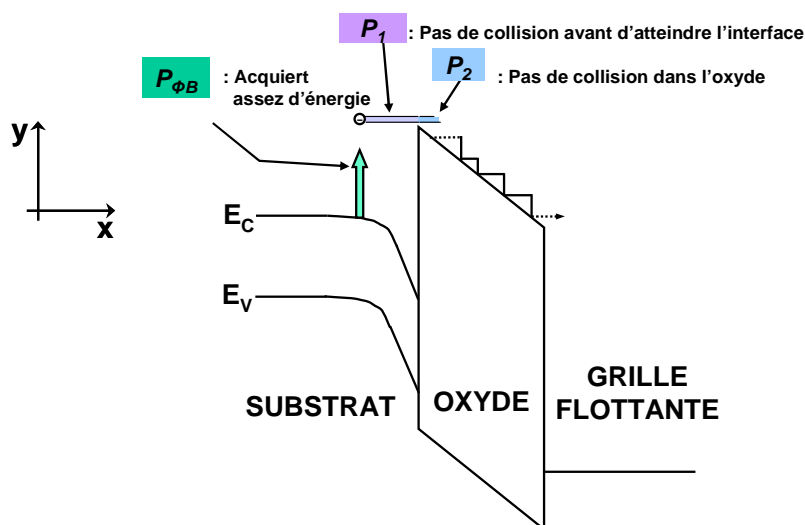


Fig. IV-35 : Schéma du diagramme des bandes en énergie décrivant les trois étapes de l'injection d'électrons chauds selon le modèle de « l'électron chanceux ».

Les expressions des trois probabilités sont les suivantes :

$$P_{\Phi_B} = 0,25 \cdot \left(\frac{E_x \cdot \lambda}{\phi_B} \right) \exp\left(- \frac{\phi_B}{E_x \cdot \lambda} \right) \quad \text{Équation IV-31}$$

Avec λ : libre parcours moyen des électrons, E_x : champ électrique latéral

$$P_1 \approx 1 - \alpha \exp(\alpha) E_1(\alpha) \quad \text{Équation IV-32}$$

Avec $\alpha = \frac{6kT}{q\lambda E_{ox}}$, k constante de Boltzmann, $E_{ox} \approx \frac{Vgs - Vds}{t_{tun}}$, t_{tun} : épaisseur de l'oxyde

tunnel et $E_1(z) = \int_z^\infty \frac{e^{-t}}{t} dt$

$$P_2 = \exp\left(- \frac{y_0}{\lambda_{ox}} \right) \quad \text{Équation IV-33}$$

Avec $y_0 = \sqrt{\frac{q}{16 \cdot \pi \epsilon_{ox} E_{ox}}}$

Le courant d'injection est la probabilité que les électrons soient injectés, multipliée par le courant de drain :

$$I_g = I_{ds} \int_0^L P_{\Phi_B} \cdot P_1 \cdot P_2 \frac{dx}{\lambda_r} \quad \text{Équation IV-34}$$

Avec L : longueur du canal, λ_r : libre parcours moyen entre deux collisions redirectrices, $\frac{dx}{\lambda_r}$

représente la probabilité qu'un électron soit redirigé sur une distance dx .

Bibliographie

- [Cappelletti'99] P. Cappelletti et al., "Flash Memories", Kluwer Academic Publishers, 1999.
- [Chang'84] C.Chang, "Tunneling in thin gate oxide MOS structure", Thèse de doctorat de l'Université de Berkeley, Californie, 1984.
- [Concannon'95] A. Concannon, F. Piccinini », A. Mathewson, C. Lombardi, "The Numerical Simulation of Substrate and Gate Currents in MOS and EPROMs", IEDM Tech. Dig., p. 289, 1995.
- [Cotrell'79] P.E. Cotrell, R.R. Troutman and T.H. Ning, 'Hot-electron emission in n-channel IGFET's", IEEE Trans. Electron Devices, vol. ED-26, no 4, pp. 520-532, 1979.
- [DeSalvo'00] B. DeSalvo, G. Ghibaudo, G. Pananakakis, B. Guillaumot, T. Baron, "Investigation of charging/discharging phenomena in nano-crystal memories", Superlatt. Microstruct., vol. 28, no. 5-6, pp. 339-344, 2000.
- [DeSalvo'01] B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, B. Guillaumot "Experimental and theoretical investigation of nanocrystals and nitride-trap memory devices", IEEE Trans. on El. Dev., Vol. 48, No. 8, p. 1789, 2001.
- [DeSalvoTh] B. De Salvo, "Etude du transport électrique et de la fiabilité des isolants des mémoires non volatiles à grille flottante", Thèse de doctorat INPG, 1999.
- [Tsividis'99] Y. Tsividis, "Operation and modeling of the MOS Transistor", 2nd ed. New York: Columbia Univ., p. 74, 1999.
- [Driussi'04] F. Driussi, D. Esseni, L. Selmi, "Performance, Degradation Monitors, and Reliability of the CHISEL Injection Regime", IEEE Trans. Dev.Mat. Rel., vol. 4, no. 3, p.327, 2004.
- [Fiegna'90] C. Fiegna, F. Venturi, E. Sangiorgi and B. Ricco, "Efficient non-local modelling of the electron energy distribution in sub-micron MOSFET's", IEDM Tech. Dig., p. 451, 1990.
- [Fiegna'91] Fiegna, F. Venturi, M. Melanotte, E. Sangiorgi and B. Ricco, "Simple and efficient modelling of EPROM writing", IEEE Trans. Electron Devices, vol. ED-38, no. 3, pp.603-610.
- [Frohman'69] D. Frohman-Betchkoswy, M. Lenzlinger, "Charge transport and storage in metal-nitride-oxide-silicon (MNOS) structures," J. Appl. Phys., vol. 408, no. 8, p. 3307, 1969.

- [Gasiorowicz'95] S. Gasiorowicz, "Quantum Physics", John Wiley & Sons, 1995.
- [Groeseneken'98] G. Groeseneken, H. E. Maes, J. VanHoudt, J. S. Witters, "Basics of nonvolatile semiconductor memory devices", in NVSM Technology. Piscataway, IEEE Press, 1998.
- [Hadjadj'01] A. Hadjadj, G. Salace, C. Petit, "Fowler-Nordheim Conduction in Polysilicon (n+)-Oxide-Silicon(p) Structures: Limit of the Classical Treatment in the Barrier Height Determination", J.Appl.Phys., vol. 89, no. 12, pp. 7994-8001, 2001.
- [Hagenbeck'04] R. Hagenbeck, S. Decker, F. Lau, P. Haibach, J.M. Schley, M. Isler, T. Mikolajick, G. Tempel, "Modeling and Simulation of Electron Injection during Programming in Twin Flash Devices Based on Energy Transport and the Non-Local Lucky Electron Concept", J. of Comp. Elec., vol. 3, pp. 239-242, 2004.
- [Hagenbeck'06] R. Hagenbeck et al., "Monte Carlo Simulation of Charge Carrier Injection in Twin Flash Memory Devices during Program and Erase", Proc. of SISPAD, pp. 322-325, 2006.
- [Hu'79] C. Hu, "Lucky-electron model for channel hot-electron emission", IEDM Tech. Dig., p. 22, 1979.
- [Ingrosso'02] G. Ingrosso, L. Selmi and E. Sangiorgi, "Monte Carlo simulation of program and erase charge distribution in NROM devices", Proc. Of ESSDERC, p. 187, 2002.
- [Jungemann'97] C. Jungemann, S. Yamaguchi, H. Goto, "Investigation of the influence of impact ionization feedback on the spatial distribution of hot carriers in a nMOSFET," Proc. of ESSDRC., 1997.
- [Kencke'98] D. L. Kencke, X.Wang, H.Wang, Q. Ouyang, S. Jallepally, M. Rashed, C. Maziar, A. Tasch Jr., S. K. Banerjee, "The origin of secondary electron gate current: A multiple stage Monte Carlo study for scaled, low power flash memory," IEDM Tech. Dig., p. 889, 1998.
- [Landau'65] L.D. Landau, E.M. Lifshitz, "Quantum mechanics: non-relativistic theory", Volume 3 de Course of Theoretical Physics, Second Edition, Pergamon Press, Oxford, 1965.
- [Landau'84] L.D. Landau, E.M. Lifshitz, L.P. Pitaevskii, "Electrodynamics of Continuous Media", 2nd ed. London, U.K., Butterworth-Heinemann, 1984.
- [Larcher'02] L. Larcher and P. Pavan, "A new analytical model of Channel Hot Electron (CHE) and Channel Initiated Secondary Electron (CHISEL) current suitable for compact modelling", International Conference on Modeling and Simulation of Microsystems Tech. Proc., 2002.

- [Nagano'94] S. Nagano, M. Tsukiji, E. Hasegawa, A. Ishitani, "Mechanism of Leakage Current Through the Nanoscale SiO₂ Layer", *J.Appl.Phys.*, vol. 75, no. 7, pp. 3530-3535, 1994.
- [O'Dwyer'73] J. J. O'Dwyer, "The Theory of Electrical Conduction and Breakdown in Solids Dielectric", Oxford, U.K. : Clarendon, 1973.
- [Perniola'06] L. Perniola, G. Iannaccone, G. Ghibaudo, "Subthreshold behavior of dual-bit nonvolatile memories with very small regions of trapped charge", *IEEE Tech. Nano.*, vol. 5, no. 4, p. 373, 2006.
- [Roux'92] O. Roux, G. Ghibaudo, J. Brini, « Model for drain current RTS amplitude in small area MOS transistors », *Solid State Elec.*, vol. 35, p. 1273, 1992.
- [Schuegraf'92] K. Schuegraf, C. King, C. Hu, "Ultra thin silicon dioxide leakage current and scaling limit," in *Symp. VLSI Technol. Dig. Tech. Papers*, pp. 18–19, 1992
- [Selberherr'80] S. Selberherr, A. Shutz and H.W. Potzl, "MINIMOS-A two-dimensional MOS transistor analyzer", *IEEE Trans. Electron Devices*, vol. ED-27, pp. 1540-1550, 1980.
- [Selberherr'84] S. Selberherr, "Analysis and Simulation of Semiconductor Devices", Wien, New York: Springer-Verlag, 1984.
- [Silvaco] <http://www.silvaco.com/>
- [Shockley'61] W. Shockley, "Problems related to p-n junctions in silicon", *Solid-State Electron.*, vol. 2, p. 35, 1961.
- [Stratton'62] R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers", *Phys. Rev.*, vol. 126, no. 6, p. 2002, 1962.
- [Stratton'69] R. Stratton, "Tunneling in Schottky barrier rectifiers", *Tunneling Phenomena in Solids*. New York : Plenum, pp. 105–125, 1969.
- [Stratton'72] R. Stratton, "Semiconductor Current-Flow Equations (Diffusion and Degeneracy)", *IEEE Trans. Electron Devices*, vol. 19, no. 12, pp. 1288-1292, 1972.
- [Sun'04] L. Sun et al., "Effects of CHE and CHISEL Programming Operation on the Characteristics of SONOS Memory", *Proc. of ICSICT*, p. 695, 2004.
- [Tam'84] S. Tam, P.K. Ko, C. Hu, "Lucky-electron model of channel hot-electron injection in MOSFET's", *IEEE Trans. Electron Devices*, vol. ED-31, no. 9, 1984.
- [Tempel'06] G. Tempel, R. Hagenbeck, M. Strassburg, "Quantitative Model for Data Retention Loss at NROM Nitride Charge Trapping Devices after Program / Erase Cycling", *Proc. of NVSMW*, pp. 78-80, 2006.

- [Tsu'73]** R. Tsu, L. Esaki, "Tunneling in a Finite Superlattice", Appl. Phys. Lett., vol. 22, no. 11, pp. 562-564, 1973.
- [Weinberg'77]** Z. A. Weinberg, "Tunneling of electrons from Si into thermally grown SiO₂", Solid State Electron., vol. 20, no. 11, 1977.

CONCLUSIONS ET PERSPECTIVES

Ce travail a porté sur l'intégration, la caractérisation et la modélisation des mémoires non-volatiles à nanocristaux de silicium. Un démonstrateur complet Flash NOR de 32 Mb à nanocristaux de silicium a été réalisé à partir d'un produit ATMEL. Nous avons présenté un ensemble de mesures expérimentales illustrant de manière exhaustive les performances électriques de la cellule simple et les caractéristiques statistiques des matrices mémoires à nanocristaux de silicium. Des modèles analytiques et des simulations TCAD ont permis la compréhension des résultats expérimentaux.

Dans le **chapitre I**, nous avons montré l'essor considérable que le marché des mémoires non-volatiles et en particulier celui des mémoires Flash a connu ces dernières années, tiré par les applications portables grand public. Nous avons montré que les dispositifs mémoires conventionnels à grille flottante avaient des difficultés à pousser davantage la miniaturisation, notamment avec la diminution de l'épaisseur du diélectrique tunnel qui a atteint aujourd'hui sa taille minimale. Afin de prolonger la viabilité de ces dispositifs, une solution est le remplacement de la grille flottante continue par des sites de piégeage discrets, tels qu'une couche de nitrure (SONOS, TANOS, NROM) ou bien des nanocristaux de silicium qui présentent une bonne immunité face aux défauts dans l'oxyde tunnel. Cependant, les mémoires à base de nitrure présentent des inconvénients que n'ont pas les mémoires à nanocristaux de silicium. C'est pourquoi ces dernières qui ont fait l'objet de ce manuscrit. Nous avons enfin effectué une revue des réalisations de mémoires à Si-NCs faites par les industriels ces dernières années.

Dans le **chapitre II**, nous avons d'abord présenté les différentes méthodes de fabrication des nanocristaux de silicium. La technique que nous avons choisie pour la fabrication des nanocristaux de silicium dans nos dispositifs est la croissance par LPCVD (Low Pressure Chemical Vapor Deposition). Dans la deuxième partie du chapitre, nous avons présenté l'intégration des nanocristaux dans un produit ATMEL Flash NOR 32 Mb, basé sur une technologie 130 nm. A partir du procédé de fabrication d'une mémoire Flash standard à grille flottante continue, nous avons étudié les différentes façons d'intégrer les nanocristaux de silicium dans le procédé de fabrication à la place de la grille flottante en polysilicium. Un des points clés de l'intégration est de limiter l'oxydation des Si-NCs, les étapes d'oxydation les plus critiques étant la croissance des oxydes de grille des transistors de périphérie. Deux procédés différents ont été réalisés. Dans le premier, les oxydes de grille des transistors sont réalisés après le dépôt des Si-NCs. Cette solution est celle qui se rapproche le plus du procédé standard, mais les nanocristaux risquent d'être oxydés. Un deuxième procédé plus complexe, mais permettant d'éviter toute oxydation parasite des Si-NCs a donc été mis en place. Dans ce cas, les nanocristaux sont déposés après avoir réalisé les oxydes de grille des transistors de la périphérie.

Enfin, nous avons démontré que trois niveaux de masques pourraient être économisés, entraînant une réduction non négligeable des coûts de fabrication.

Dans le **chapitre III**, nous nous sommes intéressés à la caractérisation électrique des mémoires à nanocristaux de silicium. Nous avons d'abord présenté les résultats électriques obtenus sur des cellules mémoires simples. Une étude exhaustive de l'influence des méthodes de programmation ainsi que des paramètres technologiques sur les caractéristiques d'écriture par électrons chauds et d'effacement par Fowler-Nordheim a été réalisée. On a montré que l'on pouvait obtenir une fenêtre de programmation de 4 V avec un temps d'écriture de 10 μ s en utilisant des conditions d'écriture adéquates. Dans une deuxième partie, nous nous sommes intéressés aux caractéristiques électriques de matrices mémoires. Pour la première fois, un démonstrateur Flash NOR 32 Mb avec la logique CMOS de périphérie fonctionnelle a été réalisé. Une fenêtre de programmation de plus de 3 V a été obtenue dans les conditions de programmation standard du produit Flash. Il est possible d'augmenter cette fenêtre jusqu'à 4 V et surtout de séparer suffisamment les distributions effacées et écrites en modifiant les conditions de programmation, ce qui est similaire, voir meilleur que ce qui a été présenté dans la littérature (chapitre I). Le contrôle de la taille des Si-NCs permet également de rendre les distributions des tensions de seuil plus étroites. Enfin, une étude de fiabilité a été réalisée, démontrant la robustesse des dispositifs à Si-NCs contre les défauts générés par les stress d'écriture/effacement dans l'oxyde.

Le **chapitre IV** traite de la modélisation des mémoires Flash à nanocristaux de silicium. Dans une première partie, nous avons présenté le modèle dit de la quasi-grille flottante [DeSalvo'01] qui nous a permis de simuler l'effacement Fowler-Nordheim et le « gate disturb ». Nous avons montré qu'un bon accord était obtenu entre les mesures et les simulations. Concernant l'effacement, ces modélisations ont permis de cerner l'influence des épaisseurs de l'oxyde tunnel et de contrôle, ainsi que l'influence du taux de couverture des Si-NCs. Concernant le gate disturb, nous avons mis en évidence l'influence de la tension de grille de lecture et de l'épaisseur du HTO. Les simulations ont prédit qu'un HTO de 12 nm au minimum était nécessaire pour minimiser l'effet du « gate disturb » après de 10 ans de lecture. Dans la deuxième partie du chapitre, nous nous sommes intéressés à l'écriture par électrons chauds et en particulier à la localisation de la charge dans les nanocristaux. En effet, dans le cas d'une mémoire à pièges discrets, la charge injectée près du drain reste localisée au-dessus de la jonction canal/drain. Nous avons étudié l'influence des conditions d'écriture sur la localisation de la charge à l'aide de simulations TCAD et d'un modèle analytique [Perniola'06] couplé à des mesures expérimentales. La région de localisation des électrons injectés reste la même quelque soit la tension de grille d'écriture. Au contraire, lorsque le temps d'écriture augmente ou lorsqu'une tension négative est appliquée sur le substrat pendant l'écriture, la charge devient moins localisée et l'injection est décalée en direction du canal.

Nous avons donc montré que les avantages majeurs des mémoires à Si-NCs étaient les suivants :

- La compatibilité du procédé Si-NCs avec le procédé CMOS. La partie mémoire est donc facile à intégrer avec de la logique CMOS. Nous avons également montré dans le chapitre II qu'il était possible d'économiser plusieurs niveaux de masques avec le procédé Si-NCs par rapport au procédé de la Flash standard, ce qui permet de réduire le coût de fabrication par plaque tout en ayant des performances similaires.

- La réduction des dimensions est plus facile. Du point de vue de la fabrication, la grille flottante n'est pas définie par lithographie et du point de vue électrique, le couplage capacitif entre cellules voisines est fortement réduit. On peut donc atteindre de grandes densités d'intégration.

- La rapidité de l'effacement de la cellule (de l'ordre de 1 ms).

- La rapidité de l'effacement complet de la matrice mémoire. Une particularité est la saturation de la tension de seuil au niveau neutre au cours de l'effacement, contrairement à la cellule à grille flottante continue qui peut être sur-effacée. Ceci permet de simplifier l'algorithme d'effacement de trois étapes (pré-programmation, effacement et programmation douce) à une seule étape (effacement). L'effacement complet de la matrice mémoire est donc finalement beaucoup plus rapide que celui d'une Flash standard.

- La robustesse des dispositifs à Si-NCs contre les défauts générés par les stress d'écriture/effacement dans l'oxyde. Les mesures sur matrices mémoires après endurance (10K cycles), rétention à 150°C et rétention à 150°C après endurance n'ont montré aucun bit erratique.

Ces avantages correspondent aux caractéristiques requises pour les applications microcontrôleurs embarqués (**Tableau 1**).

Application	Stockage de données	Code
Capacité	8 Kb-1 Mb	256 Kb-16 Mb
Programmation rapide	Obligatoire pour des mises à jour fréquentes	Souhaitable pour minimiser les coûts de test
Endurance	Jusqu'à 1M de cycles	10K cycles

Tableau 1 : Caractéristiques des applications microcontrôleurs embarqués.

On peut souligner que les mémoires à Si-NCs ont aussi quelques faiblesses. Tout d'abord, la saturation de la tension de seuil effacée qui permet de simplifier l'algorithme d'effacement constitue aussi un inconvénient pour l'endurance. En effet, l'augmentation de la tension de seuil effacée au cours de l'endurance ne peut pas être compensée par le sur-effacement, comme c'est le cas pour la Flash standard. Nous avons montré que l'on pouvait atteindre 10K cycles, ce qui est suffisant pour les applications embarquées dédiées au code (**Tableau 1**). Toutefois, pour des applications embarquées de stockage de données qui nécessitent d'atteindre les 1M de cycles, il semble nécessaire d'optimiser les conditions de programmation et les paramètres technologiques de la cellule afin d'obtenir de meilleures caractéristiques d'endurance. Une autre solution pourrait être d'augmenter la fenêtre de programmation afin de garantir une marge suffisante et ce, après la dérive de la tension de seuil effacée liée à l'endurance. Cependant, la fenêtre de programmation reste une des principales limitations des mémoires à Si-NCs. Pour résoudre ce problème, plusieurs technologies sont actuellement étudiées : les nanocristaux métalliques [**Lui'02-a**] [**Liu'02-b**] et des templates organiques pour former des matrices ordonnées de nanocristaux [**Guarini'03**].

Bibliographie

- [DeSalvo'01]** B. De Salvo, G. Ghibaudo, G. Pananakakis, P. Masson, T. Baron, N. Buffet, A. Fernandes, B. Guillaumot « Experimental and theoretical investigation of nanocrystals and nitride-trap memory devices », IEEE Trans. On El. Dev., Vol. 48, No. 8, p. 1789, 2001.
- [Guarini'03]** Guarini et al., “Low voltage, scalable nanocrystal FLASH memory fabricated by templated self assembly”, IEDM Tech. Dig., p.541, 2003.
- [Lui'02-a]** Z. Liu et al., “Metal Nanocrystal Memories—Part I: Device Design and Fabrication”, IEEE Tr. on El. Dev, Vol. 49, No. 9, p. 1606, 2002.
- [Liu'02-b]** Liu et al., “Metal Nanocrystal Memories—Part II: Electrical Characteristics”, Z. IEEE Tr. on El. Dev., Vol. 49, No. 9, p. 1614, 2002.
- [Perniola'06]** L. Perniola, G. Iannaccone, G. Ghibaudo, “Subthreshold behavior of dual-bit nonvolatile memories with very small regions of trapped charge”, IEEE Tech. Nano., vol. 5, no. 4, p. 373, 2006.

BIBLIOGRAPHIE DE L'AUTEUR

Conférences et articles de revues

- G. Molas, D. Deleruyelle, B. De Salvo, G. Ghibaudo, M. Gely, **S. Jacob**, D. Lafond et S. Deleonibus, "Impact of few electron phenomena on floating-gate memory reliability", IEEE IEDM, pp. 877-880, 2004, San Francisco, USA.
- **S. Jacob**, B. De Salvo, G. Le Carval, S. Deleonibus, "Modeling of a Double-Gate FinFlash memory", ICMTD (International Conference on Memory Technology and Design), pp. 153-156, 2005, Giens, France.
- **S. Jacob**, L. Perniola, P. Scheiblin, E. Jalaguier, B. De Salvo, F. Boulanger, S. Deleonibus, G. Festes, R. Coppard, "TCAD Modeling and Data of NOR Nanocrystal Memories", NVMTS (Non Volatile Memory Technology Symposium), pp. 31-33, 2006, San Mateo, USA.
- **S. Jacob**, L. Perniola, B. De Salvo, E. Jalaguier, G. Festes, R. Coppard, F. Boulanger, S. Deleonibus, "On the localization of the trapped charges in Silicon nanocrystal NOR Flash devices", Proc. of ICMTD, pp. 247-250, 2007, Giens, France.
- **S. Jacob**, L. Perniola, G. Festes, S. Bodnar, R. Coppard, J.F. Thiery, T. Pedron, E. Jalaguier, F. Boulanger, B. De Salvo, S. Deleonibus "Investigation of Reliability Characteristics of Si Nanocrystal NOR Memory Arrays", NVSMW (Non Volatile Semiconductor Memory Workshop), pp. 71-72, 2007, Monterey, USA.
- **S. Jacob**, G. Festes, S. Bodnar, R. Coppard, J.F. Thiery, T. Pate-Cazal, T. Pedron, B. De Salvo, L. Perniola, E. Jalaguier, F. Boulanger, S. Deleonibus, "Integration of CVD Silicon Nanocrystals in a 32Mb NOR Flash Memory", ESSDERC (European Solid State Device Research Conference), pp. 410-413, 2007, Munich, Allemagne.
- **S. Jacob**, B. De Salvo, L. Perniola, G. Festes, S. Bodnar, R. Coppard, J.F. Thiery, T. Pate-Cazal, C. Bongiorno, S. Lombardo, J. Dufourcq, E. Jalaguier, T. Pedron, F. Boulanger, S. Deleonibus, "Integration of CVD Silicon Nanocrystals in a 32Mb NOR Flash Memory", papier ESSDERC sélectionné pour un article dans Solid-State Electronics : accepté.

Autres communications

S. Jacob, « Intégration, caractérisation et modélisation des mémoires non volatiles avancées à sites de stockage discrets », présentation d'un poster aux journées de l'Ecole Doctorale, 22 janvier 2007, Marseille.

TITRE : Intégration, caractérisation et modélisation des mémoires non-volatiles à nanocristaux de silicium

Résumé :

Depuis une vingtaine d'années, l'industrie de la microélectronique et en particulier le marché des mémoires non-volatiles connaît une évolution considérable, en termes d'augmentation de la capacité d'intégration et de diminution du prix de revient. Ceci a permis au grand public d'accéder aux produits électroniques (téléphones portables, baladeurs MP3, clés USB, appareils photos numériques...) qui connaissent actuellement un énorme succès. Cependant, la miniaturisation des mémoires Flash risque de rencontrer des limitations. C'est pourquoi les industriels et les laboratoires recherchent actuellement de nouvelles voies qui permettraient de prolonger la durée de vie de ces dispositifs. Dans ce contexte, l'objectif premier de cette thèse est l'étude expérimentale et théorique des mémoires non-volatiles à nanocristaux de silicium. Nous avons montré les différentes possibilités d'intégration des nanocristaux de silicium à partir d'un procédé de fabrication standard. Un démonstrateur Flash NOR 32 Mb à nanocristaux de silicium a été réalisé à partir d'un produit ATMEL. Nous nous sommes ensuite intéressés à la caractérisation électrique des cellules et matrices mémoires. Une étude exhaustive de l'influence des conditions de programmation ainsi que des paramètres technologiques sur les performances électriques a été menée. La modélisation de l'effacement Fowler-Nordheim et du « gate disturb » a permis de comprendre l'influence de certains de ces paramètres. Concernant l'écriture par porteurs chauds, nous avons étudié l'influence des conditions d'écriture sur la localisation de la charge à l'aide de simulations TCAD et d'un modèle analytique couplé à des mesures expérimentales.

MOTS-CLES

Microélectronique, mémoire Flash, non-volatile, grille flottante, nanocristaux de silicium, intégration, caractérisation électrique, modélisation, NOR, effacement Fowler-Nordheim, écriture par électrons chauds.

TITLE : Integration, characterization and modeling of silicon nanocrystal non volatile memories

Abstract:

Over the last 20 years, the industry of microelectronics and particularly the non-volatile memory market has known a considerable growth, in terms of integration capacity increasing and cost reduction. Consumers have been able to access to electronic products (mobile phones, MP3 players, flash drives, digital cameras...) which are currently very successful. However, scaling of standard Flash memories will face in a near future several limitations. Consequently, new paths are investigated in order to push the scaling limits of these devices. Within this context, the main purpose of this PhD is the experimental and theoretical study of non-volatile silicon nanocrystal memories. First, several options of silicon nanocrystal integration using a standard process have been shown. A 32Mb NOR silicon nanocrystal Flash memory demonstrator has been fabricated from an ATMEL product. Then, electrical characterization of memory cells and arrays has been performed. An exhaustive study of the influence of programming conditions and technological parameters has been carried out. The influence of some parameters has been understood through modeling of Fowler-Nordheim erasing and gate disturb. Finally, the localization of the trapped charges in silicon nanocrystal devices written by Hot Electron injection has been investigated through TCAD simulations and an exhaustive set of experimental data explained by an analytical model.

KEY WORDS

Microelectronics, Flash memory, non-volatile, floating gate, silicon nanocrystals, integration, electrical characterization, modeling, NOR, Fowler-Nordheim erasing, hot electron writing.