

Table des matières

Remerciements.....	iii
Table des matières	1
Introduction générale	7
CHAPITRE I : LES CELLULES MÉMOIRES DYNAMIQUES	11
I.1. Les technologies mémoires.....	12
I.1.1 Mémoires Non Volatiles.....	13
I.1.2 Mémoires Volatiles	14
I.1.2.1. Cellule SRAM.....	14
I.1.2.2. Cellule DRAM.....	14
I.1.3 Point mémoire	15
I.1.4 Description d'un point mémoire sous matrice.....	16
I.2. Mémoires Dynamiques 1T – 1C	17
I.2.1. Réduction des dimensions du transistor d'accès	18
I.2.2. Réduction des dimensions de la capacité de stockage.....	19
I.3. La cellule mémoire 1T – DRAM	21
I.3.1. Transistor MOSFET.....	21
I.3.1.1. Architecture BULK	21
I.3.1.2. Transistor MOSFET PDSOI.....	29
I.3.1.3. Transistor MOSFET FDSOI.....	32
I.3.1.4. Transistor MOSFET NANOFIL	32
I.3.2. Mémoires 1T – DRAM	34
I.3.2.1. Programmation du '1' : Ionisation par Impact (II)	35
I.3.2.2. Programmation du '1' : transistor à jonction bipolaire (BJT)	36
I.3.2.3. Programmation du '1' : boucle de rétroaction positive de la tension seuil du transistor (V_{TH}).....	37
I.3.2.4. Programmation du '1' : Effet Tunnel Bande à Bande (B2B).....	37
I.3.2.5. Programmation du '0' ou Effacement.....	37
I.4. Pourquoi l'architecture A2RAM comme 1T – DRAM	39
I.4.1. Mémoire Z-RAM.....	39
I.4.2. Architecture mémoire 1T – DRAM de type FDSOI.....	39
I.4.2.1. Architecture mémoire MSDRAM.....	39
I.4.2.2. Architecture mémoire ARAM.....	40
I.4.2.3. Architecture mémoire A2RAM	41
I.5. Conclusion.....	43

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD	44
II.1. Stratégie de simulation.....	45
II.1.1 Description du logiciel de simulation.....	45
II.1.2 Structure A2RAM simulée.....	46
II.1.3 Chronogramme de tensions pour l'étude de la mémoire A2RAM	47
II.1.4 Définition des modèles physiques pour la simulation électrique d'opérations mémoires.....	49
II.1.5 Analyse des opérations mémoires par simulations TCAD	51
II.1.6 Comparaisons entre résultats de simulation TCAD et caractérisation électrique	53
II.2. Analyse de la sensibilité des performances mémoires : variation des polarisations .	57
II.2.1 Variation des tensions de grille face avant et de drain pendant la programmation : V_{gW1} et V_{dW1}	57
II.2.2 Variation de la tension de grille face avant pendant la phase de maintien : V_{gHold}	60
II.2.3 Variation de la tension de grille face avant pendant la lecture : V_{gRead}	61
II.2.4 Variation des tensions de grille face avant pendant la lecture et la phase de maintien avec $V_{gRead}=V_{gHold}$	64
II.2.5 Variation de la tension de drain pendant la lecture : V_{dRead}	67
II.2.6 Variation de la polarisation de grille arrière V_b	68
II.2.7 Variation de la tension de grille face avant pendant l'effacement : V_{gErase}	70
II.3. Conclusion.....	71
CHAPITRE III : OPTIMISATION DE LA CELLULE A2RAM PAR UNE ETUDE DE SENSIBILITE TCAD	72
III.1. Mise en évidence du nouveau mécanisme de programmation de l'état '1' dans cellule A2RAM	73
III.1.1. Présentation du nouveau mécanisme de programmation de l'état '1'	73
III.1.2. Impact des modèles de transport de la charge sur le nouveau mécanisme d'écriture de l'état '1'	80
III.1.3. Tentative de mise en évidence expérimentale du nouveau mécanisme de programmation de l'état '1' de l'A2RAM	82
III.2. Étude des variations des paramètres technologiques d'A2RAM sur ses performances	84
III.2.1. Réduction de la longueur de la grille (L_g).....	85
III.2.2. Variation de l'épaisseur d'oxyde de la grille face avant (EOT).....	86
III.2.3. Variation de l'épaisseur du body : T_{body}	87

III.2.4.	<i>Variation de l'épaisseur du bridge : T_{bridge}</i>	89
III.2.5.	<i>Variation du dopage du bridge : N_{bridge}</i>	90
III.3.	D'autres voies d'optimisation de performances de l'A2RAM	93
III.3.1.	<i>Intégration d'une mono couche de bore pur dans une cellule A2RAM : la structure A2RAM B++</i>	93
III.3.2.	<i>Hétérostructures et performances mémoires de l'A2RAM : cas du SiGe</i>	99
III.4.	Conclusion	104
CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM		105
IV.1.	Technique d'extraction des paramètres technologiques de l'A2RAM	106
IV.1.1.	<i>Étude de la caractéristique C_{gg}-V_g de la cellule A2RAM</i>	106
IV.1.2.	<i>Extraction du dopage du bridge</i>	109
IV.1.3.	<i>Extraction de l'épaisseur du body et du bridge</i>	116
IV.2.	Validation expérimentale de la technique de caractérisation du profil de dopage de la cellule A2RAM	118
IV.2.1.	<i>Description des échantillons mesurés</i>	118
IV.2.2.	<i>Extraction de profils de dopage sur des échantillons A2RAM par caractérisation électrique</i>	119
IV.2.3.	<i>Analyse des performances mémoires A2RAM</i>	120
IV.3.	Conclusion	124
CHAPITRE V : MODELISATION DE LA CELLULE MEMOIRE A2RAM		125
V.1.	Modèle compact	126
V.1.1.	<i>Qu'est-ce qu'un modèle compact ?</i>	126
V.1.2.	<i>Modèle compact de l'A2RAM : Macro-modélisation</i>	126
V.2.	Modélisation compacte du comportement DC de l'A2RAM	127
V.2.1.	<i>Stratégie de modélisation</i>	127
V.2.2.	<i>Modélisation de la résistance du bridge de l'A2RAM</i>	128
V.3.	Modélisation de la tension de seuil du bridge	131
V.3.1.	<i>Expression potentiel électrostatique dans le film de silicium</i>	131
V.3.2.	<i>Validation par TCAD</i>	133
V.3.3.	<i>Modélisation V_{thb} d'une cellule A2RAM</i>	134
V.3.4.	<i>Validation de la tension seuil du bridge V_{thb} par TCAD</i>	135
V.3.4.1.	<i>Variation du dopage du bridge (N_{bridge})</i>	135
V.3.4.2.	<i>Variation de l'épaisseur totale du film de silicium (T_{sc})</i>	136
V.3.4.3.	<i>Variation de la polarisation de grille face arrière (V_b)</i>	137

Table des matières

V.4.	Modèle compact DC de l'A2RAM.....	138
V.4.1.	<i>Extraction carte modèle du transistor sur substrat SOI.....</i>	<i>138</i>
V.4.2.	<i>Validation du modèle quasi statique de l'A2RAM.....</i>	<i>138</i>
V.5.	Modélisation d'opérations mémoires de l'A2RAM.....	139
V.5.1.	<i>Circuit équivalent : stratégie de modélisation de l'effacement et de l'écriture</i> <i>139</i>	
V.5.2.	<i>Modélisation de l'opération de lecture de l'information : stratégie.....</i>	<i>141</i>
V.5.3.	<i>Extraction du paramètre SWMEMV.....</i>	<i>143</i>
V.5.4.	<i>Validation de la modélisation des opérations mémoires.....</i>	<i>144</i>
V.5.5.	<i>Etude des opérations mémoires d'A2RAM dans une matrice 2x2: simulation</i> <i>SPICE 145</i>	
V.6.	Conclusion.....	147
CHAPITRE VI : FABRICATION D'UNE CELLULE A2RAM NANOFIL TRIPLE-GRILLES		148
VI.1.	Intégration de la cellule A2RAM de type 3D	149
VI.1.1.	<i>Etapas de fabrication de la cellule A2RAM 3D</i>	<i>149</i>
VI.2.	Simulations prédictives TCAD	152
VI.2.1.	<i>Dimensionnement de la longueur de la grille et des épaisseurs du body et du</i> <i>bridge 152</i>	
VI.2.2.	<i>Dimensionnement des zones de recouvrement source et drain.....</i>	<i>154</i>
VI.2.3.	<i>Evaluation de la concentration optimale du germanium dans SiGe.....</i>	<i>157</i>
VI.2.4.	<i>Dimensionnement de recette d'implantation du bridge</i>	<i>158</i>
VI.3.	Résumé des choix des procédés pour la fabrication des cellules A2RAM 3D	161
VI.4.	Caractérisation électrique de la cellule A2RAM de type 3D	162
VI.4.1.	<i>Caractéristique statique de la cellule A2RAM 3D</i>	<i>163</i>
VI.4.2.	<i>Caractéristique petit signal de la cellule A2RAM 3D</i>	<i>164</i>
VI.4.3.	<i>Essai d'extraction du dopage du bridge</i>	<i>167</i>
VI.4.4.	<i>Analyse des performances mémoires</i>	<i>169</i>
VI.4.4.1.	<i>Caractérisation de l'écriture par GIDL sur nos échantillons.....</i>	<i>171</i>
VI.4.4.2.	<i>Lecture de l'état '1' : impact de l'énergie d'implantation du bridge</i>	<i>173</i>
VI.4.4.3.	<i>Lecture de l'état '1' : impact de la longueur de la grille.....</i>	<i>174</i>
VI.4.4.4.	<i>Lecture de l'état '1' : impact de l'épaisseur du bridge et/ou du body.....</i>	<i>175</i>
VI.4.4.5.	<i>Lecture de l'état '1' : impact de l'implantation LDD</i>	<i>176</i>
VI.4.4.6.	<i>Lecture de l'état '1' : impact de la polarisation face arrière.....</i>	<i>177</i>
VI.5.	Conclusion.....	178
Conclusion générale		180

Table des matières

BIBLIOGRAPHIE.....	186
PUBLICATIONS DE L'AUTEUR	197



INTRODUCTION GENERALE

L'industrie de la microélectronique continue de suivre la loi de Moore, qui prévoit que la surface des circuits intégrés doit être réduite de moitié tous les deux ans. Ceci se traduit par une division par 2 de la surface occupée pour les composants logiques comme les transistors MOS, mais aussi pour les points mémoires. Ces derniers peuvent être composés de plusieurs transistors (SRAM) ou plusieurs dispositifs (une DRAM qui est composée d'un transistor et d'une capacité).

Les mémoires sont classées dans 2 familles distinctes :

- Les mémoires volatiles qui perdent l'information stockée une fois l'alimentation coupée, et peuvent être statique (SRAM) ou dynamique (DRAM). Elles sont utilisées comme mémoire vive et/ou cache dans tout système microélectronique, et sont donc proches du microprocesseur.
- Les mémoires non volatiles qui gardent l'information stockée indéfiniment même lorsqu'elles sont hors alimentation. Elles sont utilisées comme mémoire secondaire dans tout système microélectronique et permettent de stocker une grande quantité d'information mais avec des temps d'accès élevés.

La latence causée par les interactions entre le microprocesseur et la mémoire est devenue tout aussi importante, sinon plus, pour la performance du système entier que la vitesse des circuits logiques. Pour résoudre cela, les mémoires sont organisées en hiérarchie. Partant d'une mémoire centrale pouvant stocker un très grand nombre de données mais avec des temps d'accès élevés ($> \text{ms}$) à la mémoire cache pouvant stocker un nombre très limité de données mais un temps d'accès court ($\sim \text{ns}/\mu\text{s}$) comparable au temps d'exécution d'une instruction par le microprocesseur. Puisque que les mémoires caches sont les plus proches du microprocesseur, leur densité d'intégration et leur vitesse d'accès doivent d'être maximisées. Pour les processeurs multicœurs et multithreads d'aujourd'hui, le défi majeur est d'avoir une mémoire cache à forte densité d'intégration avec une faible consommation d'énergie tout en maintenant un faible temps d'accès. Traditionnellement, la mémoire SRAM était utilisée couramment pour la mémoire cache car elle permettait d'avoir des temps d'accès rapides. Cependant, étant constitué de six transistors, elle occupe une surface substantielle sur la puce et consomme un courant important. La figure i.1.a montre la proportion de la surface occupée par un cache niveau 3 (L3) à base de cellules SRAM dans un processeur Intel i-7. La surface de la mémoire SRAM sur la puce est aussi importante que celle occupée par les circuits logiques.

D'un autre côté, la cellule DRAM intégrée (eDRAM) utilise un transistor et un condensateur. Elle garantit donc 2.5 à 3 fois plus de mémoires intégrées pour le même espace de silicium, avec pour avantages supplémentaires une faible dissipation de puissance et un taux d'erreur 1 000 fois inférieur [G. Fredeman 16]. La figure i.1.b montre une organisation d'un processeur IBM Power 7 où la eDRAM est utilisée au niveau de la mémoire cache. Chaque cœur du processeur a sa propre mémoire cache, ce qui réduit considérablement les temps d'accès mémoire et la vitesse d'exécution des instructions. De plus, la surface occupée par la cellule eDRAM est inférieure à celle de la figure i.1.a. Le gain apporté par la surface occupée sur le silicium par la cellule eDRAM est un atout majeur pour optimiser les performances des processeurs.

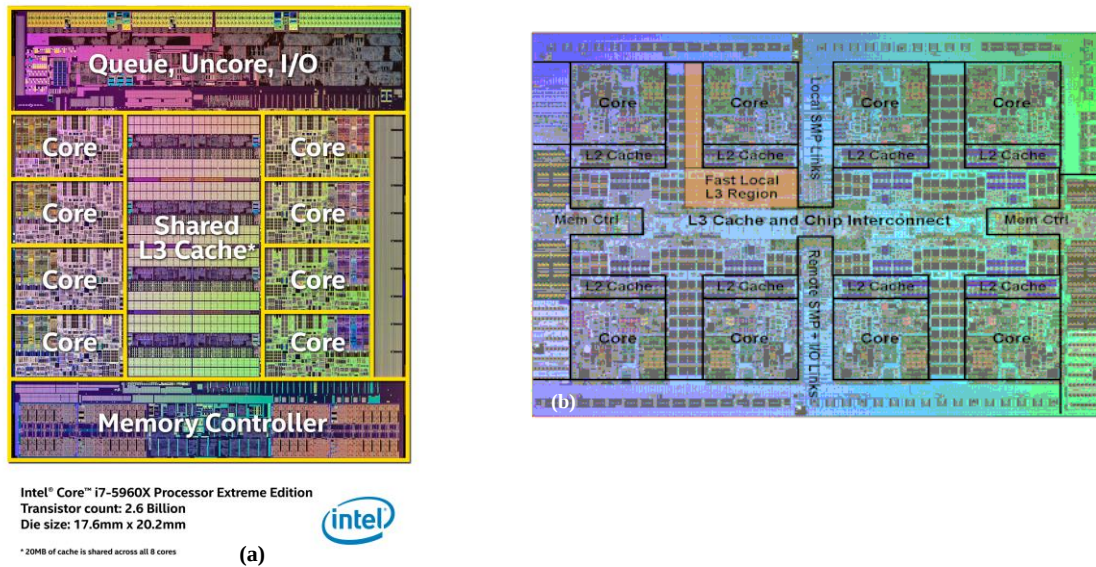


Figure i.1 : (a) Image d'un processeur Intel i7-5960X utilisant un bloc mémoire SRAM en cache. (b) Image d'un processeur IBM Power 7 avec un bloc mémoire eDRAM en cache.

Cependant, comme pour tout composant électronique, les fabricants de cellules DRAM sont confrontés à un défi de taille : réduire les dimensions de la cellule mémoire lorsque la longueur caractéristique de la technologie utilisée (F) continue de diminuer. Les cellules DRAM actuelles intègrent une cellule 1 T / 1 C ayant une aire de $6F^2$ [Park 15], et le principal problème de la réduction de la taille de la cellule réside au niveau de l'intégration du condensateur de stockage qui doit rester à une valeur constante de capacité [SZE 06]. Les challenges technologiques que doivent faire face les industriels pour aller vers une cellule DRAM avec une aire de $4F^2$ sont les suivants [IRTS 15] :

- Maintenir la capacité constante tout en réduisant sa longueur caractéristique (F),
- Maintenir une faible fuite du transistor d'accès et du condensateur de stockage,
- Maintenir une faible résistivité sur les lignes de mots et de données,
- Augmenter la densité de stockage tout en maintenant un faible coût de production.

Une solution pour contourner les problèmes liés à l'intégration de la capacité de stockage est de se tourner vers des architectures mémoires sans capacité de stockage : 1T-DRAM. Pour ces architectures mémoires, l'information est stockée dans le transistor et l'accès à l'information se fait via le même transistor. [Wann 93] propose pour la première fois le concept de mémoire 1T-DRAM issu des études de simulation, mais ce n'est qu'en 2001 que le concept a été démontré expérimentalement [Okhonin 01]. Cependant, la dégradation des performances de cette architecture mémoire lorsqu'on réduit ses dimensions a rendu sa commercialisation impossible, et a également mené à la prolifération de nombreuses nouvelles architectures 1T-DRAM dans les laboratoires de recherche. Parmi elles, l'A2RAM [Rodriguez 11], dont la démonstration expérimentale a été réalisée dans [Rodriguez 12]. Cependant, l'A2RAM est encore à optimiser pour la rendre compétitive par rapport à la 1T/1C DRAM standard.

Le but de ce travail de thèse est d'effectuer une étude approfondie de l'A2RAM par simulation à éléments finis (TCAD), caractérisation électrique et modélisation compacte. Le présent manuscrit est organisé en 6 chapitres.

Dans le premier chapitre, le fonctionnement d'une cellule mémoire 1T-DRAM sera décrit, essentiellement à partir de la physique du transistor MOSFET et de ses évolutions technologiques. Ensuite les évolutions technologiques qui ont menées à l'introduction de la cellule A2RAM seront retracées.

Dans le second chapitre, le comportement de la cellule A2RAM durant les différentes opérations mémoires sera détaillé via des simulations TCAD. Dans un premier temps, nous présenterons notre stratégie de simulation qui sera validée par des données expérimentales. Pour finir, le comportement de l'A2RAM sera décrit au travers d'une analyse du rôle et de l'impact de chaque polarisation pendant l'opération mémoire.

Dans le troisième chapitre, une étude de sensibilité des performances de l'A2RAM à ses paramètres technologiques (longueur, épaisseur, dopage et matériau) sera menée via des simulations TCAD. Des structures A2RAM optimisées seront alors proposées. Enfin, un nouveau mécanisme de programmation à basse tension sera mis en évidence.

Dans le quatrième chapitre, une méthode d'extraction du profil de dopage du bridge de l'A2RAM sera développée à partir de simulations TCAD, et mise en œuvre en caractérisation électrique. Cette technique d'extraction électrique nous a permis d'évaluer les performances mémoires des échantillons de cellules A2RAM disponibles.

Dans le cinquième chapitre, le développement d'un modèle compact d'une A2RAM sera décrit. Pour pouvoir évaluer les performances de la cellule A2RAM à l'échelle circuit, un modèle compact est indispensable. Le modèle sera implémenté en VerilogA, et puis validé par simulation TCAD. Enfin, une matrice 2x2 A2RAM sera simulée.

Dans le sixième chapitre, le procédé de fabrication d'A2RAM basé sur un procédé de fabrication de nanofil, défini par simulations TCAD sera présenté. Enfin, les caractérisations électriques des dispositifs fabriquées seront détaillées afin de mettre en évidence l'opération mémoire et de quantifier les performances.

CHAPITRE I : LES CELLULES MÉMOIRES DYNAMIQUES

I.1. Les technologies mémoires

L'industrie des technologies de l'information et communication est toujours en évolution, partant d'un ordinateur en poste fixe pour effectuer des simples opérations localement ; il a fallu attendre l'arrivée d'internet en 1960 pour pouvoir effectuer des opérations en mode non local. L'avènement de l'internet a non seulement permis de rendre possible des opérations sur des données à distance et en temps réel, mais aussi l'apparition de dispositifs digitaux. Grâce à la digitalisation de l'industrie des technologies de l'information et communication il est alors désormais possible d'avoir des dispositifs dits intelligents : les smartphones, smart TV, GPS, appareils photos numériques, montre connectée et bien d'autres. Aujourd'hui, nous sommes à l'aire de : l'intelligence artificielle, la réalité virtuelle/augmentée, et l'internet des objets.

- (i) L'intelligence artificielle : une définition claire de cette thématique ne s'aurait être exhaustive, mais elle s'inscrit dans le développement de technologies qui imitent le comportement humain. Ces technologies touchent : l'énergie, l'agriculture, la fabrication d'objets, la logistique, la santé, la construction, les transports et presque toutes les autres industries imaginables.
- (ii) La réalité virtuelle est une simulation ou une reproduction artificielle générée par ordinateur, d'un environnement ou d'une situation de la vie réelle. Il donne à l'utilisateur le sentiment de vivre directement la réalité simulée, principalement en stimulant sa vision et son audition. Cependant, la réalité augmentée est une technologie qui superpose les améliorations générées par ordinateur au-dessus d'une réalité existante afin de la rendre plus significative grâce à la possibilité d'interagir avec celle-ci. La réalité augmentée est développée en applications, et est utilisée sur les appareils mobiles pour intégrer des composants numériques dans le monde réel, de manière qu'ils se renforcent mutuellement, tout en pouvant être différenciés facilement.
- (iii) L'internet des objets est un système d'interfaces informatiques, de machines et d'objets numériques reliés entre eux, dotés d'identifiants uniques et permettant de transférer des données sur un réseau sans nécessité l'interaction de l'Homme/ordinateur.

Pour la mise en œuvre de ces trois nouveaux axes technologiques (i, ii, et iii), les dispositifs électroniques mis en jeu devront nécessairement, traiter de l'information ensuite la transférer. Il est alors important de pouvoir stocker temporairement ou à long terme l'information dans ces dispositifs, d'où une nécessité grandissante en besoin de mémoire dans l'industrie des technologies de l'information et de la communication. Aujourd'hui, l'idéal serait d'avoir mémoire unique pouvant être utilisée par n'importe quels dispositifs électroniques. Cette mémoire doit répondre au cahier de charges suivants : avoir une haute densité d'intégration, avoir des vitesses de programmation et de lecture les plus élevées possibles, pouvoir stocker l'information le plus longtemps possible, on parle de mémoire universelle [Bawedin 11]. De nos jours, une mémoire qui répond à tous ces critères n'existe pas, mais deux catégories de mémoires issues de la technologie CMOS (Complementary Metal Oxyde Semiconducteur) sont utilisées : les mémoires dites volatiles (V) et les mémoires non volatiles (NV) [ITRS 15]. Les mémoires volatiles (V) perdent l'information une fois l'alimentation électrique coupée, et les mémoires non volatiles (NV) ne perdent pas

l'information même hors alimentation électrique. Ces architectures mémoires dites RAM pour Random Access Memory, se distinguent aussi par leur temps d'accès, leur encombrement, et leur coût de fabrication imposant ainsi des compromis pour leur utilisation. En effet, pour répondre aux exigences requises par les nouveaux axes technologiques, les dispositifs électroniques doivent avoir une vitesse de calcul quasiment instantanée, et ceci est garanti par un processeur (*CPU: Central Processor Unit*) ultra performant. Cependant, l'élément qui limite la vitesse des exécutions du calcul d'un processeur est le temps d'accès à la mémoire. Pour réduire autant que possible cette latence, les mémoires sont organisées en hiérarchie comme montré sur la figure 1.1 [Ghetti 18]. Les mémoires à très forte densité d'intégration (NV) ont un temps d'accès élevé, c'est pourquoi elles sont éloignées du processeur. Plus on se rapproche du processeur, les temps d'accès diminuent mais au prix d'une faible densité d'intégration et un coût élevé de fabrication (V).

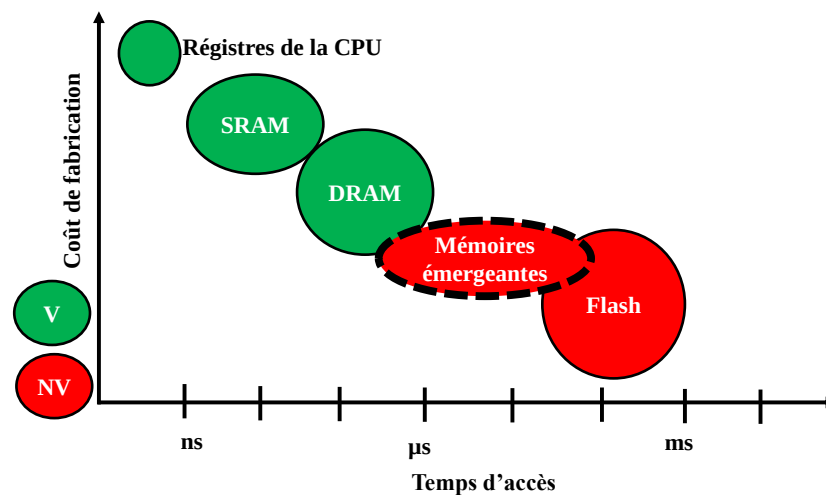


Figure 1.1 : Compromis coût, temps d'accès pour les différentes architectures mémoire. Les architectures les plus denses sont les plus rapides et sont les plus proches du processeur. L'écart entre la mémoire Flash et la mémoire DRAM offre une opportunité aux mémoires émergents. V : volatile, et NV : non volatile.

I.1.1 Mémoires Non Volatiles

Elles sont essentiellement dominées par la technologie 'Flash', qui est constituée d'un transistor servant à la fois de stockage (piégeage de la charge dans une grille flottante) et de moyen d'accès à l'information. Mais, puisque la mémoire flash commence à atteindre ses limites en termes de miniaturisation, aujourd'hui de nouvelles cellules mémoires non-volatiles et non-conventionnelles dites mémoires émergentes [Guy 15] sont proposées, pour pouvoir combler le gap existant entre la Flash et les mémoires volatiles. Pour ces mémoires, le principe de stockage de l'information n'est pas basé sur le 'piégeage' de la charge, mais sur le changement de propriétés du matériau. De ces mémoires émergentes deux technologies sont déjà en phase de commercialisation : la MRAM et la PCRAM. La MRAM (*Magnetic RAM*) a été fabriquée par Intel [Golonzka 18] et GlobalFoundries [Lee 18] pour des applications liées à l'intelligence artificielle et l'internet des objets qui demandent une faible consommation énergétique. La PCRAM (*Phase Change RAM*) est proposée par STmicroelectronics [Arnaud 18]

principalement pour le marché de l'automobile car cette architecture mémoire démontre une forte endurance après de multiple cycles de lecture et une intégrité de l'information stockée sur plusieurs années, et sous une température de 150°C [Ghetti 18].

I.1.2 Mémoires Volatiles

Ces types de mémoires, aussi appelées mémoires vives, sont les mémoires principales de tout système informatique. Elles permettent de stocker des données lors de l'exécution d'un programme, qui sont irrémédiablement effacées lorsqu'on éteint le système. Les mémoires volatiles peuvent être statiques (*Static RAM pour SRAM*) ou dynamique (*Dynamic RAM pour DRAM*).

I.1.2.1. Cellule SRAM

La cellule SRAM (figure 1.2) est schématiquement constituée de deux inverseurs tête-bêche (Inv1 et Inv2) et de deux transistors d'accès (T1 et T2). La cellule SRAM a pour avantages d'avoir un temps de maintien de l'information infinie (tant qu'elle reste alimentée par une source de tension), et un temps d'accès de l'ordre de la nanoseconde (le plus court des mémoires actuelles). Ainsi, l'utilisation des mémoires SRAM est dédiée aux applications où la vitesse est le critère principal tel que le niveau un de la cache de l'ordinateur [TAUR 09]. Toutefois, ses principales limites sont : la surface occupée par la cellule élémentaire qui est constituée de six transistors, et la consommation statique élevée due aux performances en vitesse demandées.

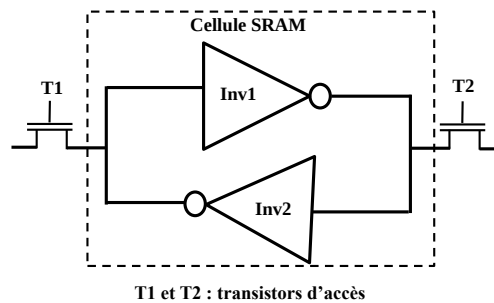


Figure 1.2 : Schéma d'une cellule SRAM.

I.1.2.2. Cellule DRAM

Le concept de mémoire DRAM a été proposé pour la première fois par [Dennard 68]. L'accès à l'information se fait via un transistor et l'information est stockée dans une capacité : on parle alors de 1T – 1C DRAM (figure 1.3). La cellule est à l'état '0' lorsqu'il n'y a pas de charge stockée dans le condensateur, et à l'état '1' lorsqu'il y a de la charge. La charge une fois stockée est perdue avec le temps via les courants de fuites du transistor d'accès, et des circuits environnants car le condensateur est laissé flottant. C'est pourquoi des cycles de rafraîchissement périodiques sont nécessaires pour rétablir l'état mémoire ('0' ou '1') dans la cellule. Contrairement à une SRAM, la DRAM a un temps de rétention fini (~100ms [TAUR 09]) même lorsqu'elle reste branchée à une source de tension. L'accès à l'information dans une DRAM est plus lent que dans une SRAM, mais le point mémoire est bien plus petit en DRAM qu'en SRAM.

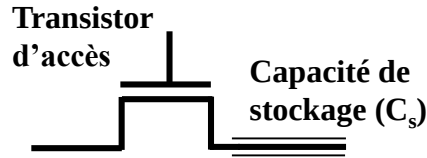


Figure 1.3 : Schéma d'une cellule DRAM.

I.1.3 Point mémoire

Le point mémoire est un dispositif isolé pouvant stocker de l'information communément appelé *bit* qui a pour valeur 0 ou 1. Les principales opérations pouvant être effectuées dans un point mémoires sont les suivantes :

- Programmation de l'état '1' : c'est le fait de pouvoir mettre le point mémoire dans un état '1' c'est-à-dire écrire dans le point mémoire le mot 1. Le mécanisme utilisé est fonction de la nature du point mémoire :
 - ⇒ La présence de la charge piégée dans le point mémoire. Ce mécanisme se retrouve en particulier dans les points mémoires DRAM et Flash.
 - ⇒ Un changement d'état du matériau : on peut avoir par exemple un état de métastabilité (un oxyde devient un métal) ou le passage à un état de faible résistivité. Comme mentionnée dans la section 1.1, ce mécanisme est utilisé dans les mémoires émergentes.
- Programmation de l'état '0' : c'est le fait de pouvoir mettre le point mémoire dans un état '0', c'est-à-dire écrire dans le point mémoire le mot 0. Les mécanismes utilisés sont les suivants :
 - ⇒ L'évacuation de la charge en excès dans le point mémoire si le mécanisme utilisé est le piégeage de la charge.
 - ⇒ Le retour à un état stable du matériau ou le retour à un état de forte résistivité, dans le cas d'écriture par changement d'état du matériau.
- Lecture de l'information : le mécanisme de lecture est toujours fonction de la nature du point mémoire en considération, et consiste soit à vérifier la quantité de charge stockée dans le point mémoire ou bien à vérifier l'état du matériau constituant le point mémoire. Dans les deux cas, cela va se traduire par une forte valeur de courant en sortie du point mémoire lorsqu'on va lire l'état '1', ou une faible valeur de courant lorsqu'on lira l'état '0'. Le courant lu de l'état '1' nous l'appellerons I_1 , et le courant lu de l'état '0' sera I_0 ; on considère qu'il y'a un effet mémoire lorsque $I_1 \neq I_0$. La 'marge' qui existe entre les valeurs de I_1 et I_0 est appelée la fenêtre de programmation (FP). Cependant, selon les auteurs, elle peut être définie comme étant la différence entre les courants lus ($I_1 - I_0$) [Okhonin 08], ou comme étant le rapport des courants lus (I_1/I_0) [Bawedin 08].
- Le maintien de l'information : c'est l'intervalle de temps durant lequel aucune opération d'écriture, effacement et lecture n'est effectuée sur le point mémoire. C'est une opération fondamentale dans un point mémoire. Car, il est important de pouvoir maintenir le plus longtemps possible l'information stockée ('1' ou '0') dans le point mémoire. Durant cette phase de maintien, l'information stockée doit garder toute son intégrité c'est-à-dire qu'on ne doit pas la perdre. Pour les mémoires non volatiles, ce

souci du maintien de l'information ne se pose pas, de même que la mémoire SRAM tant qu'elle est sous une alimentation électrique. Mais, avec une mémoire dynamique, on considère qu'après un certain temps l'information sera automatiquement perdue. Ce temps est appelé le temps de rétention, il est égal à la durée d'une écriture ou un effacement parasite induits par les courants de fuites des dispositifs qui environnent la capacité de stockage.

I.1.4 Description d'un point mémoire sous matrice

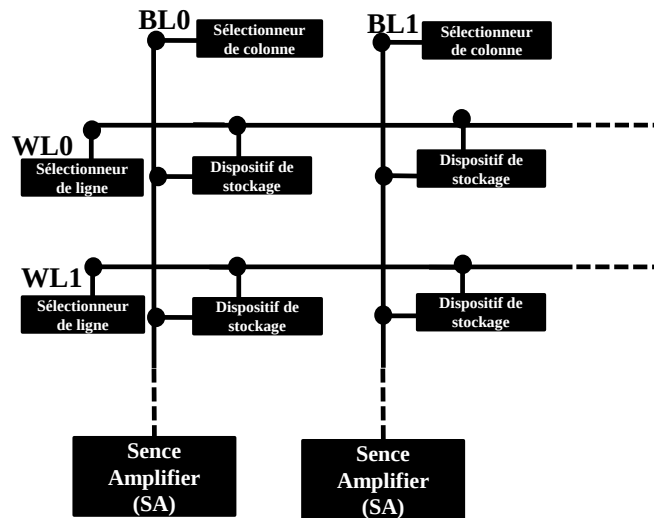


Figure 1.4 : Exemple simplifié de la disposition de cellules mémoires en matrice.

Un point mémoire mis sous matrice est une composition de deux éléments appelée dispositif de stockage. Il est constitué d'un 'nœud de stockage' qui est une des architectures mémoires présentées sur la figure 1.1, et un 'sélectionneur' qui donne accès au nœud de stockage. Le choix du 'sélectionneur' est étroitement lié à l'architecture mémoire choisie pour le nœud de stockage. Le dispositif le plus utilisé comme 'sélectionneur' est le transistor pour les mémoires volatiles et émergentes. La mémoire flash est un cas particulier où le transistor qui sert de nœud de stockage est aussi utilisé pour avoir accès à l'information. Durant les opérations mémoires (section 1.3), pour avoir accès à un mot du dispositif de stockage, on devra sélectionner la ligne de mot ou Word Line (WL) via le sélectionneur de ligne. Pour la sélection de la ligne de donnée ou Bit Line (BL) correspondante, elle se fera via le sélectionneur de colonne. La programmation de l'information se fera ligne par ligne, et la lecture de l'information colonne par colonne. De plus, pour faciliter la lecture de l'information un circuit de lecture appelé 'current sense amplifier' ou SA est utilisé. Comme son nom l'indique, il va amplifier la valeur du courant lu (I_1 ou I_0) afin de réduire autant que possible le temps de lecture.

Pour garantir une bonne fonctionnalité du point mémoire sous matrice, les concepteurs de circuits imposent des critères de fonctionnalités sur le point mémoire seul :

- I_1 doit être suffisamment élevé pour pouvoir être détecté par le circuit de lecture (SA). En effet, I_1 doit être supérieur à la valeur du bruit intrinsèque en courant du SA.

Aujourd'hui pour les mémoires dynamiques, on considère que le SA est en même de pouvoir détecter un courant non inférieur à 6 μA [Malinge 05].

- La fenêtre de programmation $I1/I0 > 20$ [Bawedin 08] ou $I1 - I0 > 6 \mu\text{A}$ [Okhonin 05] pour pouvoir mieux distinguer l'état '1' de l'état '0'.
- Le temps de rétention : c'est un critère critique pour une mémoire dynamique. Il est évalué en considérant la configuration du point mémoire qui favorise la perte de l'information due à des effets parasites. C'est pourquoi les conditions d'évaluation du temps de rétention sont fonction du type d'architecture de la cellule DRAM, car les effets parasites se manifestent différemment.

I.2. Mémoires Dynamiques 1T – 1C

Le nombre de composants connectés sera de 35 milliards d'ici 2020 [ITRS 15], ce qui aura pour conséquence d'augmenter drastiquement le trafic de données échangées, et une demande croissante de dispositifs de stockage notamment la DRAM. La figure 1.5 montre les prédictions de la demande des mémoires DRAM pour le marché de DATA CENTER d'ici 2020. Afin de répondre à cette demande il faut augmenter la densité d'intégration de la cellule mémoire DRAM tout en maintenant un bas coût de fabrication. Ceci est possible en réduisant la surface occupée par la cellule DRAM, comme le montre la figure 1.6 où la longueur caractéristique du point mémoire devrait avoisiner 10 nm d'ici 2022 [ITRS 07].

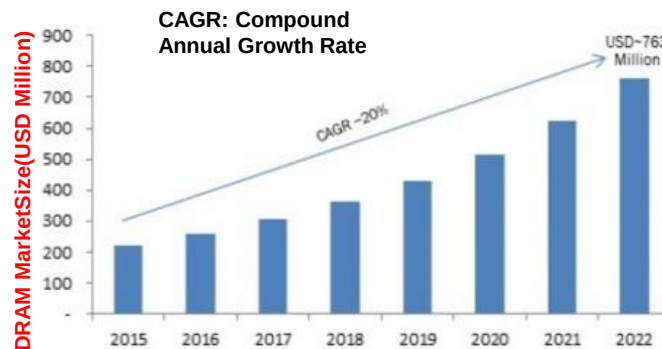


Figure 1.5 : Evolution du marché de DRAM pour les centres de données (data center en anglais) [MarketersMEDIA 18].

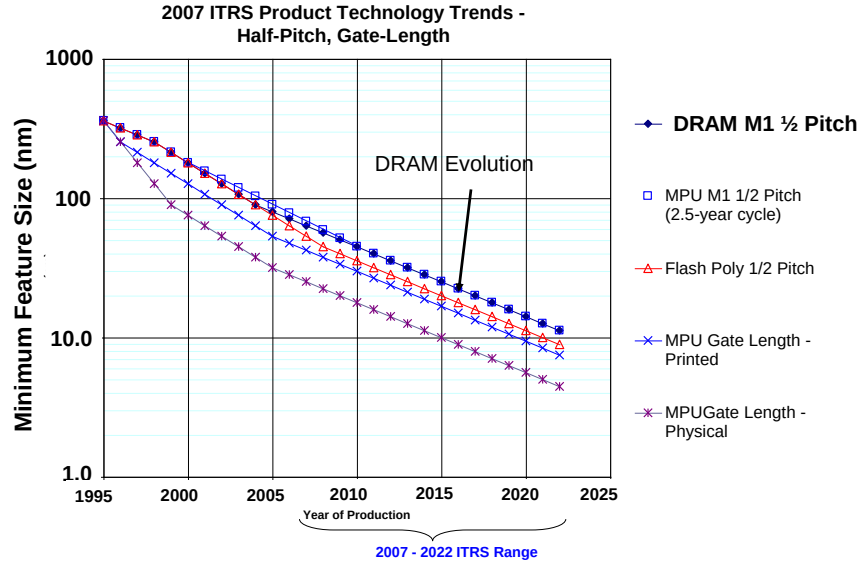


Figure 1.6 : Evolution de la longueur caractéristique (F) des nœuds technologiques DRAM en fonction des années [ITRS 07].

I.2.1. Réduction des dimensions du transistor d'accès

La réduction des dimensions du transistor d'accès suit les tendances des circuits logiques. L'élément principal qui gouverne la réduction de ses dimensions est le maintien du contrôle électrostatique par la grille. Nous aborderons cette thématique plus en profondeur dans la section 3. Notons tout de même que les premières architectures de point mémoire DRAM avaient des structures planaires pour le transistor d'accès jusqu'aux nœud technologique 110 nm (figure 1.7). Cependant, pour réduire la surface occupée par ce dernier, il faut réduire ses dimensions, ce qui a pour conséquence d'augmenter les courants de fuites des diodes source-substrat et drain-substrat, et donc de dégrader le temps de rétention de la cellule DRAM. Pour améliorer le temps de rétention, les architectures planaires avec une ingénierie des jonctions source (drain)/substrat ou de gravure de la grille telles que les transistors STAR (Step gaTed AsymmetRic) [Jang 05] et S-RCAT (sphere-shaped-recess-channel-array transistor) [Park 05] ont été proposées. Mais la lithographie associée à la fabrication de ces architectures mémoires augmente le coût de fabrication, c'est pourquoi les industriels se sont tournés vers les architectures de transistors 3D telles que le FinFet, et la grille verticale [Kawasaki 09].

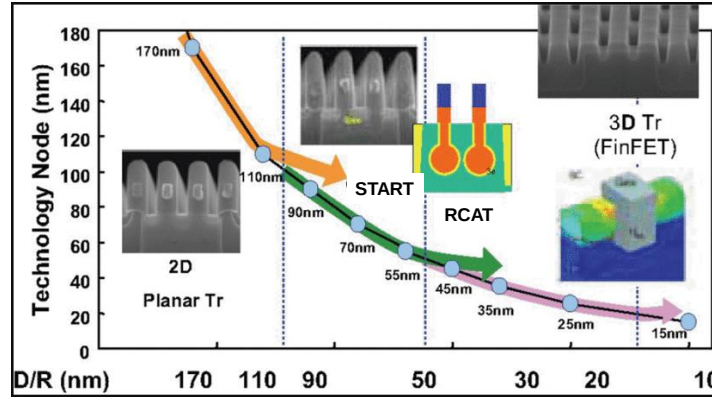


Figure 1.7 : Evolution des nœuds technologiques en fonction de la longueur caractéristique de la technologie (F) du transistor [Park 06].

Bien que la réduction du transistor d'accès impose des challenges technologiques, il est encore possible de réduire ses dimensions pour les prochains nœuds technologiques ; mais qu'en est-il de la capacité de stockage ?

I.2.2. Réduction des dimensions de la capacité de stockage

Dans un premier temps, pour augmenter la densité d'intégration de la DRAM, il suffisait de réduire la surface occupée par la capacité de stockage tout en conservant une valeur constante (environ 20/30 fF) [TAUR 09]. Cette valeur de capacité était garantie d'abord par une réduction de l'épaisseur de l'oxyde puis par l'introduction de matériaux à forte permittivité (High-K). Mais, cette stratégie s'est rapidement avérée inefficace lorsque les épaisseurs d'oxyde ont atteint des valeurs ultimes [Wendell 85]. Pour continuer à réduire la surface occupée par le point mémoire tout en maintenant la valeur de la capacité de stockage constante, les industriels se sont alors tournés vers des architectures types 3D (figure 1.8) :

- En enterrant la capacité de stockage sous le transistor d'accès (figures 1.8.a et 1.8.b) : on parle de *trench capacitance* [Intel, EETimes 14].
- En empilant la capacité de stockage au-dessus du transistor d'accès (figures 1.8.c et 1.8.d) : on parle de *stacked capacitance* [IBM SRC eDRAM 09].

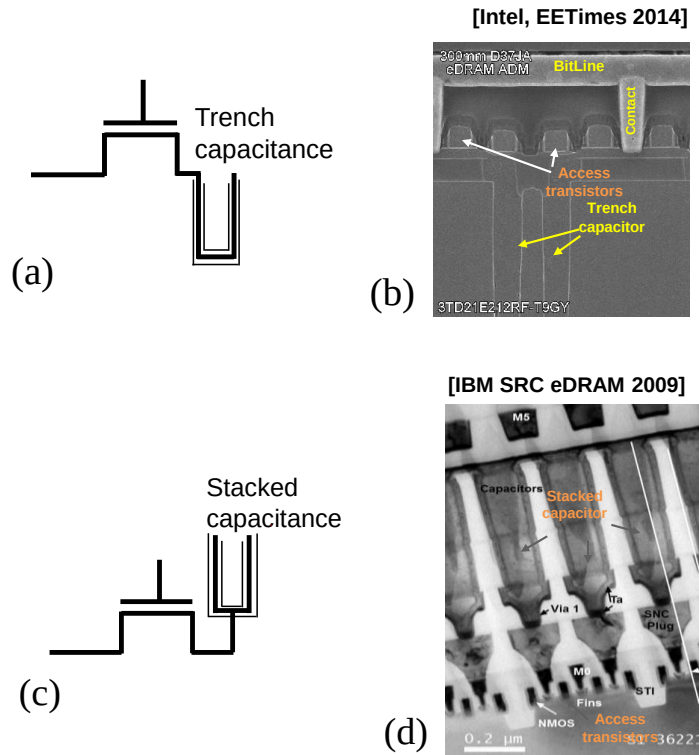


Figure 1.8 : Schéma et vue au MET (Microscope Electronique en Transmission) de la mémoire 1T-1C DRAM à : (a,b) capacité enterrée (trench DRAM), (c,d) capacité empilée (stacked DRAM).

On comprend aisément que de telles architectures ont nécessité d'importants développements technologiques : lithographie, introduction de nouveaux matériaux notamment les oxydes à grande permittivité ou encore de nouvelles morphologies en passant par l'ingénierie des électrodes comme le montre la figure 1.9 pour le cas d'une capacitance empilée. Bien qu'aujourd'hui commercialisées, les architectures DRAM de types 3D présentent une forte complexité en termes de procédé de fabrication, donc une augmentation importante du coût de fabrication [Park 06].

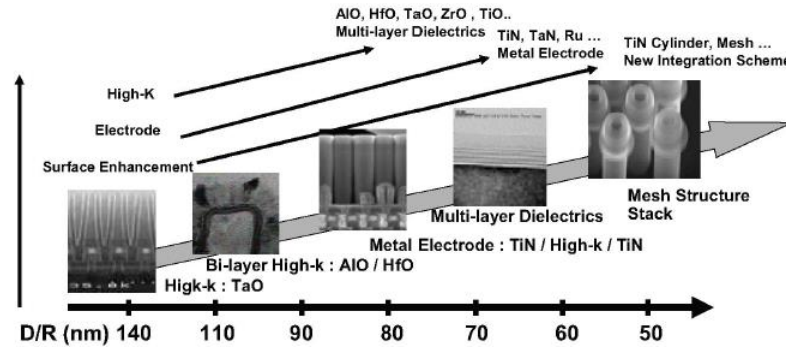


Figure 1.9 : Evolution de technologie de la *Stacked capacitance* en fonction des nœuds technologiques [Park 06].

Dans la course à la miniaturisation du point mémoire 1T-1C DRAM, le point critique est donc la capacité de stockage. C'est pourquoi les industriels semblent aujourd'hui ouverts pour envisager un changement de concept de point mémoire DRAM, et se tourner vers les DRAM sans capacité de stockage : les 1T-DRAM.

I.3. La cellule mémoire 1T – DRAM

Dans cette partie, nous allons tout d'abord parler du transistor MOSFET qui est essentiel pour la compréhension du fonctionnement d'un point mémoire 1T-DRAM. Nous introduirons des notions de base liées au fonctionnement du transistor MOSFET : caractéristiques statiques, effets parasites et challenges technologiques pour sa miniaturisation. Pour finir, nous introduirons les notions d'opérations mémoires dans une 1T-DRAM.

I.3.1. Transistor MOSFET

La brique de base et fondamentale de la mémoire 1T-DRAM est le transistor MOSFET. Il sert à la fois de nœud de stockage et point d'accès à l'information. Ainsi, pour comprendre le fonctionnement d'une 1T-DRAM, il est important de comprendre le fonctionnement d'un transistor MOSFET. Le but de cette partie est de détailler les phénomènes physiques du transistor MOSFET qui sont ensuite réutilisés dans les architectures 1T-DRAM, mais aussi de retracer les éléments qui ont conduit aux évolutions technologiques de ce dernier.

I.3.1.1. Architecture BULK

Le transistor MOSFET est issu de la technologie MOS (Metal Oxide Semiconductor). Ses différents régimes électrostatiques sont ceux d'une capacité MOS, d'où l'importance de bien comprendre son fonctionnement.

a- Capacité MOS

La structure MOS (Metal Oxide Semiconductor) se compose d'un empilement de métal, d'oxyde qui peut d'être du dioxyde de silicium (SiO_2), et d'une couche de semiconducteur tel que du silicium dopé de type p (figure 1.10). Durant son fonctionnement, selon la différence des travaux de sortie entre le métal et le semiconducteur, et de la tension appliquée sur le métal, le potentiel de surface ϕ_s à l'interface oxyde-semiconducteur va être modulé. On peut

donc distinguer trois régimes (figure 1.10), délimités par des valeurs particulières du potentiel de surface :

- Le régime d'accumulation (figure 1.10.a) : défini par la présence de porteurs majoritaires dans le semiconducteur juste quelques nanomètres en dessous de l'oxyde. L'évolution de la charge induite par les variations des porteurs minoritaires dans le semi-conducteur est une fonction exponentielle du potentiel de surface [SZE 07].
- Le régime de désertion (figure 1.10.b) : définie par une épaisseur de désertion T_{dep} en dessous de l'oxyde qui est totalement dépourvue de porteurs majoritaires du semiconducteur. On peut démontrer que l'épaisseur de désertion T_{dep} , et la charge dans la zone de désertion sont proportionnelles à la racine carrée du potentiel de surface ϕ_s [SZE 07].
- Le régime d'inversion (figure 1.10.b) : défini par la présence d'une couche de porteurs minoritaires juste en dessous de l'oxyde. Cependant, en fonction de la dynamique de variation des porteurs minoritaires, on distingue :
 - La faible inversion définie par $\phi_s < 2\phi_f$, où la charge induite par les porteurs minoritaires est fonction du racine carrée du potentiel de surface ϕ_s [SZE 07].
 - Et le régime de forte inversion définie par $\phi_s > 2\phi_f$; dans ce cas la charge induite par les porteurs minoritaires est une fonction exponentiel du potentiel de surface ϕ_s [SZE 07].

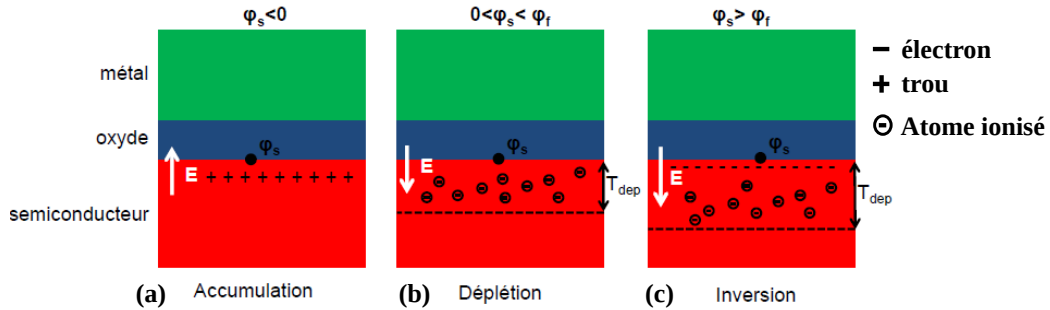


Figure 1.10 : Capacité MOS en régime (a) d'accumulation, (b) déplétion, et (c) d'inversion [Lacord 12].

Notons le cas particulier où $V_g = V_{FB}$ qui marque la condition de bandes plates, c'est-à-dire la limite entre le régime d'accumulation et le régime de désertion. Dans cette condition, tout le semiconducteur reste neutre, et le diagramme à bande d'énergie de l'empilement Métal-Oxyde-Semiconducteur est aligné comme montré sur la figure 1.11. Cet état est atteint soit lorsque le travail de sortie du métal se trouve au milieu de la bande interdite du matériau semiconducteur : on parle de *mid-gap metal*, ou lorsque la tension appliquée sur le métal est l'opposé de la tension de bande V_{FB} [SZE 07].

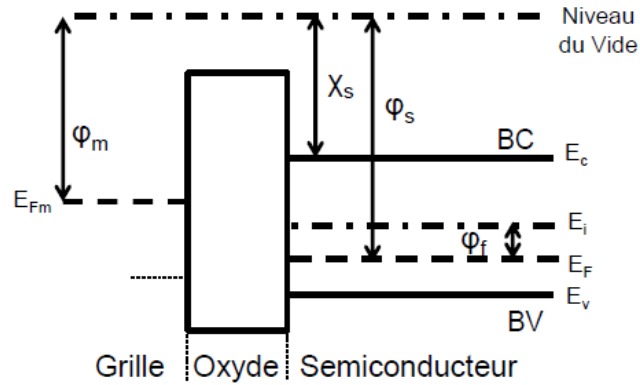


Figure 1.11 : Structure de bandes de la capacité MOS en régime de bandes plates [Lacord 12].

b- Tension seuil du transistor MOSFET

La structure de base d'un transistor MOSFET est illustrée sur la figure 1.12. Il s'agit d'un dispositif à quatre terminaux : grille, source, drain et substrat. Dans le cas d'un transistor MOSFET à canal n ou nMOSFET comme sur la figure 1.12, le substrat de silicium est de type p, et les implantations des zones source et drain sont de type n⁺⁺. L'électrode de grille est généralement constituée de polysilicium fortement dopé, et elle est séparée du substrat par un mince film de dioxyde de silicium : l'oxyde de grille. De plus, pour éviter des courts circuits entre les contacts de source et drain avec celui de la grille, il y'a une couche d'isolant appelée espaceur. La région sous l'oxyde de grille entre la source et le drain s'appelle le canal, et elle est essentielle à la conduction du courant dans un transistor MOSFET. Le fonctionnement de base d'un transistor MOSFET peut être facilement compris à partir du condensateur MOS décrit dans la section précédente.

En effet, en appliquant une faible tension positive sur le drain (source et substrat à la masse), le champ électrique transverse va produire un courant entre la source et le drain, qui dépendra du régime du canal, comme le montre la caractéristique statique courant de drain I_d en fonction de la tension de la grille V_g sur la figure 1.13.a.

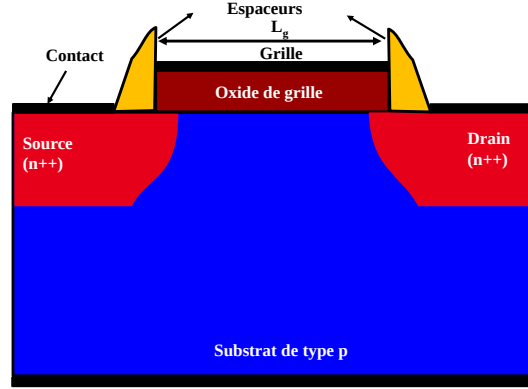


Figure 1.12 : Architecture d'un transistor MOSFET sur substrat massif ou BULK.

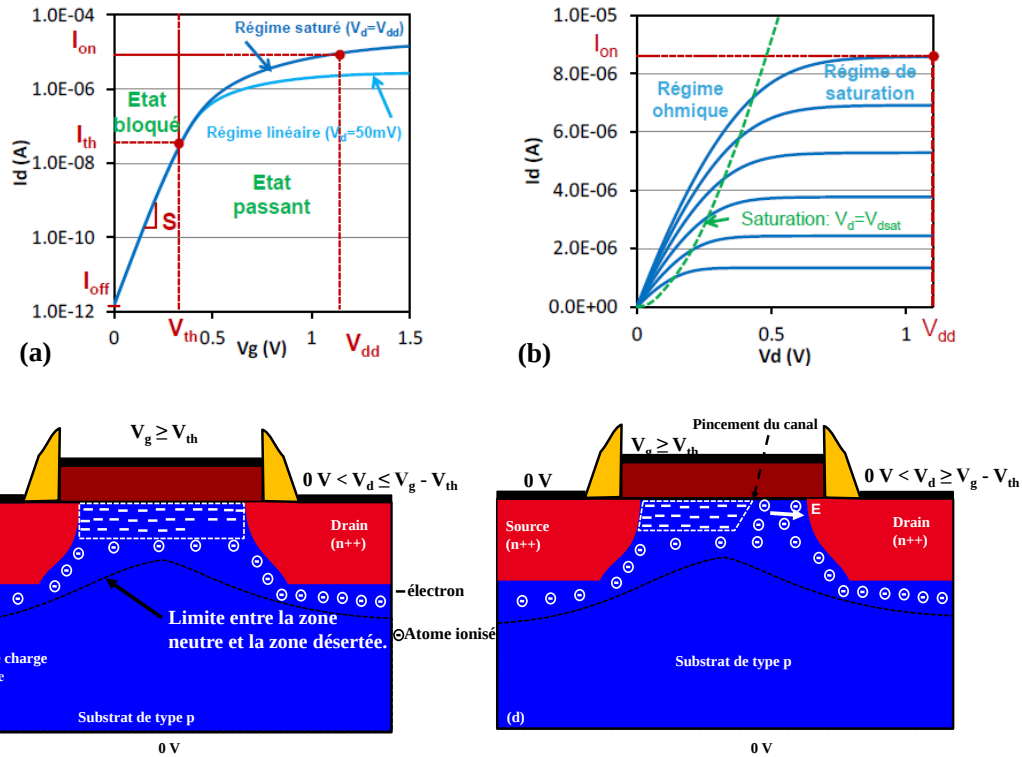


Figure 1.13 : Caractéristiques statiques (a) I_d - V_g , et (b) I_d - V_d , d'un transistor MOSFET sur substrat massif extraites des mesures électriques pour un dispositif avec une longueur de grille $L_g = 10 \mu m$ et une largeur $W = 10 \mu m$ [Lacord 12]. Schéma d'illustration du transistor MOSFET en (c) régime ohmique, (d) régime de saturation.

A faible polarisation de grille $V_g < V_{th}$ (figure 1.13.a) : nous sommes en régime de déplétion, on dit que le transistor est à l'état bloqué. Le courant I_d entre le drain et la source est un courant de diffusion, d'où la pente sous le seuil S sur la figure 1.13.a. On définit le paramètre SS (*Subthreshold Swing*) comme étant l'inverse de la pente sous seuil (S) de la caractéristique I_d - V_g en échelle logarithmique, et c'est une figure de mérite importante du transistor MOSFET. Il quantifie la

variation de la tension sur la grille qui induit une augmentation d'une décade du courant de drain sous le seuil. La valeur théorique minimale de SS est de 60 mV/dec à température ambiante (300 K) [Colinge 91].

Lorsque V_g augmente, quand $\phi_s = 2\phi_f$, une couche d'inversion est créée dans le canal lorsque $V_g = V_{th}$, ainsi un fort courant de dérive circule entre la source et le drain : on dit que le transistor est à l'état passant. Dans cette configuration, en fonction de la tension V_d appliquée sur le drain on distinguera :

- La région ohmique définie pour $V_d < V_g - V_{th}$ ($V_d = 50$ mV sur la figure 1.13.c) : elle est marquée par la présence d'une couche d'inversion tout le long du canal.
- La région de saturation définie pour $V_d > V_g - V_{th}$ ($V_d = V_{dd}$ sur la figure 1.13.d) : elle est marquée par la présence d'une zone de désertion entre le point de pincement du canal (pinch-off) et le drain (figure 1.13.d). Les électrons qui traversent cette région sont accélérés par le champ électrique transversal E .

Comme on peut le voir sur la figure 1.13.a, la valeur du courant I_d sous le seuil tends vers I_{off} . I_{off} est un courant principalement dû aux fuites des diodes entre les jonctions source/substrat et drain/substrat.

L'autre caractéristique statique du transistor MOSFET est le courant de drain I_d en fonction de la tension de drain V_d sur la figure 1.13.b, pour plusieurs tensions de grille. Nous pouvons distinguer les deux régions de fonctionnement du transistor citées précédemment délimitées par les tensions de drain : régime ohmique et régime de saturation (ligne verte en pointillé avec $V_d = V_{dsat} = V_g - V_{th}$).

c- Effets parasites dans un transistor MOSFET

1) Courant de fuite des diodes

Lorsque le transistor est polarisé en mode bloqué ($V_g < V_{th}$ et $V_d > 0$ V), les diodes parasites se trouvant entre la source-substrat et le drain-substrat (figure 1.14) sont polarisées en inverse. Nous avons un courant de saturation de porteurs minoritaires qui circule dans le transistor ; en fonction de la qualité des jonctions p-n constituant ces diodes, ce courant de saturation peut être non négligeable. La conséquence est d'avoir une augmentation de la pente sous le seuil de la caractéristique statique I_d - V_g et par conséquent du courant I_{off} .

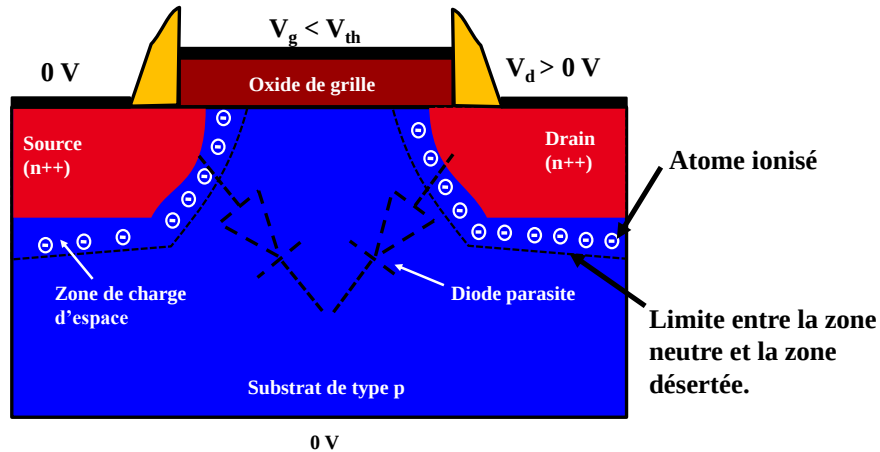


Figure 1.14 : Transistor MOSFET polarisé en mode bloqué.

2) Effet tunnel bande à bande dû au GIDL

Le GIDL (*Gate Induced Drain Leakage*) [Chang 87], [Chan 87] [Kurimoto 89], [Endoh 90], [Bouhda 97] s'active dans le cas d'un transistor MOSFET de type n, qui provoque une augmentation du courant de drain I_d pour des polarisations de grille de plus en plus négatives et une tension de drain $V_d > 0$ V (figure 1.15.a [Rafhay 12]). Ce phénomène a lieu dans la zone de recouvrement drain-oxyde de grille comme on peut le voir sur la figure 1.15.b. En effet, lorsqu'on applique une forte polarisation négative sur la grille et une forte polarisation sur le drain, la zone de recouvrement drain/oxyde de grille est mise en régime 'd'accumulation'. Pour un point P situé dans cette zone de recouvrement (figure 1.15.b), le fort champ électrique qui y règne induit une courbure des à bandes d'énergie comme illustrée sur la figure 1.15.c. Lorsque cette courbure des bandes d'énergie est supérieure à la valeur de la bande interdite d'énergie E_g du matériau semi-conducteur coté drain, nous aurons une génération des paires électrons-trous par effet tunnel bande à bande. Les électrons de la bande de valence traversent la bande d'énergie interdite (E_g) par effet tunnel pour aller vers la bande de conduction libérant des trous derrière eux. Les paires électrons-trous générées par effet tunnel bande à bande sont séparées par le champ électrique latéral entre la jonction drain-substrat. Les électrons sont collectés par le drain et contribuent au courant de fuite du drain de la partie rouge de la figure 1.15.a ; les trous sont évacués par le substrat.

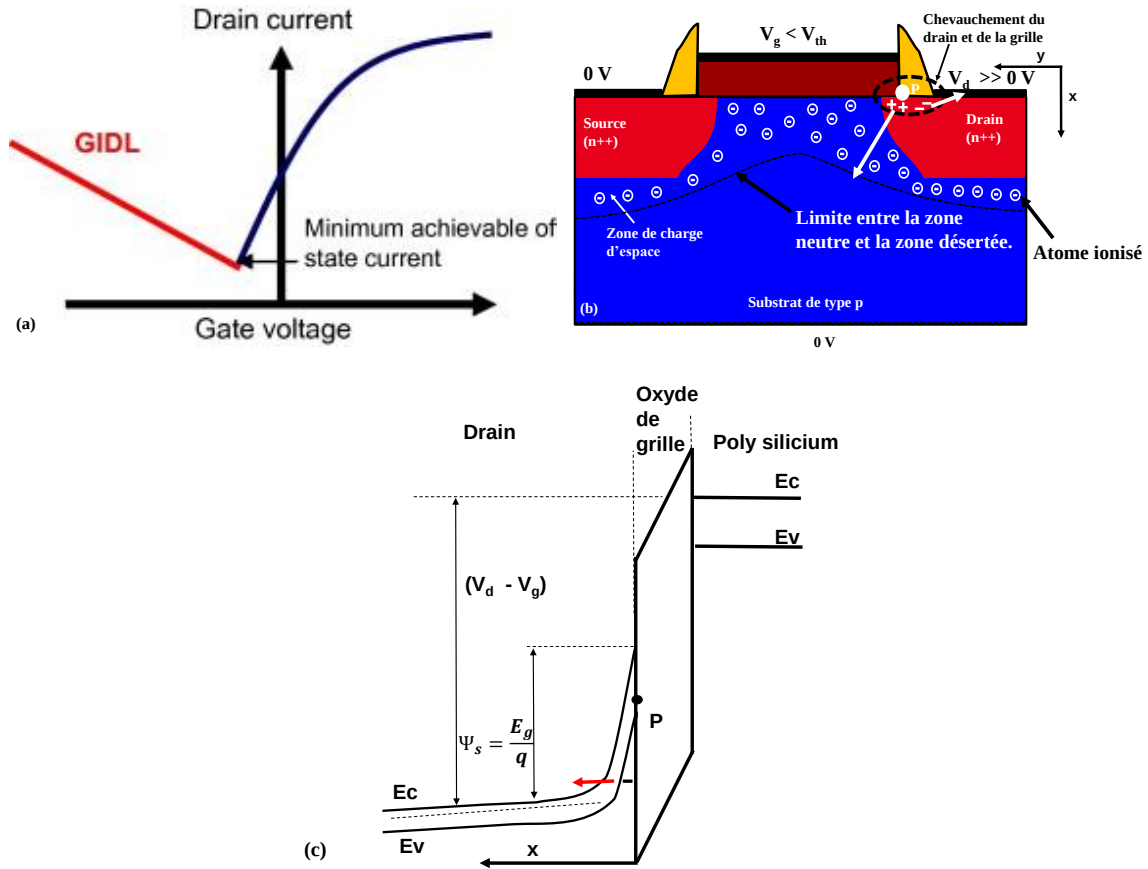


Figure 1.15 : (a) Caractéristique courant de drain en fonction de la tension de grille d'un dispositif MOSFET sujet à un courant de fuite par GIDL [Rafhay 12]. (b) Point mémoire PDSOI : programmation de l'état '1' par B2B. (c) Profil du diagramme de bandes d'énergie.

3) Ionisation par impact en régime de saturation

Contrairement aux deux autres effets parasites, celui-ci s'active lorsque le transistor est en conduction ($V_g > V_{th}$) et en régime de saturation (V_d fort). Dans cette condition, il se crée dans le canal une zone de désertion comme nous l'avons vu précédemment. Les électrons arrivant de la source, une fois au niveau du point de pincement seront accélérés vers le drain. Si le champ électrique est suffisamment élevé (généralement de l'ordre de 10^5 V.cm^{-1} [Mathieu 04]), l'énergie de l'électron augmente. S'il entre en collision avec un atome, il pourra céder une partie de son énergie, ce qui donnera lieu à la génération d'une paire électron trou.

Les trous générés par ionisation par impact seront évacués via le substrat. Dans le cas (1) de la figure 1.16, les électrons générés dans la zone de désertion seront accélérés et recueillis au drain augmentant le courant de saturation. Mais, il arrive que certains électrons une fois dans la zone de charge d'espace acquièrent de l'énergie et la conservent car, ils n'entrent pas en collision avec le réseau cristallin : on parle de génération de porteurs chauds (*hot carriers generation*, en anglais) [SZE 07] : c'est le cas (2) sur la figure 1.16. Si l'énergie acquise par ces porteurs chauds est supérieure à la barrière d'énergie entre l'interface face avant et oxyde de grille, ils pourront s'échapper du canal et être piégés dans l'oxyde de grille. Ce phénomène a

pour conséquence de détériorer l'oxyde grille et entraîne avec le temps une perte au niveau de la fiabilité de ce dernier.

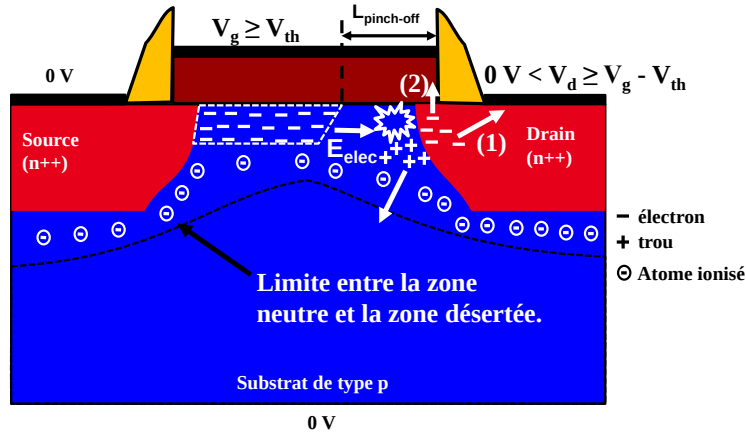


Figure 1.16 : Point mémoire PDSOI : programmation de l'état '1' par ionisation par impact. (1) Courant d'électrons collecté au drain, (2) porteurs chauds piégés dans l'oxyde de grille.

d- Réduction des dimensions du transistor MOSFET sur substrat massif

Les règles de miniaturisation du transistor MOSFET sont régies par la loi de MOORE [SZE 07] qui prédisait que le nombre de transistors par plaque devait doubler tous les deux ans. En d'autres termes la surface occupée par le transistor est réduite de moitié d'un nœud technologique à l'autre. Cependant, pour des nœuds technologiques plus avancés (en-dessous de 100 nm), la réduction de la longueur de la grille L_g va entraîner une perte du contrôle électrostatique de la grille sur le canal. Ceci aura pour conséquence une réduction de la tension seuil V_{th} et une augmentation des effets parasites dans le transistor.

En effet, pour un transistor à canal long comme sur la figure 1.17.a, bien qu'il existe une zone de charge d'espace entre les jonctions source-substrat et drain-substrat, le régime de désertion dans le canal est intégralement contrôlé par la grille. Ceci se traduit par le profil de barrière de potentiel de la figure 1.17.d : les électrons qui se trouvent dans la source et le drain doivent traverser cette barrière de potentiel pour former le canal d'inversion. Mais avec la réduction de la longueur de la grille, ces zones de charge d'espace se rapprochent de plus en plus, et le régime de désertion dans le canal n'est plus totalement contrôlé par la grille (figure 1.17.b). Ainsi, la barrière de potentiel le long de la source/substrat/drain s'abaisse (figure 1.17.e), les électrons ont besoin de moins d'énergie pour former le canal. Dans cette condition, la tension seuil est alors réduite à cause d'un effet canal court ou SCE (*Short Channel Effect*). Si en plus d'un canal court, une tension positive est appliquée sur le drain, la jonction drain-substrat est polarisée en inverse ce qui augmente encore plus la désertion dans le substrat, d'où le second abaissement de la barrière de potentiel sur la figure 1.17.f. Nous aurons une fois de plus une réduction de la tension seuil à cause de la polarisation du drain ou DIBL (*Drain Induced Barrier Lowering*). La réduction de la tension seuil se traduit par l'équation 1.1, et les expressions du SCE et du DIBL sont données par les équations 1.2 et 1.3 respectivement.

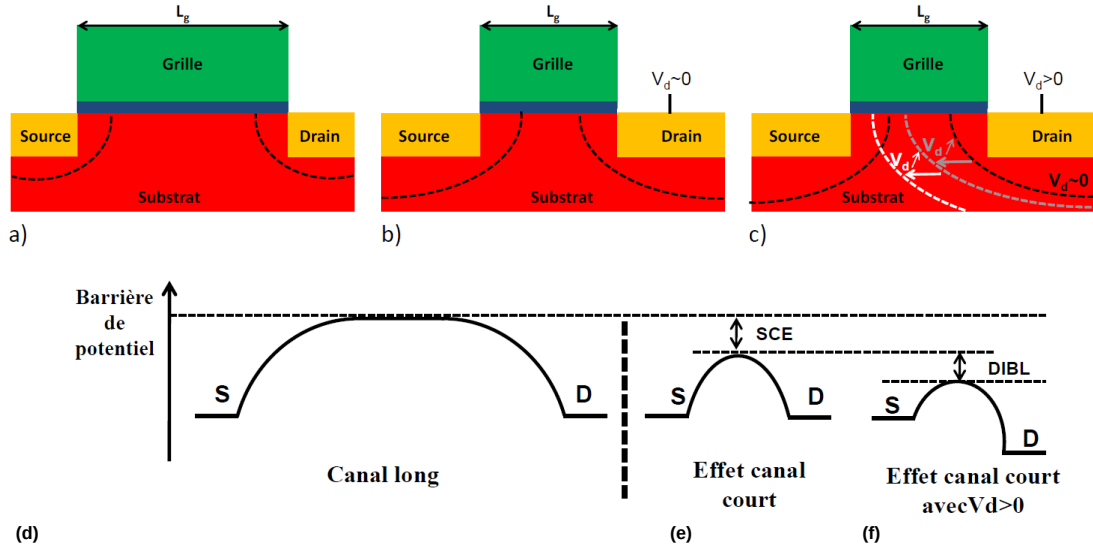


Figure 1.17 : Illustration de l'extension de la zone de charge d'espace coté source et drain pour (a) un transistor à canal long (b) un transistor à canal court avec une polarisation de drain V_d négligeable (c) un transistor à canal court avec une polarisation de drain V_d forte. Les pointillés indiquent l'extension de la zone de charge d'espace. (d, e, f) Illustration de l'abaissement de barrière de potentiel entre source et drain, donc de la tension de seuil due à la réduction de longueur de grille (SCE) puis à la polarisation du drain (DIBL) [Lacord 12].

$$V_{th_canalcourt} = V_{th_canallong} - SCE - DIBL \quad (1.1)$$

$$SCE = \frac{\epsilon_{Sc}}{\epsilon_{ox}} \frac{t_{ox}}{L_g} \frac{T_{dep}}{L_g} \phi_d \quad (1.2)$$

$$DIBL = \frac{\epsilon_{Sc}}{\epsilon_{ox}} \frac{t_{ox}}{L_g} \frac{T_{dep}}{L_g} V_d \quad (1.3)$$

Il est donc important de réduire l'impact du SCE et DIBL sur la tension seuil. En se basant sur leurs expressions, dans un premier temps la solution consistait à réduire l'épaisseur de l'oxyde de grille et à utiliser l'utilisation de matériaux high-K pour l'oxyde de grille pour augmenter sa permittivité. Pour réduire la valeur de T_{dep} il fallait augmenter le dopage du substrat, ce qui a pour effet d'induire la fluctuation des dopants (*RDF: Random dopant fluctuation*) dans le substrat [Weber 08]. Toutes ces améliorations ont contribué à complexifier le procédé de fabrication du transistor MOSFET sur massif augmentant le coût de fabrication. C'est pourquoi les industriels ont changé de stratégie et se sont tournés vers un changement d'architecture de transistor. Parmi ces choix, nous avons les technologies de transistor MOSFET sur substrat SOI.

I.3.1.2. Transistor MOSFET PDSOI

L'architecture du transistor MOSFET PDSOI (*Partially Depleted Silicon On Insulator*) a pour particularité une couche d'oxyde en dessous des zones source et drain : on parle d'oxyde enterré ou BOX (Buried Oxide). La partie en-dessous du BOX est appelée le substrat et la

partie au-dessus du BOX est appelée film de silicium ou body. Une partie du film de silicium contrôlée par la grille est désertée, et l'autre est neutre. Si le film de silicium n'est pas connecté à une source de tension, la partie neutre est dite flottante (*floating body*). Cette architecture a pour avantage de ne pas avoir de courant de fuite des diodes parasites car les zones source et drain sont isolées du substrat par la couche d'oxyde enterré.

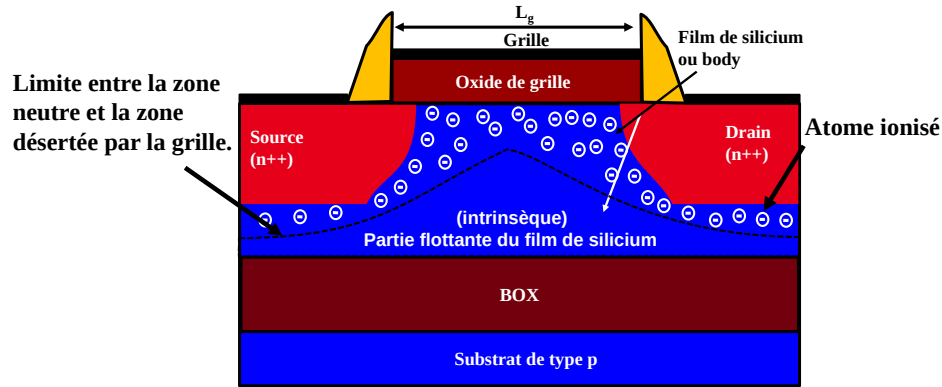


Figure 1.18 : Architecture d'un transistor MOSFET PDSOI, la partie supérieure du film de silicium est contrôlée par la grille est totalement désertée, et la partie en-dessous est neutre.

Néanmoins, lorsque le transistor PDSOI fonctionne en régime de saturation, comme pour le transistor MOSFET sur substrat massif, nous avons une génération de charge dans le canal. Dans un transistor classique sur silicium massif, les porteurs minoritaires ainsi générés sont collectés par le drain et les majoritaires sont évacués par le substrat. Dans un transistor PDSOI, du fait de la présence de l'oxyde enterré, les porteurs majoritaires ne peuvent plus être évacués par le substrat et se trouvent alors stockés dans le substrat flottant. Dans le cas d'un nMOS, les trous stockés dans le substrat flottant ont pour effet d'augmenter la charge dans cette région et par conséquent de réduire la tension de seuil du transistor, et donc d'augmenter son niveau de courant débité. On parle alors d'effet Kink [Colinge 85], [Colinge 88], il se traduit par une réduction de la tension seuil et une augmentation de la pente sous le seuil [Fossum 87] sur la caractéristique de transfert du transistor courant de drain I_d en fonction de la tension grille-source V_g (I_d - V_g) de la figure 1.19.a. Sur la caractéristique courant de drain en fonction de la tension drain-source V_{DS} (I_d - V_d) de la figure 1.19.b il est marqué par une bosse en régime de saturation.

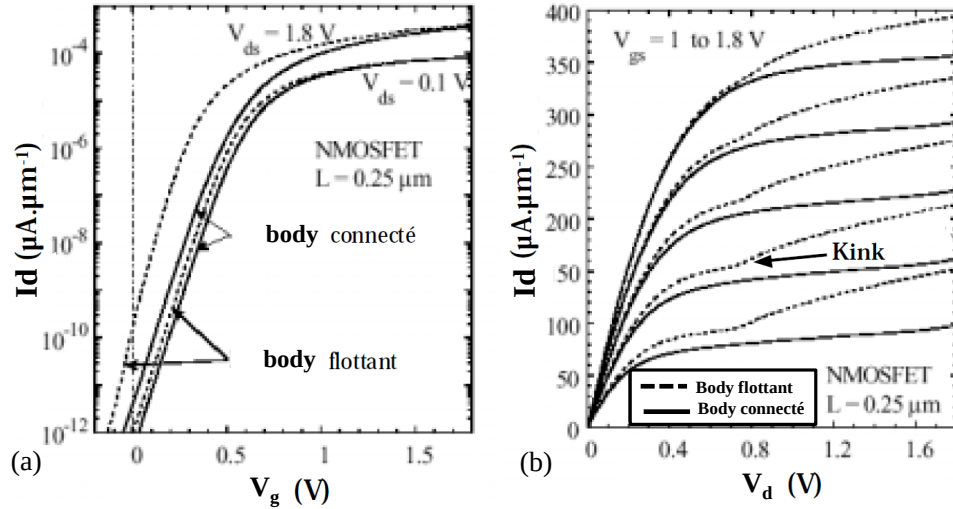


Figure 1.19 : (a) Caractéristique de transfert I_d - V_g d'un transistor nMOS PDSOI avec effet « kink » : la pente sous le seuil augmente au-delà de la limite théorique classique des 60 mV/décade à cause de l'ionisation par impact. (b) Caractéristique de sortie I_d - V_d du même transistor montrant l'effet « kink » : le déclenchement de l'ionisation par impact se traduit par un rebond de courant. [Villaret 04].

A noter que la charge majoritaire qui est générée et piégée dans le film de silicium flottant va induire une augmentation du potentiel électrostatique du film de silicium, ce qui a pour effet de réduire la barrière de potentiel de la diode entre le film de silicium et la source (figure 1.20). Dès que cette barrière est suffisamment faible, la diode passe en direct, le surplus de charge généré est évacué dans la source, et le niveau de charge du film de silicium flottant reste constant. La charge majoritaire piégée dans le film de silicium a donc un effet auto limité, et on parle d'auto-polarisation du film de silicium.

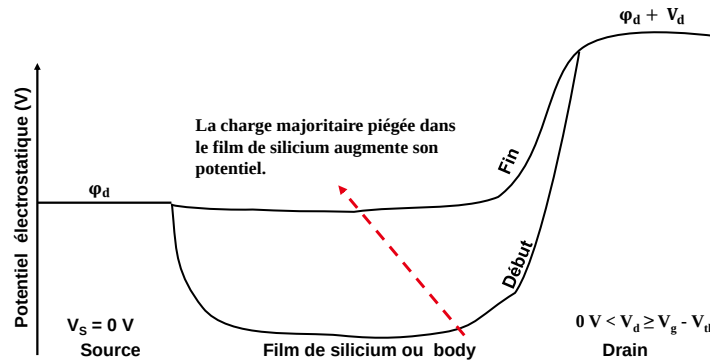


Figure 1.20 : Mise en évidence de l'auto-polarisation du film de silicium. En début de la génération de la charge, la jonction source-body a une barrière ϕ_d d'une jonction pn. A cause de la charge piégée dans, le body son potentiel électrostatique va augmenter mettant en direct la diode pn.

En résumé, la partie flottante du film de silicium dans un transistor PDSOI dégrade les performances du transistor à cause de l'autopolarisation du body. Notons que l'effet Kink est aussi observé lorsque le transistor est polarisé en mode bloqué. Car, comme nous l'avons vu précédemment, nous pouvons avoir une génération de la charge par le courant de fuite dû au GIDL. Il est possible de résoudre ce problème d'autopolarisation du body en le connectant à la masse. Mais, l'ajout d'un contact sur le body augmente la surface occupée par le transistor, de plus il augmente aussi le nombre de jeux de masque nécessaire à la fabrication du transistor, et par conséquent le coût de fabrication. C'est pourquoi il était intéressant pour les industriels de se tourner vers des technologies de transistor MOSFET sans une zone neutre dans le film de silicium : on parle de transistor MOSFET FDSOI (Fully Depleted Silicon On Insulator).

I.3.1.3. Transistor MOSFET FDSOI

L'architecture du transistor MOSFET FDSOI est similaire à celui d'un transistor PDSOI mais le film de silicium est complètement déserté car très mince (une épaisseur inférieure à 10 nm) [Planes 12]. Par rapport à un transistor MOSFET PDSOI ou sur substrat massif, nous avons un meilleur contrôle des effets de canaux courts, pas de variabilité du dopage (RDF) car le film de silicium est 'intrinsèque' [Weber 08]. Nous avons aussi une flexibilité apportée par la polarisation face arrière si le BOX est mince, dans ce cas on parle de transistor FDSOI UTBOX (*Ultra Thin BOX*) [Fenouillet 12].

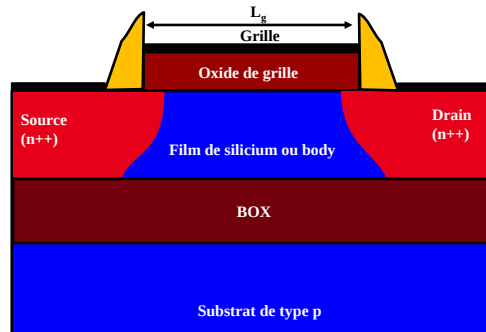


Figure 1.21 : Architecture d'un transistor MOSFET FDSOI.

Cependant, pour des nœuds technologiques plus avancés (figure 1.7), même le transistor FDSOI planaire ne peut plus être utilisé à cause de la dégradation du contrôle électrostatique de la grille (dégradation de la pente sous le seuil et I_{off}). C'est pourquoi les transistors MOSFET de type 3D tel que le FinFET sont utilisés aujourd'hui. Pour atteindre la fin de la roadmap [IRTS 2015], les technologies nanofils semblent être les plus adaptées.

I.3.1.4. Transistor MOSFET NANOFIL

La figure 1.22 nous montre la structure d'un transistor MOSFET nanofil : le film de silicium (canal) est totalement enrobé par la grille. Cette structure particulière du transistor nanofil lui confère un bon contrôle électrostatique. [Barraud 13] démontre qu'il est possible d'obtenir l'inverse de la pente sous le seuil (SS) de 80 mV/dec avec une structure nanofil de longueur $L_g = 10$ nm.

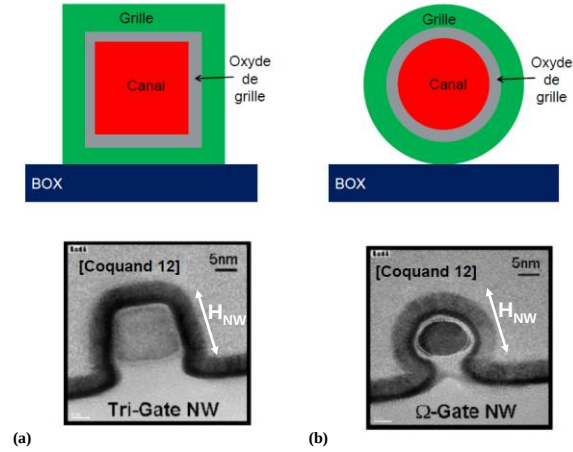


Figure 1.22 : Architectures et images TEM d'un transistor MOSFET NANOFIL (a) à section carré, et (b) à section circulaire [Lacord 12].

La figure 1.23 montre l'inverse de la pente du sous-seuil (SS) sur un pMOS pour $H_{NW} = 11$ nm pour une longueur du nanofil de 20 nm en fonction de sa largeur, de 10 nm (structure 3D) à 250 nm (dispositif planaire similaire au FDSOI). Le SS varie de 70 mV/dec pour le nanofil le plus étroit, à 125 mV/dec pour la plus grande largeur. Ce résultat met en évidence l'amélioration du contrôle électrostatique qu'on peut obtenir avec une structure 3D par rapport à une structure totalement planaire. L'évolution du DIBL pour un nanofil de type n et p confirme aussi cette observation, car on réussit à maintenir sa valeur constante jusqu'à une longueur de grille de 30 nm.

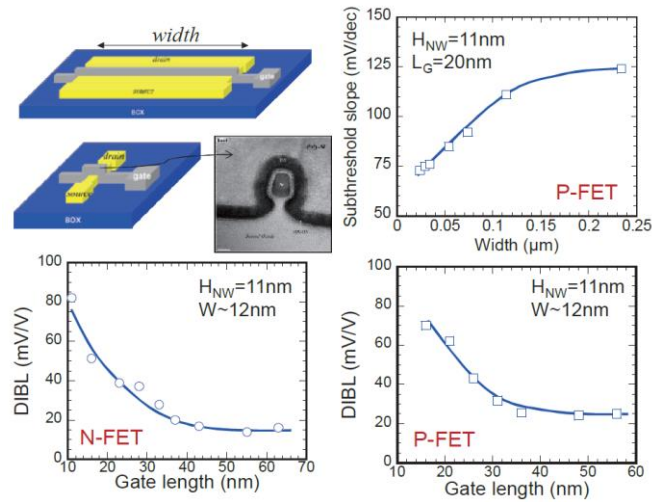


Figure 1.23 : Variation de la pente sous le seuil, et du DIBL en fonction de la longueur de grille mesurée pour des nMOS et pMOS [Barraud 13].

I.3.2. Mémoires 1T – DRAM

Nous avons vu que la mémoire DRAM standard 1T-1C impose des challenges technologiques principalement à cause de la capacité de stockage [Bawedin 11]. C'est pourquoi il y a de cela plusieurs années [Wann 93], une nouvelle génération de DRAM utilisant un seul transistor, appelée 1T-DRAM a été proposée comme alternative à l'architecture DRAM 1T-1C actuelle. L'architecture d'un point mémoire 1T-DRAM peut donc être un transistor MOSFET de type PDSOI [Wann 93], FDSOI [Bawedin 05] ou encore sur substrat massif [Malinge 05]. On utilise les effets parasites du transistor MOSFET pour créer un effet mémoire. L'accès au point mémoire se fait toujours via le même transistor MOSFET. Etant composé seulement d'un transistor, les cellules 1T-DRAM sont compatibles avec la technologie CMOS. Pour la suite nous allons parler d'architecture 1T-DRAM avec un transistor MOSFET sur substrat SOI (PDSOI et FDSOI), mais les opérations mémoires qui seront décrites pour ces architectures sont aussi valable pour une architecture 1T-DRAM avec un transistor MOSFET sur substrat massif.

Lorsque le film de silicium d'un transistor sur substrat SOI n'est pas connecté à une source de tension on dit qu'il est flottant. Durant le fonctionnement du transistor, comme illustré précédemment, on peut avoir génération de la charge par des effets parasites. Dans le cas d'un transistor MOSFET PDSOI de type n par exemple, lorsque la charge est stockée dans le film de silicium flottant (figure 1.24.a), son potentiel électrostatique va augmenter (courbe en bleu sur la figure 1.24.c). Par conséquent la tension seuil diminue, et le courant débité par le transistor augmente et atteint la valeur I_1 (figure 1.24.d). A contrario, si le film de silicium n'est pas chargé (figure 1.24.b), son potentiel électrostatique est inchangé (courbe en rouge figure 1.24.c) et il n'y a aucune variation de la tension seuil : le courant du transistor reste à sa valeur I_0 (figure 1.24.b). C'est cette différence de courant qui est utilisée pour lire l'information.

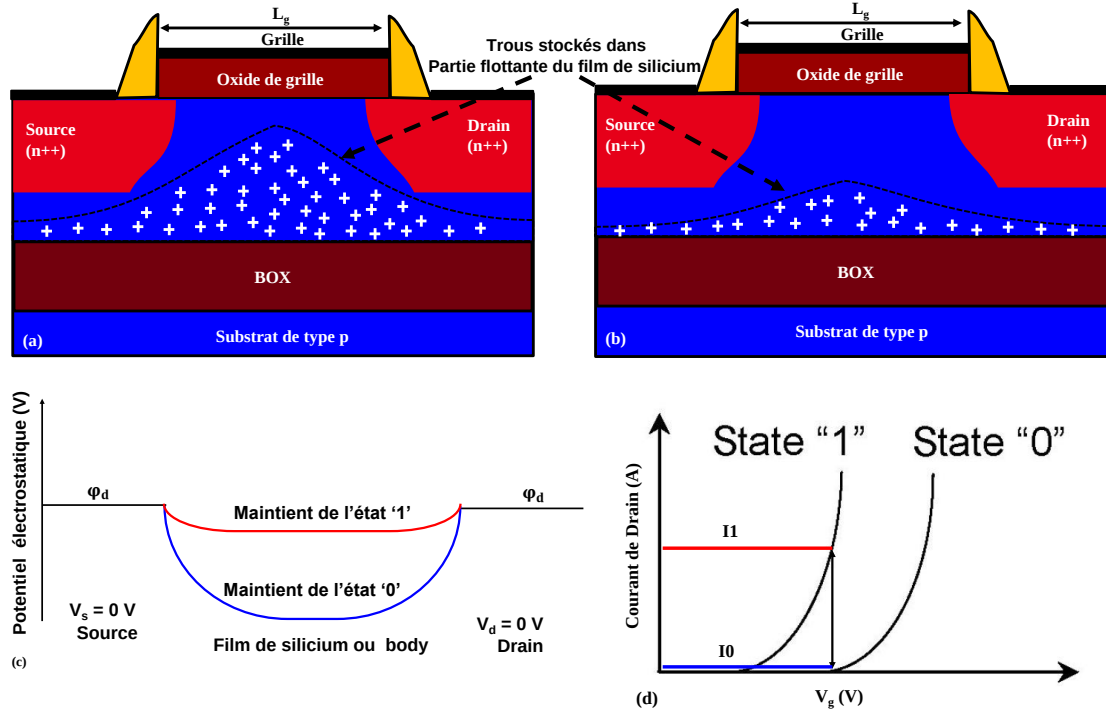


Figure 1.24 : (a) Film de silicium avec un surplus de charge : Stockage du '1'. (b) Film de silicium sans surplus de charge : Stockage du '0'. (c) Profil du potentiel électrostatique entre source et drain pour le stockage du '0' (courbe en bleu) pour le stockage du '1' (courbe en rouge). (d) Caractéristique courant de drain-tension de grille : pour les états '0' et '1'.

Pour la suite, nous allons présenter brièvement les mécanismes de programmation des états '1' et '0' dans une 1T-DRAM. Il convient de signaler que durant la programmation de l'état '1', la quantité de charge pouvant être 'stockée' est limitée par l'auto-polarisation du film silicium [Malinge 05] comme nous l'avons souligné dans la section 3.1.2.

I.3.2.1. Programmation du '1' : Ionisation par Impact (II)

Ce principe de programmation de l'information par ionisation par impact est basé sur l'effet 'Kink' (figure 1.19). C'est un mécanisme très utilisé dans les 1T-DRAM car il garantit une vitesse de stockage très élevée. Durant la programmation de l'état '1', des tensions positives et relativement élevées sont appliquées sur les contacts de drain V_{DW1} et de grille V_{GW1} pour mettre le transistor en régime de saturation. Grâce au phénomène d'ionisation par impact dans la zone de désertion du canal (zone de pincement ou pinch-off), on aura la génération de paires électrons-trous. Les trous générés sont refoulés et piégés dans la zone du film de silicium flottant car la jonction body-drain est en inverse. Durant son parcours, un trou peut aussi acquérir suffisamment d'énergie pour pouvoir générer une paire électron-trou à son tour. De même, l'électron généré dans la zone de désertion sera accéléré et recueilli au drain, et pourra lui aussi générer une paire électron-trou.

Pour garantir un bon stockage de la charge, il faut maintenir un champ électrique élevé dans la zone de charge d'espace, et sa longueur $L_{pinch-off}$ (figure 1.16) doit être suffisamment grande. Cela revient donc à maximiser la tension de drain (V_{DW1}). Par ailleurs, l'écriture de l'état '1' par ionisation par impact implique une forte consommation d'énergie électrique (car

le transistor est en régime de saturation pendant la programmation). La fiabilité peut aussi être limitée par la génération de porteurs chauds qui vont détériorer l'oxyde de grille.

I.3.2.2. Programmation du '1': transistor à jonction bipolaire (BJT)

Cette méthode de programmation de l'état '1' tire parti de l'effet parasite du transistor bipolaire [SZE 07] (en anglais *BJT : Bipolar Junction Transistor*) qui peut être activé dans la partie flottante d'un transistor MOSFET. Cet effet est aussi décrit dans la littérature comme étant un effet 'kink' secondaire [Tihanyi 75]. Mais, pour les applications 1T – DRAM, il est utilisé comme une technique de programmation de l'état '1' avec l'avantage d'avoir une faible consommation énergétique, car le transistor reste dans un état bloqué ($V_{GW1} < V_{FB} < V_{th}$). Ce mécanisme est aussi présenté comme adapté pour les architectures 1T – DRAM non-planaires telles que : FinFET, grilles multiples, et grille enrobante [Okhonin 07].

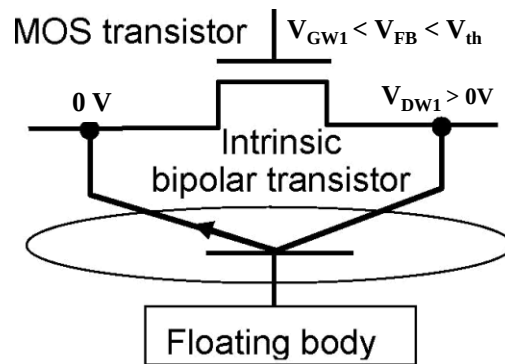


Figure 1.25 : Programmation de l'état '1' par l'amplification du transistor bipolaire parasite intrinsèque au transistor MOSFET sur substrat massif ou PDSOI sans prise de contact sur le substrat et film de silicium.

Comme le montre la figure 1.25, pour pouvoir générer de la charge majoritaire dans le film de silicium flottant (la base), la grille face avant est polarisée négativement maintenant ainsi le film de silicium flottant en régime d'accumulation. Le drain (collecteur) est polarisé à une forte tension positive, et la source (émetteur) à la masse. Dans cette condition, il existe une barrière de potentiel de jonction pn entre la source et le film de silicium flottant ; en revanche, la jonction film de silicium-drain est polarisée en inverse créant une zone de charge d'espace (ZCE). Comme pour le cas précédent, si le champ électrique dans le ZCE est élevé, les électrons venant du film de silicium flottant et/ou ceux générés thermiquement dans la ZCE seront accélérées, et s'ils acquièrent suffisamment d'énergie, ils généreront des paires électron-trou par ionisation par impact. Les trous générés dans la ZCE sont recueillis dans le film de silicium flottant, ceci a pour effet d'augmenter son potentiel électrostatique réduisant ainsi la barrière de potentiel entre la jonction source-film de silicium flottant. La réduction de la barrière va augmenter le courant d'électron venant de la source, et produira encore plus de paires électron-trou. On a donc une amplification du mécanisme de génération grâce à la présence du transistor boltaire npn (source-body-drain).

I.3.2.3. Programmation du '1' : boucle de rétroaction positive de la tension seuil du transistor (V_{TH})

Cette technique de programmation est analogue à la précédente, mais elle est activée seulement dans le cas des architectures de cellule mémoire 1T-DRAM de type FDSOI avec une 'mince' couche d'oxyde enterré [Lu 10] : on parle de UTBOX FDSOI. Elle a donc l'avantage de pouvoir utiliser la grille face arrière pour induire l'effet substrat flottant, tout en garantissant l'utilisation de faibles polarisations.

Pour l'activer, dans le cas d'un transistor de type n par exemple, la grille face avant est polarisée en régime de faible inversion ($V_{GW1} < V_{TH}$). La grille face arrière à une forte polarisation négative V_{GB} pour maintenir l'interface arrière en régime d'accumulation. Dans ce régime de fonctionnement, il existe un courant dans le canal face avant. Si la tension de drain est suffisamment grande, les électrons qui sont accélérés de la source vers le drain pourrons générer des paires électron-trou. Les trous générés par ionisation par impact dans la ZCE au niveau du drain sont attirés et accumulés au niveau de l'interface arrière du film de silicium-oxyde enterré à cause de la polarisation V_{GB} . Il en résulte une augmentation du potentiel du film de silicium, et en suivra une réduction de la tension seuil. La réduction de la tension seuil, permettra d'injecter plus d'électron dans le canal, et ainsi d'augmenter la génération des paires électrons trous dans la ZCE.

I.3.2.4. Programmation du '1' : Effet Tunnel Bande à Bande (B2B)

Ce principe de programmation de l'information par génération tunnel bande à bande utilise un effet parasite du fonctionnement classique du transistor MOS : le GIDL (*Gate Induced Drain Leakage*) vu précédemment (figure 1.15). Lors de l'écriture de l'état '1' pour une cellule de type n par exemple, on devra appliquer sur la grille une tension négative V_{GW1} , et sur le drain une tension V_{DW1} positive. La polarisation grille/drain va induire une génération par effet tunnel dans la zone de recouvrement du drain et de la grille. Les électrons seront recueillis au drain et les trous générés seront piégés dans le film de silicium flottant ce qui va constituer le stockage de l'état '1'.

Par ailleurs, le mécanisme de stockage par effet tunnel bande à bande peut aussi être imbriqué à une amplification bipolaire, et à la boucle rétroactive positive de la tension seuil du transistor [Bawedin 09] tous deux présentés aux sections 3.1.2 et 3.1.3. L'avantage de ce mécanisme de stockage est la très faible consommation énergétique car le transistor reste dans un état bloqué, et le courant issue de la génération est faible comparé aux cas précédents. Initialement présenté comme un mécanisme avec une faible vitesse d'écriture, il a été démontré qu'il peut aussi avoir des temps d'écriture très faibles [Yoshida 03], [Puget 09].

I.3.2.5. Programmation du '0' ou Effacement

Pour programmer l'état '0', il faut évacuer les porteurs majoritaires stockés du film de silicium flottant. Il en résulte une diminution du potentiel électrostatique du substrat flottant et donc la tension de seuil augmente. Dans la littérature, deux méthodes sont disponibles pour évacuer la charge majoritaire du film de silicium flottant [Bawedin 09].

- La première consiste à polariser en direct les jonctions source-film et drain-film de silicium flottant. Afin d'évacuer la charge majoritaire du film de silicium flottant, dans le cas par exemple d'un transistor de type n, la tension de la grille face avant est

- maintenue positive ($\geq V_{TH}$) alors qu'une tension négative est appliquée sur le contact de drain. Les jonctions film de silicium flottant-drain et film de silicium flottant-source deviennent respectivement polarisées en direct et en inverse. En conséquence, le courant de trous injecté depuis la source à l'intérieur du corps est beaucoup plus faible que celui extrait par la jonction de drain. L'évacuation étant lente, cette méthode ne peut être utilisée dans notre cas car la durée de l'effacement nécessaire n'est pas compatible avec une opération DRAM standard.
- La seconde technique consiste à utiliser le couplage dynamique entre les grilles face avant et arrière pour évacuer la charge majoritaire du film de silicium flottant. Considérant un transistor de type n, une impulsion positive sera appliquée sur la grille face avant, tout en maintenant le drain et la source à la masse. Au fur et à mesure que le potentiel de la grille face avant augmente, l'interface oxyde de grille-film de silicium entre en régime de désertion et les jonctions film de silicium flottant-drain et film de silicium flottant-source sont polarisées en direct. Par conséquent, les trous sont évacués à travers les jonctions. Cette technique est beaucoup utilisée, car plus efficace.

En général, le mécanisme d'effacement reste le même quel que soit l'architecture du transistor MOSFET considérée. Pour l'état '1' en revanche, il varie selon le type du transistor qui constitue la cellule mémoire. Ceci est dû principalement au fait que selon l'architecture du transistor l'effet de film de silicium flottant se manifeste différemment. La conséquence directe c'est que le mécanisme utilisé pour lire l'information est directement lié à celui du stockage de charge comme indiqué dans le tableau 1.1 [Lu 10].

	[Yoshida 03] [Okhonin 01]	[Fazan 01]	[Ranica 04]	[Matsuoka 07] [Ban 08]	[Okhonin 07] [Song 08] [Jang 09] [Mohapatra 09]	[Lu 10]
Architectures	PDSOI	PDSOI	FDSOI à BOX épais	FDSOI à BOX mince	PDSOI	FDSOI à BOX mince
Ecriture de l'état '1'	B2B	II			BJT	V _{TH}
Ecriture de l'état '0'	Diode en direct ou couplage capacitif entre les grilles face avant et arrière.					
Mécanisme de lecture de l'information	Modulation du courant du canal dû au décalage de la tension seuil du transistor.				Courant qui circule dans le BJT.	

Tableau 1.1. Comparaison des opérations mémoires entre les différents architecture 1T – DRAM de la littérature (pas exhaustive) utilisant l'effet substrat flottant pour stocker la charge.

I.4. Pourquoi l'architecture A2RAM comme 1T – DRAM

I.4.1. Mémoire Z-RAM

L'une des premières architectures de mémoire 1T-DRAM est la Z-RAM (Z pour «zéro capacité») [Okhonin 01]. Son architecture est similaire à un transistor MOSFET PDSOI, et les mécanismes de programmation de l'information sont l'ionisation par impact pour l'état '1', et les diodes en direct pour l'état '0'. L'inconvénient avec cette mémoire est le fait d'avoir une fenêtre de programmation très faible : $I_1/I_0 \approx 2$ comme on peut le voir sur la figure 1.26. Pour pouvoir améliorer sa fenêtre mémoire, la ZRAM2 (ZRAM seconde génération) a été proposé [Okhonin 07]. La différence avec l'architecture ZRAM se trouve uniquement au niveau de la programmation de l'état '1' qui se fait grâce à l'activation du transistor bipolaire intrinsèque à la cellule mémoire. Cependant, lorsque l'on réduit la longueur de la grille de la cellule ZRAM, pour garder le contrôle électrostatique sur le canal il faut réduire l'épaisseur du film de silicium. Dans cette condition, l'épaisseur du substrat flottant sera faible y compris la quantité de charge pouvant être stockée ; ceci a pour conséquence une forte diminution de la fenêtre de programmation. C'est pour cette raison que des nouvelles cellules 1T-DRAM sur substrat SOI à film mince ou FDSOI ont vues le jour.

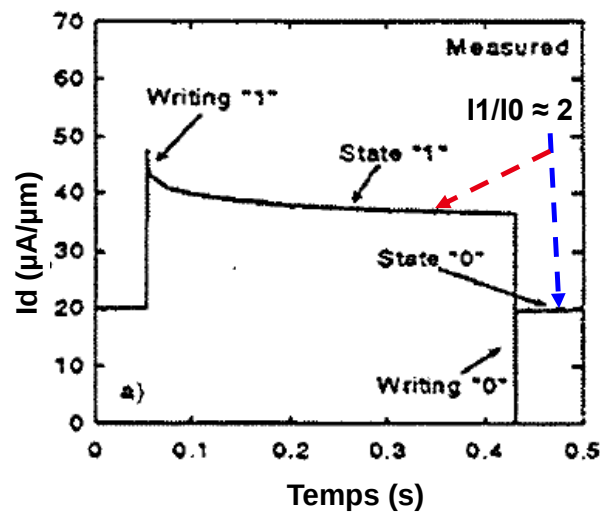


Figure 1.26 : Courant de drain en fonction du temps extrait de mesure électrique d'une cellule ZRAM avec $L_g = 0.5 \mu m$, $W = 25 \mu m$ [Okhonin 01].

I.4.2. Architecture mémoire 1T – DRAM de type FDSOI

I.4.2.1. Architecture mémoire MSDRAM

La cellule MSDRAM a été proposée par [Bawedin 05], [Bawedin 06]. Elle est similaire à un transistor FDSOI (figure 1.18). Grâce à la possibilité d'avoir une prise de contact en dessous du BOX, l'effet mémoire est induit par le couplage capacitif entre les grilles face avant (V_{G1}) et arrière (V_{G2}). La programmation de l'état '1' se fait par la génération de la charge par effet tunnel bande à bande. Pour l'état '0', on utilise le couplage dynamique entre les grilles face avant et arrière pour évacuer la charge stockée dans le film de silicium. La lecture est effectuée en utilisant la tension de la grille face arrière (V_{G2}) pour activer le canal arrière :

lorsque l'on a stocké l'état '1' il y'aura un courant dans le canal d'inversion en face arrière, sinon le canal ne s'active pas et on lira l'état '0'.

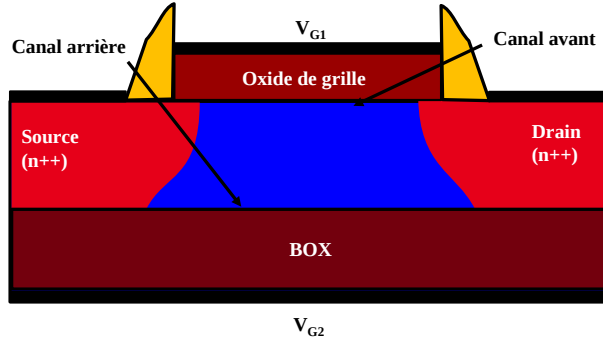


Figure 1.27 : Cellule mémoire MSDRAM.

Comme pour toutes architectures 1T – DRAM, pour augmenter la densité d'intégration de la MSDRAM il faut réduire la longueur de la grille, donc l'épaisseur du film de silicium afin de garder le contrôle électrostatique. Cependant, pour les architectures mémoires à double grille comme la MSDRAM, on est limité par le phénomène de super-couplage ou '*Super-coupling*' [Eminente 07]. Le super-couplage signifie l'impossibilité d'avoir simultanément dans un film de silicium des porteurs majoritaires (couche d'accumulation) et des porteurs minoritaires (couche d'inversion). Ce qui est le cas avec la MSDRAM, car les trous stockés forment la couche d'accumulation, et le courant lu en face arrière est dû aux porteurs minoritaires de la couche d'inversion. [Eminente 07] donne l'expression de la valeur approximative de l'épaisseur du film de silicium limite T_{si}^* en dessous de laquelle le '*Super-coupling*' apparaît :

$$T_{si}^* = \frac{k_B T \epsilon_{Si}}{q C_{ox} (V_{G1} - V_{G2})} \ln \frac{N_{body} C_{ox} C_{box} (V_{G1} - V_{G2})}{q \epsilon_{Si} n_i^2} \quad (1.4)$$

où k_B est la constante de Boltzman, T la température, ϵ_{Si} la permittivité du silicium, N_{body} le dopage du film de silicium, C_{ox} et C_{box} les capacités de l'oxyde de grille de la face avant et arrière respectivement, q la charge élémentaire, et n_i la concentration intrinsèque du silicium. L'équation 1.4 en plus d'avoir une dépendance avec les paramètres technologiques du transistor est gouvernée par les tensions. Par ailleurs, [Navarro 15] démontre que la valeur de T_{si}^* dépend aussi de la longueur de la grille à cause des effets canaux courts SCEs (*Short Channel Effects*). Il est alors difficile de maintenir les performances de la MSDRAM pour des épaisseurs de film de silicium inférieures à une vingtaine de nanomètres. Pour résoudre le problème de '*Super-coupling*', [Rodriguez 10] propose l'ARAM.

I.4.2.2. Architecture mémoire ARAM

L'architecture de l'ARAM est similaire à la MSDRAM, excepté le film mince divisé en deux parties par une couche d'oxyde (figure 1.28) [Rodriguez 10], [Rodriguez 11]. Ici, on fait une séparation physique des deux types de porteurs avec une couche d'oxyde au milieu du film le

silicium. Le film de silicium du dessus est utilisé pour le stockage des trous, et celui en dessous pour le passage du courant en phase de lecture. La programmation de l'état '1' se fait par effet tunnel bande à bande, et pour l'état '0' on utilise le couplage dynamique entre les grilles avant et arrière. Pour la lecture de l'information, la présence de la charge dans la partie supérieure du film de silicium induit électrostatiquement de la charge négative dans le film de silicium en dessous. Lorsqu'on polarise le drain légèrement positivement, il y aura un courant entre la source-film de silicium inférieur-drain : on lira le courant I_1 . Dans le cas contraire, le film de silicium inférieur reste déserté et le courant I_0 lu au drain faible.

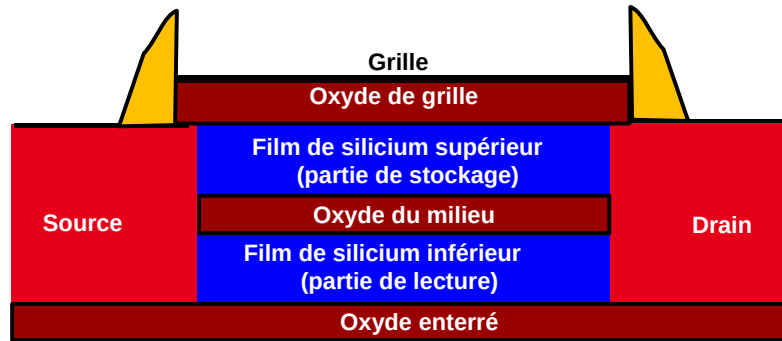


Figure 1.28 : Cellule mémoire ARAM.

Mais vue la complexité technologique que demande la fabrication de cette mémoire, [Rodriguez (1) 11], [Rodriguez (2) 11] propose l'A2RAM (avec A2 pour Seconde Génération).

I.4.2.3. Architecture mémoire A2RAM

La structure d'une cellule A2RAM de type n est présentée sur la figure 1.29, son architecture consiste en un transistor conventionnel sur substrat SOI, mais avec le film de silicium divisé en deux parties. La partie supérieure communément appelé 'body' est intrinsèque ou de dopage opposé à la source et au drain. La seconde partie au fond du film de silicium de même type de dopage que la source et drain est appelée le 'bridge'. Cette dernière est utilisée pour court-circuiter la source et le drain durant la lecture de l'information.

A l'état '1', le body de l'A2RAM est chargé de porteurs majoritaires contrairement à l'état '0'. La charge majoritaire est générée de deux manières différentes : par ionisation par impact ou par effet tunnel bande à bande. Pour l'effacement, on peut soit polariser les diodes en direct, soit utiliser le couplage dynamique entre les grilles face avant et arrière comme expliqué précédemment.

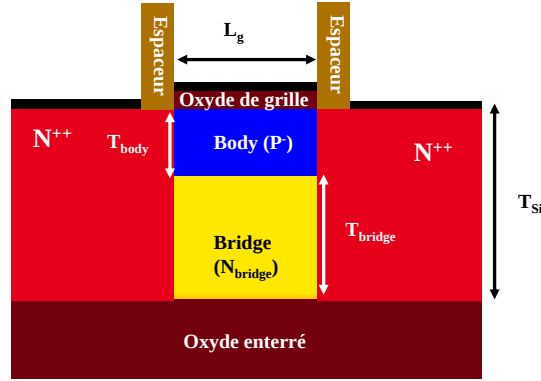


Figure 1.29 : Point mémoire A2RAM où sont indiqués la longueur de la grille L_g , l'épaisseur du body T_{body} , l'épaisseur T_{bridge} et le dopage N_{bridge} du bridge ; $T_{Si} = T_{body} + T_{bridge}$.

Durant la lecture de l'information, la grille est polarisée négativement et le drain est polarisé légèrement positivement. Si le body n'est pas chargé (état '0'), le champ électrique vertical E_{elec} due à la grille (figure 1.30.a) viendra couper électrostatiquement la conduction dans le bridge, et il n'y aura pas ou peu de courant entre la source et le drain à travers le bridge. A contrario, si de la charge est stockée dans le body (état '1'), le champ électrique vertical E_{elec} due à la grille sera écranté (figure 1.30.b). De ce fait, le champ électrique ne sera plus suffisant pour couper la conduction dans le bridge, et un fort courant circulera entre la source et drain à travers le bridge.

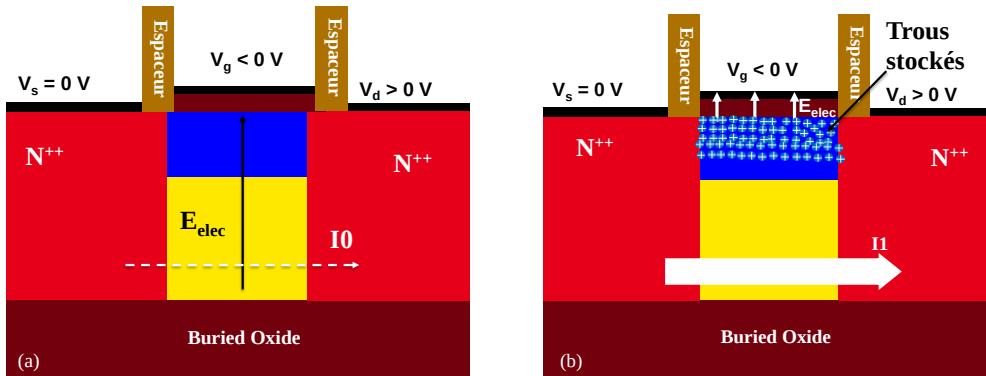


Figure 1.30 : Point mémoire A2RAM lecture de (a) l'état '0', (b) l'état '1'.

Dans le prochain chapitre, nous allons décrire le rôle et l'impact des polarisations appliquées sur l'A2RAM pendant chaque opération mémoire afin d'expliquer son fonctionnement plus finement en s'appuyant sur des simulations TCAD.

Dans cette section 4, nous avons présenté des structures 1T-DRAM dont le principe de fonctionnement et les architectures découlent du transistor MOSFET sur substrat SOI. Mais, comme nous l'avons déjà introduit de la section 3, ces architectures mémoires peuvent être

fabriquées à base de transistors sur substrat massif [Villaret 04], [Malinge 05], [Navarro 14]. Par ailleurs, il existe aussi des cellules 1T-DRAM dont les architectures sont non conventionnelles telle que la Z²-FET [Wan 12]. La cellule Z²-FET, avec Z² pour « Zéro » variation de la tension sous seuil et « Zéro » génération d'ionisation par impact, est constituée d'une diode p-i-n latérale, où la partie intrinsèque est partiellement recouverte sur une longueur L_g par la grille face avant. Cette architecture fait partie de cellules mémoires retenues dans le cadre du projet REMINDER [Lacord 18], pour évaluer ses performances dans le but de vérifier si elle peut être une candidate pour remplacer la DRAM pour les applications de type internet des objets.

I.5. Conclusion

Dans ce premier chapitre, nous avons résumé l'évolution technologique de la cellule élémentaire DRAM 1T/1C. Même si cette technologie continue à être industrialisée, l'élément principal limitant ses performances et augmentant son coût de fabrication est la capacité de stockage. Ainsi, dès le début des années 2000 la première architecture 1T-DRAM a été proposée. Cette architecture mémoire DRAM permet le stockage et la lecture de l'information par le biais du même transistor grâce à l'effet dit de 'substrat flottant'. Comparé à la cellule standard, elle n'a pas besoin d'une capacité de stockage d'où l'énorme avantage en termes de densité d'intégration et de simplicité de fabrication. Mais les études d'optimisation faites sur la ZRAM n'ont pas révélé des performances comparables à celle d'une DRAM standard. C'est dans cette optique que la ZRAM2 a été proposée, mais elle a également été rapidement rattrapée par la dégradation de ses performances lorsqu'on réduit ses dimensions. Pour résoudre ce problème, les structures à double grilles ont vu le jour. La MSDRAM en fait partie ; bien qu'elle prévoyait des meilleures performances ; lorsque l'on réduit ses dimensions, le 'Super-coupling' représente une limite fondamentale. Phrase en rouge à revoir, je ne comprends pas ce que tu veux dire. Faut dire quelle dimension est limitée par le supercoupling

Pour faire face à ce phénomène de 'Super-coupling', l'ARAM a été proposée mais présente de sérieux challenges au niveau de son procédé de fabrication. C'est pourquoi l'A2RAM a été proposée, son concept a déjà été démontrée expérimentalement. Mais des études d'optimisations sont encore nécessaires si on veut la rendre compétitive face à la cellule standard DRAM.

L'objectif de ce manuscrit de thèse est d'une part de présenter l'étude d'optimisation de l'A2RAM. Le but étant de mieux comprendre les challenges technologiques auxquels on doit faire face si l'on veut réduire ses dimensions. D'autre part, nous présenterons aussi les différents résultats de caractérisation de la cellule mémoire en vue de mieux appréhender l'évolution de ses performances en fonction de ses paramètres technologiques. Une fois ces deux éléments consolidés, nous aborderons la partie modélisation compacte de la cellule mémoire. Par ailleurs, grâce au projet REMINDER, durant cette thèse nous avons aussi travaillé sur la fabrication d'architecture mémoire de type 3D, cet élément fait objet d'étude dans le dernier chapitre de ce manuscrit.

CHAPITRE II :
DESCRIPTION DU
FONCTIONNEMENT
ELECTRIQUE D'UN POINT
MEMOIRE A2RAM VIA
SIMULATIONS TCAD

Dans ce chapitre, nous ferons une étude de performances mémoires de la cellule A2RAM via un outil de simulation TCAD. Nous commencerons par présenter l'outil de simulation TCAD. Ensuite, nous détaillerons la méthodologie de simulation mise en place, qui sera validée par la caractérisation électrique. Nous finirons avec une étude de la sensibilité des performances du point mémoire A2RAM aux variations de tensions lors des différentes opérations.

II.1. Stratégie de simulation

II.1.1 Description du logiciel de simulation

Dans la course à la miniaturisation, le nombre de composants par puce n'a pas seulement augmenté, mais le processus de fabrication de ces composants s'est également complexifié. Il devient alors de plus en plus difficile de pouvoir prédire le comportement des composants à semi-conducteurs lorsqu'ils sont 'isolés', mais surtout lorsqu'ils sont dans un environnement circuit ou matrice. Ainsi, pour réduire les coûts de fabrication et limiter le « time to market » (le temps de développement d'un produit et le proposer sur le marché), les outils de simulation de type TCAD (*Technology Computer Aided Design*) sont nécessaires. Ils font référence à l'utilisation de simulations numériques à éléments finis pour développer et optimiser les technologies de composants à semi-conducteur. Dans ce marché d'outil TCAD, on retrouve essentiellement deux grands acteurs : SILVACO [SILVACO] et SYNOPSYS [SYNOPSYS]. Les simulations ont été réalisées dans le cadre de ce travail de thèse à l'aide des outils Synopsys. Il propose notamment une suite d'outils avec lesquels l'utilisateur peut effectuer des simulations allant du procédé de fabrication jusqu'au comportement électrique du dispositif. Pour ce travail sont utilisés :

- Synopsys (Sentaurus) Process (SP) :
Il s'agit d'un outil de simulation 2D et 3D des procédés de fabrication. On peut donc évaluer l'impact des variantes technologique sur une morphologie ou sur un profil de dopant par exemple. De plus, si chaque étape est simulée en implémentant le véritable procédé de fabrication, la structure 'réelle' du dispositif peut être générée. Cette structure est ensuite maillée pour pouvoir être utilisée comme point de départ d'une simulation électrique (sauvegardée sous un fichier .msh). Cet outil TCAD-Synopsys sera utile dans le Chapitre 6 où il sera question d'optimiser le processus d'implantation et d'activation du dopage du bridge.
- Synopsys (Sentaurus) Structure Editor (SDE) :
Ces outils servent à générer des structures à partir de données géométriques. Une fois la géométrie et les matériaux dans chaque région définis, il faut tour à tour définir les électrodes, le type et le profil de dopage dans chaque région, et enfin le maillage de la structure entière ainsi qu'aux différentes interfaces entre deux régions. La structure 'dessinée' pourra être utilisée pour une caractérisation électrique. L'utilisation de cet outil en principe présuppose que l'on connaisse déjà les différents paramètres technologiques devant constituer la structure à étudier. Dans cette hypothèse, nous gagnons en temps de simulation pour la même structure simulée avec l'outil Synopsys (Sentaurus) Process.
- Synopsys (Sentaurus) Device (SD) :

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

Il permet de simuler les caractéristiques électriques, thermiques et optiques de composants à semi-conducteur. Il reçoit en entrée le composant issu de Synopsys (Sentaurus) Process ou de Synopsys (Sentaurus) Structure Editor (SDE). Selon le type de caractérisation qui sera faite, une section physique nous permet d'inclure les modèles physiques nécessaires pour garantir des résultats de simulation fiables. Les résultats de la caractérisation sont sauvegarder dans les fichiers .plt pour les courbes, et/ou .tdr pour la structure entière 2D ou 3D à un instant, et pour un jeu de polarisations donnés.

- Synopsys (Sentaurus) Visual (SV) :
Il permet de visualiser les résultats de simulation obtenus par les outils précédents. On a la possibilité de visualiser directement des sorties du simulateur (courbes ou structures) et sauvegarder avec l'option de faire des post traitements automatisés.

II.1.2 Structure A2RAM simulée

Comme structure de référence, nous utilisons l'architecture A2RAM fabriquée en 2011 au Leti [Navarro 14] dont l'opération mémoire a été démontrée par caractérisation électrique dans [Rodriguez 12]. L'image TEM (Transmission Electron Microscopy) d'un échantillon est présentée sur la figure 2.1.a ; via l'outil Synopsys (Sentaurus) Structure Editor, nous avons reproduit la morphologie de celle-ci, et la structure obtenue est représentée sur la figure 2.1.b. Sachant que le dopage dans le bridge est important pour garantir de bonnes performances mémoires, nous nous sommes inspirés du profil de dopage dans la profondeur du film de silicium de la figure 2.1.c donné par [Rodriguez 12] qui prédit un dopage de bridge $N_{\text{bridge}} = 10^{18} \text{ cm}^{-3}$, et une épaisseur de bridge $T_{\text{bridge}} = 22 \text{ nm}$.

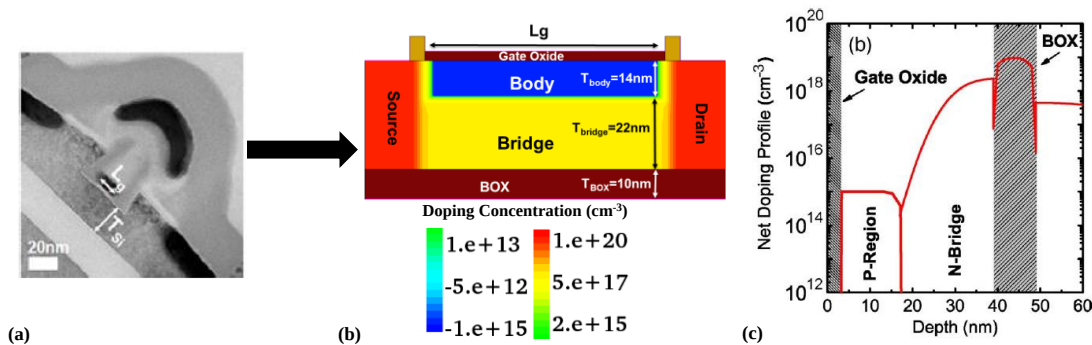


Figure 2.1 : (a) Image TEM du dispositif dont le processus de fabrication simplifié est issu de la technologie FDSOI du Leti [Navarro 14]. (b) Structure TCAD d'une A2RAM avec la définition de ses principaux paramètres. (c) Profil de dopage net final repris de [Rodriguez 12].

Dans le tableau 2.1, nous avons les valeurs des différents paramètres de la cellule A2RAM présentée sur la figure 2.1.b. Ils servent de paramètres de référence et seront optimisés au cours du Chapitre 3. Notons que les structures fabriquées ont un plan de masse dopé de type p, mais pour simplifier la structure simulée, nous avons mis en-dessous de la face arrière (en-

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

dessous de l'oxyde enterré BOX) un métal. Pour garantir la même condition de bande plate à l'interface face arrière (c'est-à-dire, oxyde enterré/métal dans notre cas, et oxyde enterré/silicium du plan de masse pour les échantillons fabriqués), nous avons pris un métal dont le travail de sortie (work-function) se trouve au milieu de la bande interdite d'un silicium intrinsèque (on parle de 'mid-gap metal').

Par ailleurs, étant donné qu'avec l'architecture A2RAM de la figure 2.1.b nous avons la possibilité d'appliquer un signal sur les deux faces, nous parlerons de grille face avant se trouvant au-dessus de l'oxyde de grille (au-dessus du le body) et de la grille face arrière en-dessous de l'oxyde enterré.

Paramètres	Valeurs
Longueur de masque de la grille (L_g)	80 nm
Largeur de masque de la grille (W)	1 μm
Longueur d'espaceur	5 nm
Épaisseur d'oxyde Equivalent (EOT)	3.1 nm
Épaisseur totale du film de silicium (T_{Si})	36 nm
Épaisseur du body (T_{body})	14 nm
Épaisseur du bridge (T_{bridge})	22 nm
Épaisseur d'oxyde enterré (T_{BOX})	10 nm
Dopage Source/Drain	$10^{18} cm^{-3}$
Dopage du body (P_{body})	$10^{15} cm^{-3}$
Dopage bridge (N_{bridge})	$10^{18} cm^{-3}$
Travail de sortie de la grille face avant	4.5 eV
Travail de sortie de la grille face arrière	Mid-gap metal

Tableau 2.1 : Définition et valeurs des différents paramètres de l'A2RAM de la figure 2.1.b.

II.1.3 Chronogramme de tensions pour l'étude de la mémoire A2RAM

Pour la caractérisation du point mémoire de la figure 2.1.b, nous allons utiliser le motif de tension présenté sur la figure 2.2; en bleu le créneau de tension appliqué sur le drain, et rouge celui appliqué sur la grille face avant ; la source et la grille face arrière sont à la masse.

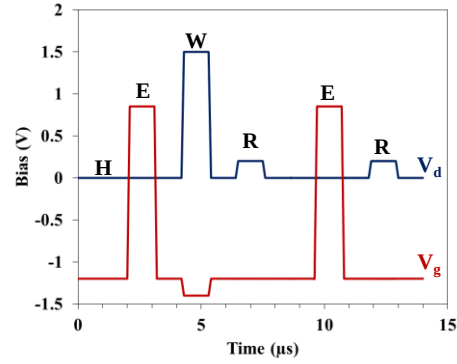


Figure 2.2 : Chronogramme de tensions appliqué sur le drain V_d (en bleu) et la grille face avant V_g (en rouge) pour la séquence d'opérations mémoire : Effacement (E) - Ecriture de l'état '1' (W) lecture de l'information (R) Effacement (E) lecture de l'information (R). entre chaque opération, une étape de maintien de l'opération (H) a lieu.

Les chronogrammes de tensions se composent des phases suivantes :

- ❖ Une phase d'effacement (E) ou programmation de l'état '0' : elle est effectuée par couplage capacitif des grilles face avant et arrière ($V_{g\text{Erase}} = 0.85 \text{ V}$, $V_{d\text{Erase}} = 0 \text{ V}$) avec la même condition définie dans [Rodriguez 12].
- ❖ La phase de programmation de l'état '1' (W). De nos études, nous avons constaté que le stockage de la charge par ionisation par impact n'était pas efficace (en plus de la forte consommation énergétique associée). C'est pourquoi nous allons utiliser la génération par effet tunnel bande à bande due au courant de fuite par GIDL. Il est donc nécessaire de maximiser le champ électrique grille-drain, ce qui signifie une tension de grille négative $V_{gW1} = -1.4 \text{ V}$ et une tension de drain positive $V_{dW1} = 1.5 \text{ V}$.
- ❖ La phase de maintien (H) est définie pour conserver des trous (informations) électrostatiquement dans le body. Il faut donc maintenir le body en régime d'accumulation, et garantir des barrières de potentiel élevées entre les jonctions body-source et body-drain. C'est pourquoi nous devons polariser la grille face avant à une tension inférieure à la tension de bande plate ici nous avons : $V_{g\text{Hold}} = -1.2 \text{ V}$.
- ❖ La phase de lecture (R) de l'état stocké ('1' ou '0'). Pendant la lecture, nous devons détecter le courant qui circule dans la résistance du bridge en maintenant le canal face avant du transistor à l'état bloqué. Le drain doit être polarisé légèrement positivement $V_{d\text{Read}} = 0.2 \text{ V}$ pour empêcher l'écriture parasite (via le courant fuite par GIDL) durant la lecture de l'état '0'. Pour lire l'état '0', il faut que la tension de la grille face avant de lecture $V_{g\text{Read}}$ soit choisi suffisamment négative pour permettre au champ électrique vertical de couper la conduction à travers le bridge. Par contre, pour lire l'état '1', la tension de lecture $V_{g\text{Read}}$ doit être suffisamment basse (en valeur absolue) afin que le champ électrique vertical soit efficacement écranté par des trous stockés dans le body pour permettre la conduction à travers le bridge. Nous avons choisi $V_{g\text{Read}} = V_{g\text{Hold}}$ pour faciliter le pilotage circuit.

Étant donné que l'état initial de la structure donnée par le simulateur n'est pas vide de charges dans le body, avant toute opération d'écriture ou de lecture il est impératif de faire une opération d'effacement (E), comme on peut le voir sur la figure 2.2. Ce motif de tensions sera utilisé pour la suite de nos études sauf mention particulière de notre part.

II.1.4 Définition des modèles physiques pour la simulation électrique d'opérations mémoires

Pour garantir des résultats de simulation proches de la réalité, il est important de bien définir les modèles utilisés dans les simulations. Ces modèles sont choisis en fonction du type de simulation, du composant à simuler et de l'application visée. Pour la caractérisation de la mémoire présentée sur la figure 2.1, et tenant compte des opérations mémoire de la figure 2.2, nous allons utiliser les modèles suivants :

- Mobilité des porteurs : du motif de tension de la figure 2.2 nous pouvons constater que durant toute les opérations mémoires il n'y a pas de courant d'inversion. Nous n'aurons pas besoin de tenir compte des modèles de mobilités liés au transport complexe en régime d'inversion. Mais, sachant que dans le film de silicium de l'A2RAM la couche de silicium fortement est dopée, nous tenons compte de la dépendance de la mobilité avec le dopage via le modèle de Masetti [Masetti 83]. Il nous sera utile pour évaluer l'impact du bridge dopé sur les performances mémoires de l'A2RAM. Notons que ce n'est pas l'unique modèle de mobilité qui tient compte de la dépendance avec le dopage, mais de nos études préliminaires nous avons noté que tous ces modèles aboutissent aux mêmes résultats.
- Effets de confinement quantique : nous n'avons pas inclus ce modèle dans nos simulations car le choix d'utiliser la programmation par le courant de fuite du GIDL implique que le canal face avant est toujours à l'état bloqué, donc il n'y a pas d'impact des effets quantiques liés à la courbure des bandes. Par ailleurs, les effets quantiques liés aux dimensions sont négligeables pour des épaisseurs supérieures à 10 nm.
- Courant de fuite de la grille : pour la simulation du temps de rétention dans une mémoire dynamique, il est important de tenir compte des fuites de la charge par effet tunnel via l'oxyde de grille face avant. Pour le choix du modèle, nous avons fait une étude d'évaluation du maintien de l'information dans la cellule A2RAM en utilisant tous les modèles de fuite de grille proposé par Synopsys. L'unique modèle qui faisait varier le courant à travers la grille était le modèle DirectTunneling. Cependant, cette étude via des simulations TCAD nous permet juste d'avoir des ordres de grandeurs, et nous savons que le comportement réel sera donné par les mesures électriques.
- Recombinaison : en plus du modèle par défaut de génération et recombinaison (SRH : Shockley-Read-Hall), nous avons inclus un modèle d'ionisation par impact et un autre de génération par effet tunnel bande à bande.
 - Synopsys propose l'utilisation de plusieurs modèles d'effet tunnel bande à bande. Pour la simulation du point mémoire A2RAM , nous avons choisi le modèle Nonlocalpath [Kane 61], avec des paramètres de simulation calibrés grâce à des projets internes au Leti. C'est un modèle qui décrit le mécanisme de génération par effet tunnel non-local direct et/ou indirect (assisté par les phonons). Il est capable de s'adapter en fonction du matériau utilisé, et il évalue

de manière dynamique le chemin plus probable pour avoir une génération de paire électron-trou.

- Synopsys comprend également plusieurs modèles d'ionisation par impact. En général, le taux de génération d'ionisation par impact s'exprime sous la forme suivante :

$$G_{ii} = \alpha_n n v_n + \alpha_p p v_p, \quad (2.1)$$

avec α_n et α_p des coefficients d'ionisation par impact, v_n et v_p la vitesse des électrons et des trous ; n et p sont les concentrations d'électrons et de trous. Les valeurs des coefficients α_n et α_p dépendent des modèles de génération par ionisation par impact et du modèle utilisé pour évaluer la composante du champ électrique parallèle à la direction de déplacement de la particule (la Driving force). Si on fixe le modèle utilisé pour évaluer cette 'Driving force' et que l'on change de modèle de génération par ionisation par impact, les valeurs de ces coefficients n'altèrent pas de manière significative la valeur de G_{ii} . Cette observation a fait l'objet d'une vérification en début de thèse ; nous avons constaté qu'en changeant de modèle de génération par ionisation par impact, les performances mémoires étaient les mêmes. Par contre, si nous fixons les modèles de génération par ionisation par impact, et que l'on change le modèle d'évaluation de la 'Driving force', les coefficients d'ionisation par impact varient énormément et on observe des différences significatives sur le G_{ii} . En effet, pour nos simulations, nous nous sommes limités aux modèles de transport de la charge suivants : Dérive-diffusion et Hydrodynamique [Martinie 09]. Selon le modèle de transport de charge, on a une valeur de la 'Driving force' bien différente.

Lorsque l'on réduit la longueur de grille du dispositif, la variation du champ électrique augmente. Dans la théorie classique de transport par Dérive-diffusion (DD), la vitesse des porteurs augmente lorsque le champ électrique augmente jusqu'à saturation, car on suppose que la charge et le réseau cristallin sont toujours en équilibre thermodynamique. L'ionisation par impact a lieu dès que le champ électrique sera élevé dans la structure (de l'ordre 10^5 V.cm^{-1}), ce qui implique une surestimation du taux de génération par ionisation par impact (G_{ii}). Cependant, les théories de transport de la charge plus avancées telle que l'hydrodynamique (HD) prédisent un surpassement de la vitesse des porteurs avant retour à l'équilibre (figure 2.3) [Martinie 09]. Lorsque la vitesse des porteurs est supérieure à la vitesse de saturation, leurs énergies sont supérieures à la valeur de l'énergie des autres porteurs se trouvant dans la bande de conduction. Dans cette configuration, la température de la charge n'est plus égale à celle du réseau cristallin. Pour un retour à l'équilibre thermodynamique (à leur vitesse de saturation), le surplus d'énergie est cédé au réseau cristallin via les collisions qui peuvent donner lieu à une génération par ionisation par impact. L'intervalle qui sépare l'état hors équilibre (surpasse de la vitesse de saturation) à l'état d'équilibre (vitesse de saturation) est le temps de relaxation de l'énergie $\tau_{q,n}$ (avec n et p pour électrons et trous). Ce paramètre est fondamental et dépend du matériau semi-conducteur. La conséquence directe de ce comportement se répercute sur le taux de génération par ionisation par impact. L'ionisation par impact se déclenche si et seulement si la température des porteurs est supérieure

à celle du réseau cristallin. Cette condition ne peut être respectée que pour des fortes polarisations. C'est pourquoi le modèle Hydrodynamique sous-estime la valeur du taux de génération par ionisation par impact (G_{ii}).

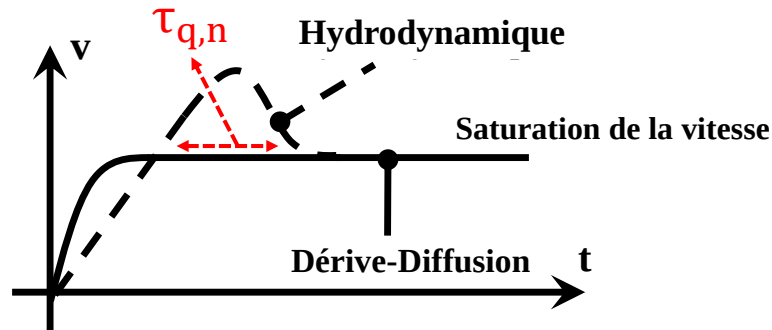


Figure 2.3 : Différence entre les modèles de transport de charge Drift-Diffusion et Hydrodynamique pour la réponse en vitesse à un échelon de champ électrique [Martinie 09].

Ainsi, dans un premier temps, nous allons considérer le modèle de transport de la charge par dérive-diffusion (*drift-diffusion*) combiné à la distribution de Fermi-Dirac, pour avoir une meilleure convergence et plus de robustesse pour nos simulations. Pour le mécanisme d'ionisation par impact, nous avons choisi celui de Van-Overstraeten [Overstraeten 70].

II.1.5 Analyse des opérations mémoires par simulations TCAD

Nous avons appliqué les chronogrammes des tensions, sur la structure de la figure 2.1.b. Sur la figure 2.4.a, nous reportons le courant du drain en fonction du temps pour la séquence d'opération mémoire (H-E-H-W-H-R-H-E-H-R-H) de la figure 2.2. Pour cette simulation, et dans la suite de nos études, nous allons activer seulement le mécanisme de génération de la charge par effet tunnel bande à bande.

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

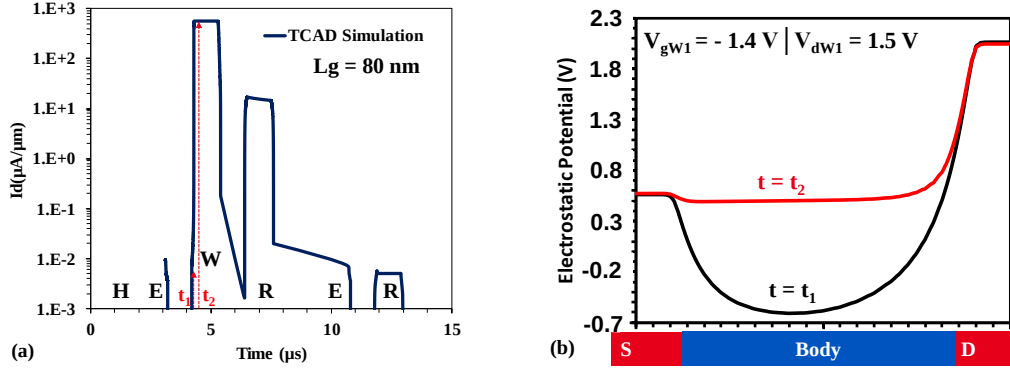


Figure 2.4 : (a) Variation du courant de drain de l'A2RAM (figure 2.1.b) en fonction du temps lorsque l'on applique sur la cellule le chronogramme de tensions de la figures 2.2. (b) Evolution du potentiel électrostatique de la phase 'écriture de l'état '1' à $t = t_1$ et $t = t_2$ comme indiqué sur la figure 2.4.a.

Durant la première phase d'effacement, le drain est à la masse, c'est pourquoi il n'y a pas de courant. Ensuite, nous avons la phase d'écriture (W) ; durant cette phase, la charge est générée par effet tunnel bande à bande, et stockée dans le body. Nous sommes sûrs d'avoir atteint la charge maximale pouvant être stockée dans le body parce que le courant d'écriture sature comme on peut le voir sur la figure 2.4.a (W). Pour confirmer cette observation, nous pouvons étudier l'évolution de la barrière de potentiel source/body durant l'écriture de l'état '1'. En effet, comme nous l'avons déjà mentionné, la charge maximale dans une 1T-DRAM est atteinte lorsqu'il y a autopolarisation du film de silicium (le body pour l'A2RAM). La figure 2.4.b montre une coupe horizontale du profil du potentiel électrostatique le long de l'ensemble source-body-drain (à 1 nm sous de l'oxyde de la grille face avant) aux deux instants t_1 et t_2 durant l'écriture de l'état '1'. A t_1 le stockage de la charge est encore à son début, la barrière de potentiel entre la source et le body est encore élevée. A t_2 , nous pouvons noter que la diode source-body passe en direct, ce qui marque l'auto-polarisation du body et par conséquent la charge maximale est atteinte.

Une fois que l'information est stockée, nous passons à la phase de lecture (R) où on observe un courant de lecture de l'état '1' $I_1 \approx 20 \mu A. \mu m^{-1}$. La figure 2.5.a démontre la présence de trous dans le body pendant la lecture de l'état '1', et la figure 2.5.b confirme que du courant circule uniquement dans le bridge pendant cette même opération.

Après la lecture de l'état '1', la charge dans le body est évacuée. Nous programmons l'état '0' toujours par couplage capacitif des grilles face avant et arrière. Durant la phase de lecture de l'état '0', le courant lu I_0 étant très faible ($5 nA. \mu m^{-1}$), nous pouvons dire qu'il n'y a plus de charge dans le body. Ceci est confirmé par la figure 2.5.c, c'est pourquoi durant la lecture de l'état '0' il n'y a pas de courant à travers le bridge (figure 2.5.d).

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

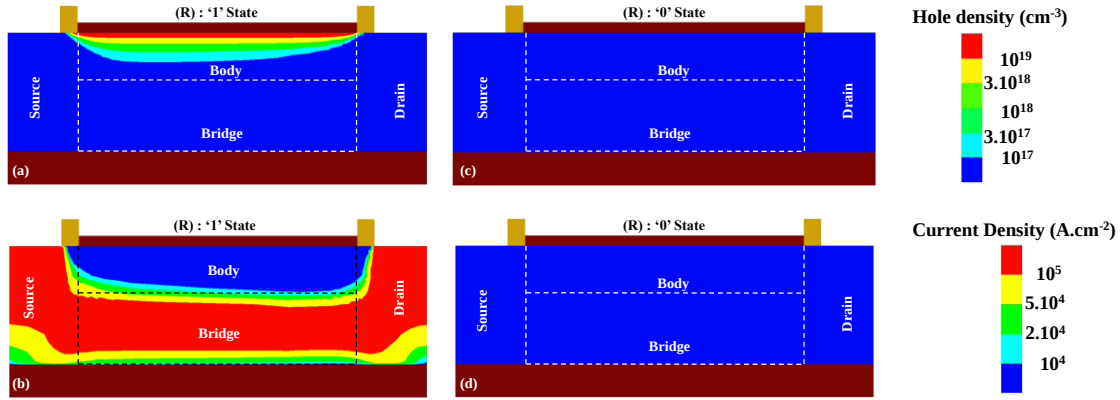


Figure 2.5 : (a, c) Profils de la densité de trous, et (b, d) Profils de la densité du courant extraits durant la phase de lecture des états '1' et '0'. Toutes les cartographies sont extraites des simulations TCAD.

Nous avons pu mettre en évidence la présence d'un effet mémoire ($I_1 \gg I_0$) dans la structure A2RAM simulée. Nous allons par la suite valider notre méthodologie de simulation en montrant que les résultats obtenus en TCAD reflètent le comportement attendu d'une structure A2RAM réelle.

II.1.6 Comparaisons entre résultats de simulation TCAD et caractérisation électrique

La caractérisation (mémoire, statique ou petits signaux) du point mémoire A2RAM durant ce travail de thèse a été effectuée sur la station sous pointe semi-automatique Elite 300 de l'IMEP-LAHC. C'est un équipement indispensable, car il garantit une rapidité d'acquisition et une précision de données de mesure incomparable. Couplé à un analyseur de paramètres de dispositifs à semi-conducteur (l'analyseur B1500A), il est facile de mesurer des amplitudes de courant inférieur à 1 fA et des capacités dont les ordres de grandeur sont autour de quelques femtoFarad (fF).

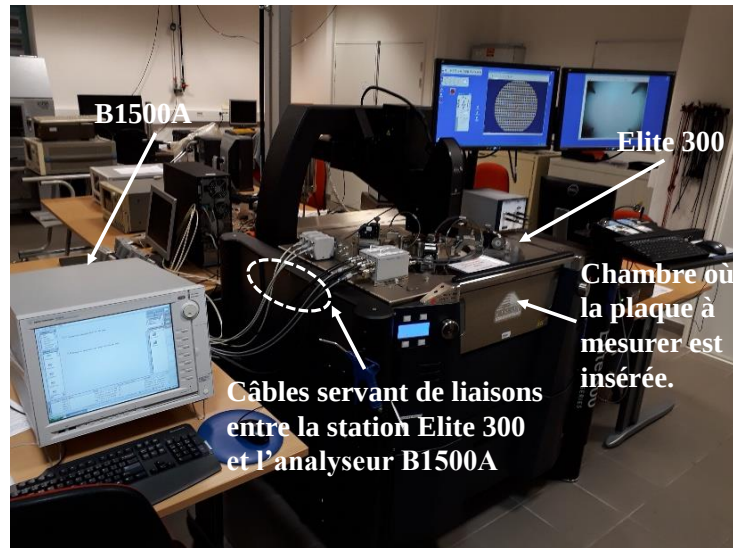


Figure 2.6 : Environnement utilisé pour la caractérisation du point mémoire A2RAM.

La figure 2.6 présente l'environnement complet, où nous pouvons distinguer la station Elite 300 et l'analyseur B1500A. Pour mesurer le dispositif, nous devons l'introduire dans la station Elite 300 (figure 2.7.a). Une fois la plaque à l'intérieur de la station semi-automatique, il s'ensuit différentes étapes pour bien la centrer et l'aligner sur le chuck. La station Elite 300 nous permet aussi de faire une cartographie des puces à mesurer. Avant de déposer les pointes sur le dispositif, il y'a plusieurs étapes à suivre.

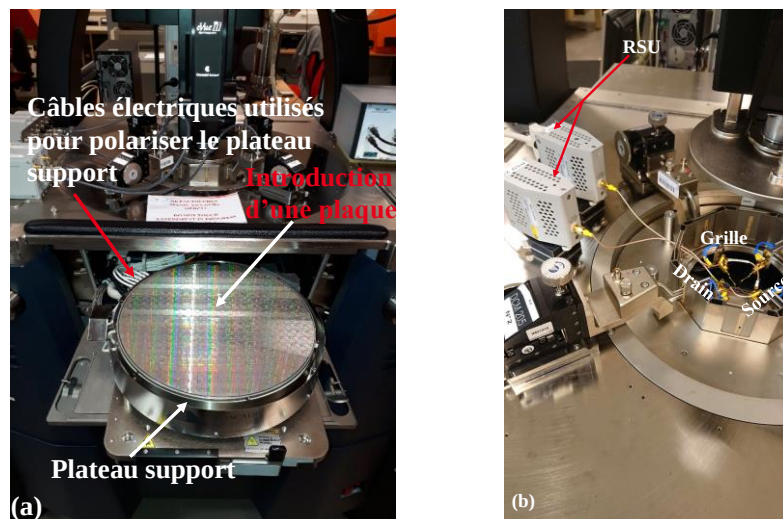


Figure 2.7 : (a) Introduction d'une plaque dans la station Elite 300. (b) Connection Drain, Grille (grille face avant), et Source ; la grille face arrière est polarisée via le plateau support.

- La première étape très importante pour une mesure électrique, c'est le câblage. Il varie selon la mesure que l'on veut effectuer. Pour la caractérisation mémoire, étant

donné que les variations du signal sont rapides, le câblage est particulier ; de plus, durant ce travail de thèse, nous n'avions à disposition que deux RSU (Remote-Sense and Switch Unit). Le RSU est un boîtier adapté pour des mesures pulsées faisant la liaison entre l'analyseur B1500A et le dispositif à mesurer (DUT : Device Under Test). Il fallait donc maintenir la source, le drain et la grille à la même référence sachant que la grille face arrière est maintenue à la masse en appliquant un signal sur le chuck. Le câblage dans notre cas est montré sur la figure 2.7.b.

- Après le câblage, l'étape suivante est le calibrage de la mesure. Cette étape se fait via l'analyseur B1500A et est aussi fonction de la mesure à effectuer. Elle consiste à annuler de nos mesures l'offset dû au câble électrique sur la valeur du courant ou de la capacité.

Une fois le câblage et la calibration effectués, nous pouvons poser les pointes sur le dispositif à mesurer. La mesure est définie dans un environnement dédié de l'analyseur B1500A, il faut entrer les différents chronogrammes de tensions (dans notre cas, c'est celui indiqué sur la figure 2.2). La précaution particulière dont il faut tenir compte pour la caractérisation mémoire est le temps de chargement et déchargement des capacités parasites des fils faisant les liaisons entre l'analyseur B1500A et le dispositif mesuré DUT (figure 2.5) en passant par les RSU. C'est pourquoi, pour garantir une bonne résolution sur les valeurs des courants lus par l'instrument de mesure, nous avons pris une durée de chaque impulsion de 1 μ s et un temps de montée et de descente de 100 ns.

Pour valider notre stratégie de simulation TCAD, nous avons comparé nos résultats obtenus en simulation avec les mesures électriques sur des points mémoires A2RAM précédemment fabriqués au Leti [Rodriguez 12]. Par ailleurs, les échantillons que nous avons caractérisés ont un plan de masse en silicium (sous le BOX) dopé de type p à 10^{18} cm⁻³. C'est pourquoi nous avons refait la même simulation de la section 1.5, en ajoutant sous la structure de la figure 2.1.b un plan de masse d'épaisseur 1 μ m et dopé de type p à 10^{18} cm⁻³ (figure 2.8.a). Néanmoins, nous n'avons pas connaissance des profils de dopage réel de la source, du drain, du substrat et surtout du bridge. C'est pourquoi nous n'avons pas calibré les modèles de génération de la charge. Le but ici n'est pas d'avoir des résultats de simulations identiques à celles de nos échantillons, mais de vérifier plutôt si nos résultats de simulations suivent les tendances de la mesure électrique. La comparaison entre les résultats de simulations TCAD et la caractérisation électrique est illustrée sur la figure 2.8.b.

Nous observons que les courants lus des états '1' et '0', I1 et I0 sont différents de ceux de la mesure électrique. La différence observée sur I1 ne peut être attribuée à la quantité de charge stockée durant l'écriture de l'état '1', car bien que les amplitudes des courants en écriture soient différentes, elles atteignent toutes la saturation. L'amplitude du courant I0 est sous-estimée en simulation comme pour le courant I1. Ceci confirme effectivement que les différences observées ne sont pas liées à la quantité de charge stockée dans la structure simulée, mais que les structures comparées sont différentes.

De nos travaux, nous avons noté que la variation des profils de dopage dans la profondeur du film de silicium (body + bridge) avait un impact important sur les performances mémoires de l'A2RAM. En appliquant des variations de la valeur du dopage du bridge (N_{bridge}), nous nous approchons de plus en plus des données expérimentales. Sur la figure 2.8.d, nous comparons

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

une fois de plus la caractéristique électrique mémoire avec une structure A2RAM comme celle de la figure 2.8.a, mais avec un dopage de bridge de $1.3 \times 10^{18} \text{ cm}^{-3}$ (figure 2.8.c).

La variation du courant d'écriture dans ce cas suit les tendances de celle de la mesure électrique, mais les amplitudes sont différentes. Cette différence est due au fait que les modèles physiques inclus dans nos simulations ne sont pas calibrés sur l'échantillon mesuré. La valeur du courant I1 est quasiment identique à celle de l'échantillon mesuré. En ce qui concerne la valeur du courant I0, comparée à la figure 2.8.b, nous observons une amélioration, mais elle reste faible par rapport à la mesure électrique.

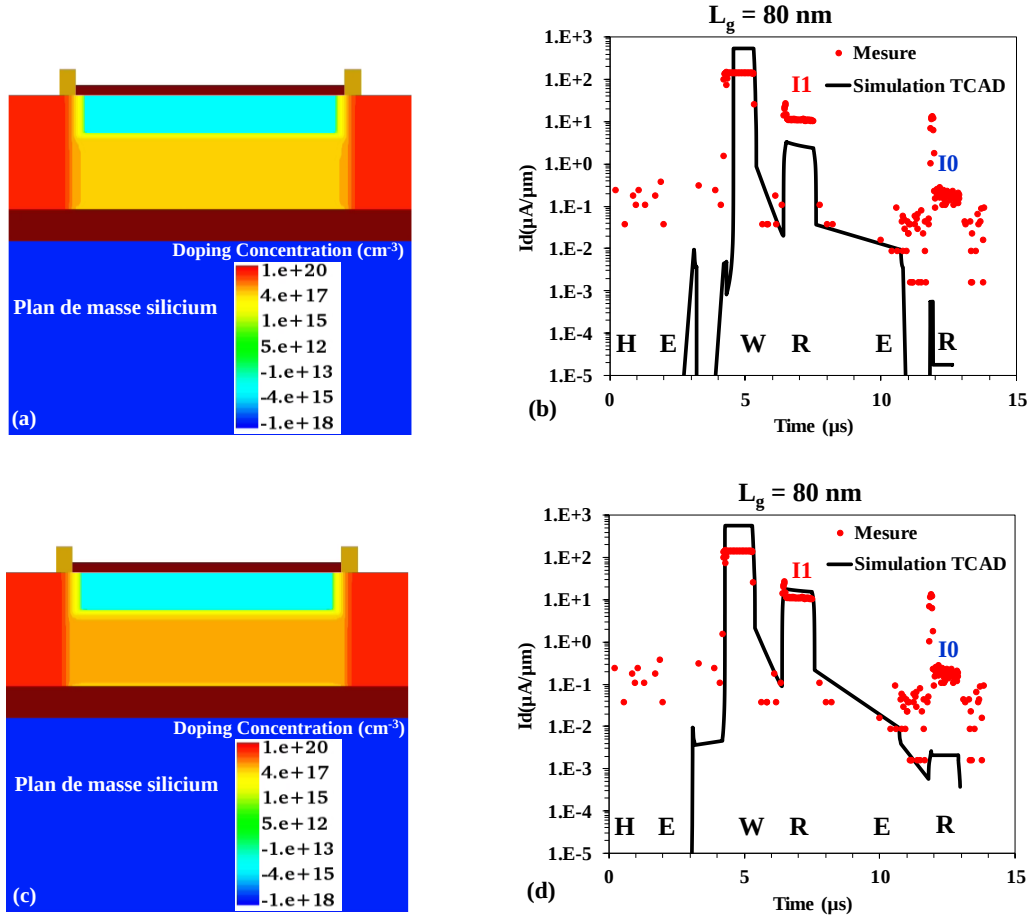


Figure 2.8 : Structures A2RAM simulées avec plan de masse silicium et comparaison de la variation du courant de drain obtenus par TCAD et par mesure pour la séquence effacement – programmation – lecture – effacement – lecture ((a,b) $N_{\text{bridge}} = 1.3 \times 10^{18} \text{ cm}^{-3}$ et (c,d) $N_{\text{bridge}} = 10^{18} \text{ cm}^{-3}$).

D'après les comparaisons des figures 2.8.b et 2.8.d, nous avons conclu que la connaissance du profil de dopage le long du film de silicium (body + bridge) est un élément fondamental pour mieux évaluer la performance mémoire de l'A2RAM. C'est pourquoi dans le Chapitre 3, nous étudions sur la sensibilité des performances mémoires de l'A2RAM en fonction de son profil

de dopage. Par la suite, dans le Chapitre 4, nous allons proposer une méthodologie d'extraction du profil de dopage dans une structure A2RAM.

II.2. Analyse de la sensibilité des performances mémoires : variation des polarisations

L'objectif de cette partie est d'étudier l'impact des polarisations appliquées pendant chacune des opérations mémoires une par une sur les performances mémoires de l'A2RAM. Pour chaque variante, nous allons extraire la valeur de I1 et I0, et étudier leurs comportements. Nous allons utiliser les mêmes chronogrammes de tensions présentés sur la figure 2.2, mais avec une durée d'impulsion de 10 ns et un front montant et descendant de 1 ns. L'intérêt d'utiliser ces temps courts est de se rapprocher de ceux d'une cellule mémoire dynamique standard 1T-1C [G. Fredeman 16]. Comme précédemment nous choisissons le mécanisme de génération de la charge par le courant de fuite par GIDL, c'est pourquoi dans les simulations nous maintenons seulement activé le mécanisme de génération de la charge par effet tunnel bande à bande.

II.2.1 Variation des tensions de grille face avant et de drain pendant la programmation : V_{gW1} et V_{dW1}

Le but de cette partie est de trouver un couple de tensions à appliquer sur la grille face avant V_{gW1} et sur le drain V_{dW1} pour garantir la valeur maximale du courant de l'état '1' lu I1 avec un temps d'écriture de 10 ns.

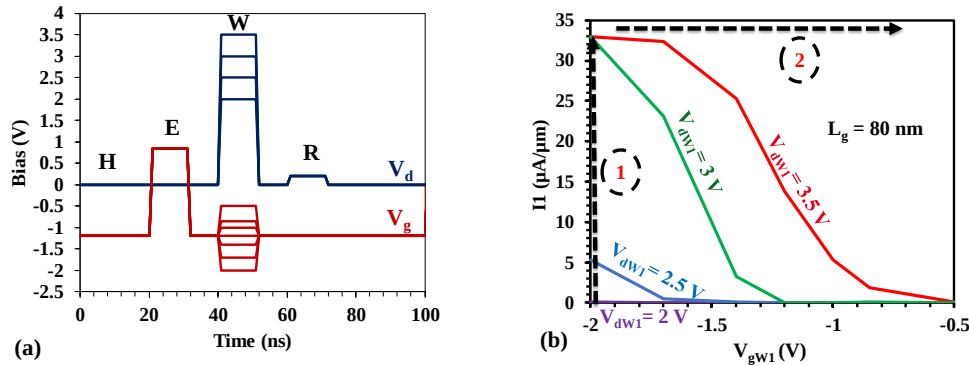


Figure 2.9 : (a) V_d et V_g en fonction du temps pour la séquence E-W-R avec des variations de tension appliquées pendant la programmation (V_{gW1} , V_{dW1}) avec une largeur d'impulsion de 10 ns et un front montant / descendant de 1 ns. (b) Courant de lecture de l'état I1 en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) extrait des simulations TCAD.

Sur la figure 2.9.a, nous avons le chronogramme de tensions utilisé pour la caractérisation de l'A2RAM (figure 2.1.b). Sur la figure 2.9.b, nous reportons la variation du courant de l'état '1' I1 en fonction des couples de tensions (V_{gW1} , V_{dW1}).

Dans un premier temps, nous allons fixer la valeur de la tension de la grille face avant d'écriture V_{gW1} à -2 V, et faire varier la tension de drain d'écriture V_{dW1} comme indiqué sur la

figure 2.9.b (flèche no.1). Nous constatons que I_1 augmente lorsque V_{dW1} augmente, ce qui veut dire qu'on stocke plus de charges lorsque la tension de drain d'écriture V_{dW1} est plus élevée. En effet, le champ électrique dans la zone de recouvrement drain/oxyde de grille face avant nécessaire pour avoir un effet tunnel bande à bande est proportionnel à $V_{dW1} - V_{gW1}$, d'où l'amélioration de la programmation.

Cette observation est confirmée par les cartographies de la figure 2.10 qui montrent le taux de génération par effet tunnel bande à bande lorsque $V_{gW1} = -2$ V et V_{dW1} varie entre 2 V à 3.5 V, extraites de TCAD en début de l'écriture lorsque V_g passe à V_{gW1} et V_d à V_{dW1} . La génération est plus forte lorsque le couple (V_{gW1}, V_{dW1}) est élevé.

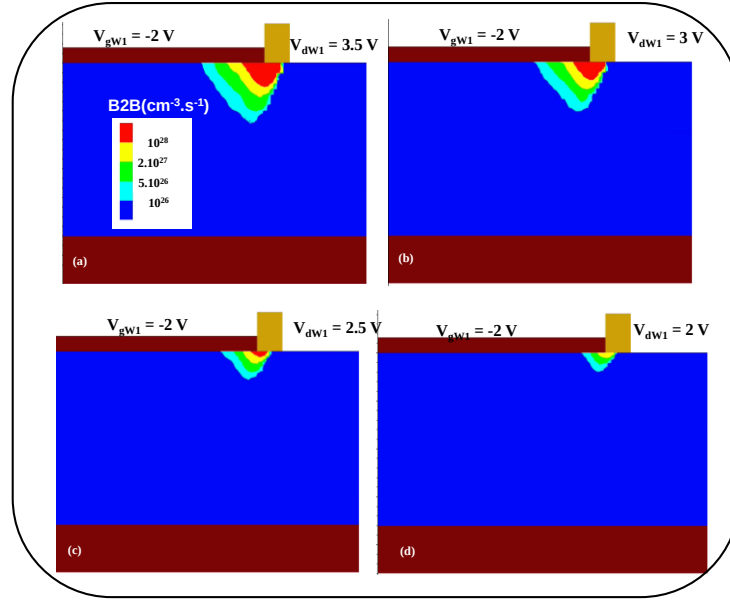


Figure 2.10 : Profils du taux de génération B2B extraits en début de la phase d'écriture de l'état '1' pour différentes valeurs de tensions de drain (V_{dW1})

D'autre part, pour s'assurer d'atteindre la quantité de charge maximale pouvant être stockée, nous reportons sur la figure 2.11.a les chronogrammes de courants extraits des simulations TCAD lorsqu'on applique le motif de la figure 2.9.a pour $V_{gW1} = -2$ V. Durant la phase d'écriture (W), nous constatons effectivement que le courant sature largement pour $V_{dW1} = 3.5$ V, et quasiment pour $V_{dW1} = 3$ V, d'où le courant I_1 élevé et pratiquement égal dans ces 2 cas sur la figure 2.9.b. Pour $V_{dW1} < 3$ V, le courant d'écriture reste faible, d'où le courant I_1 de modeste intensité sur la figure 2.9.b. Ainsi, pour des polarisations de drain $V_{dW1} < 3$ V, pour garantir l'écriture, nous devons augmenter le temps d'écriture. Pour le confirmer, nous avons refait la même simulation que celle de la figure 2.9.a, mais avec une durée de la phase d'écriture de 300 ns, et pour $V_{dW1} = 2.5$ V et $V_{dW1} = 3.5$ V. Sur la figure 2.11.b, nous constatons que I_1 est le même pour $V_{dW1} = 2.5$ V et $V_{dW1} = 3.5$ V lorsque le temps de programmation de l'état '1' (W) est égal à 300 ns.

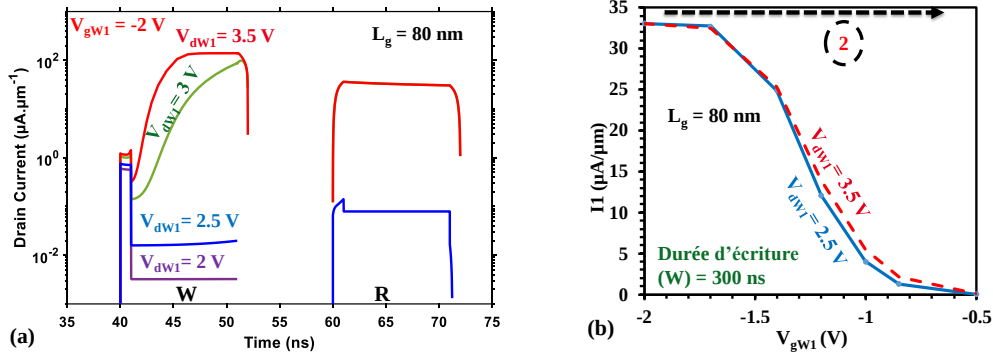


Figure 2.11 : (a) Chronogramme de courant durant les phases d'écriture (W) et de lecture (R) lorsqu'on applique le motif de tension de la figure 2.9. (b) Courant de lecture de l'état I1 en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) extrait des simulations TCAD lorsqu'on applique le motif de la figure 2.9.a mais avec une phase d'écriture (W) qui dure 300 ns.

Cependant, si nous fixons la valeur de $V_{dW1} = 3.5 \text{ V}$ et varions la tension de la grille face avant en écriture V_{gW1} de -2 V à -0.5 V, sur la figure 2.9.b (flèche no.2), l'amplitude du courant I1 diminue. Nous devons vérifier s'il y a une charge suffisante stockée dans le body. Dans un premier temps, nous avons extrait les chronogrammes des courants durant les simulations de la figure 2.9, pour $V_{dW1} = 3.5 \text{ V}$ et V_{gW1} variant de -2 V à -0.5 V comme reporté sur la figure 2.12.a. Nous constatons que les courants en écriture saturent pour toutes les valeurs de V_{gW1} , par conséquent on peut dire que la charge maximum dans le body a été atteinte. De plus, la réduction de I1 ne peut être attribuée au temps de stockage car, sur la figure 2.11.b (flèche no.2), où le temps d'écriture est de 300 ns, nous avons les mêmes tendances de la figure 2.9.b. Par conséquent, la réduction de I1 serait plutôt due à une diminution de la quantité de trous pouvant être stockée dans le body. En effet, lorsqu'on réduit V_{gW1} , la barrière de potentiel source-body est réduite. Cette observation est confirmée par les courbes du potentiel électrostatique le long de la source-body-drain en début de l'écriture et à la fin de l'écriture (figure 2.12.b). Plus la barrière de potentiel est faible, plus le réservoir de trous est réduit, et moins on stocke de la charge dans le body, d'où la réduction de I1 sur les figures 2.9.b et 2.11.b lorsque V_{gW1} augmente.

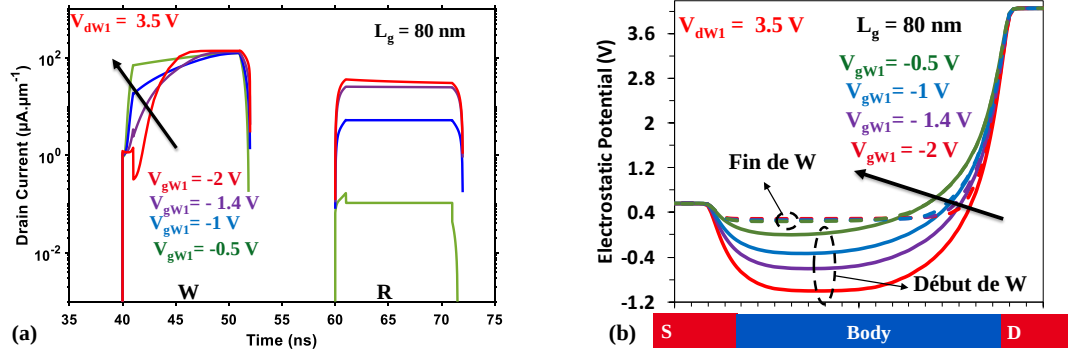


Figure 2.12 : (a) Chronogramme de courant durant les phases d'écriture (W) et de lecture (R) lorsque l'on applique le motif de tension de la figure 2.9.a sur la structure 2.1.b. (b) Evolution du potentiel électrostatique le long des jonctions source-Body-drain durant la phase d'écriture de l'état '1'.

En conclusion, le stockage de la charge par effet tunnel bande-à-bande est garanti pour un temps d'écriture de 10 ns lorsque $|V_{dW1} - V_{gW1}| \gg 0 \text{ V}$. Dans cette étude, nous avons : $V_{dW1} \geq 3 \text{ V}$ et $V_{gW1} \leq -1.7 \text{ V}$.

II.2.2 Variation de la tension de grille face avant pendant la phase de maintien : $V_{g\text{Hold}}$

Dans cette partie, nous allons faire varier la tension de maintien de la grille face avant $V_{g\text{Hold}}$ et étudier les variations de I_1 et I_0 . La figure 2.13.a montre le chronogramme de tension que nous allons utiliser pour cette étude.

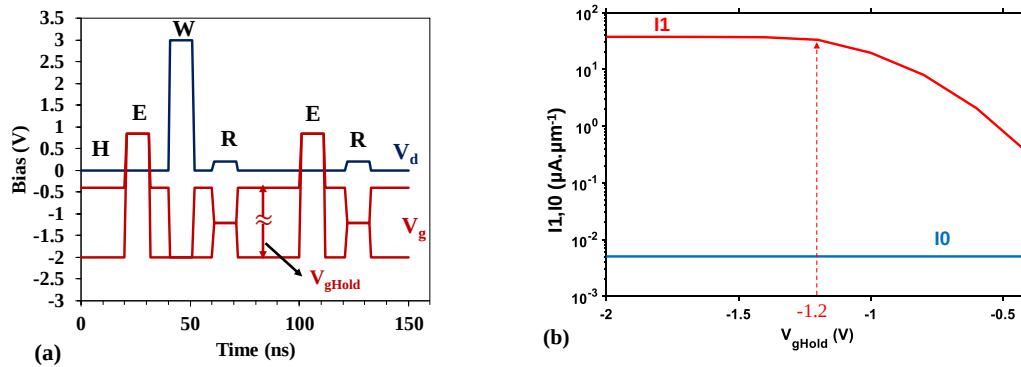


Figure 2.13 : (a) Chronogramme de tensions appliqué sur le drain V_d et la grille face avant V_g ; la tension de maintien de la grille $V_{g\text{Hold}}$ varie de -2 à -0.4 . (b) Courants lus des états '1' (I_1) et '0' (I_0), en fonction de la tension de maintien $V_{g\text{Hold}}$ extraits des simulations TCAD.

Sur la figure 2.13.b, nous reportons les variations de I_1 en fonction de $V_{g\text{Hold}}$; le premier constat est que la variation de $V_{g\text{Hold}}$ n'a pas d'impact sur I_0 . Mais, I_1 diminue lorsque $V_{g\text{Hold}} > -1.2 \text{ V}$; cette réduction ne peut être attribuée à la phase d'écriture ou de lecture de

l'information car, les conditions d'écriture et de lecture sont les mêmes dans tous les cas (figure 2.13.a). Pour expliquer cette différence, nous avons extrait les coupes verticales de la densité de trous pour 3 valeurs de V_{gHold} (figure 2.14) durant la phase de lecture de l'état '1'. On constate que la quantité de charge dans le body diminue lorsque V_{gHold} augmente. En effet, après la phase d'écriture nous avons une phase de maintien intermédiaire, et c'est durant cette phase que la charge varie. Lorsque l'on augmente V_{gHold} , on s'éloigne du régime d'accumulation dans le body, ceci se traduit par une réduction des barrières de potentiels source-body et drain-body, d'où la diminution de la charge stockée dans le body. Ainsi, durant la lecture, le champ électrique vertical induit par V_{gRead} est moins bien écranté lorsqu'on a moins de charge dans le body, d'où la réduction de I_1 .

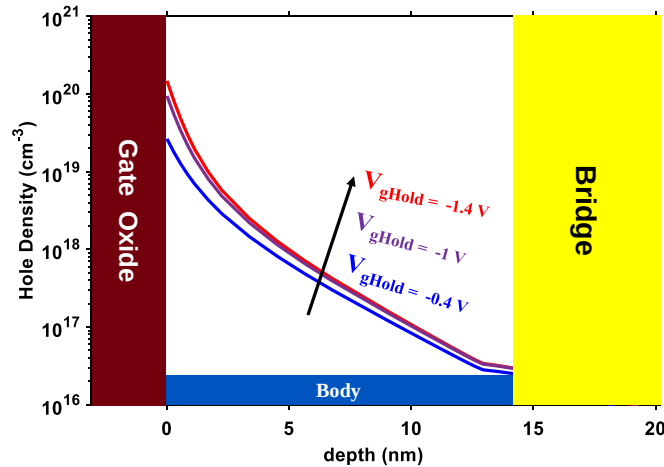


Figure 2.14 : Evolution de la densité de trous dans la profondeur du body pour différentes polarisations V_{gHold} , extraites durant la lecture de l'état '1'.

En conclusion, cette étude nous permet de mettre en évidence l'importance d'être en régime d'accumulation dans le body pour pouvoir mieux retenir la charge stockée après l'écriture. Cette condition est respectée lorsque $V_{gHold} \leq V_{FB}$. Les variations de la figure 2.13 nous devons imposer $V_{gHold} \leq -1.2$ V.

II.2.3 Variation de la tension de grille face avant pendant la lecture :

$$V_{gRead}$$

Nous allons à présent faire la même étude que précédemment mais en faisant varier V_{gRead} . Nous considérons le chronogramme des tensions de la figure 2.15.a, où nous fixons $V_{gHold} = -1.2$ V et faisons varier V_{gRead} de -2 V à 0 V. Pour chaque valeur de polarisation de lecture V_{gRead} , nous extrayons les valeurs des courants lus de l'état '1' (I_1 , figure 2.15.b) et l'état '0' (I_0 , figure 2.15.c).

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

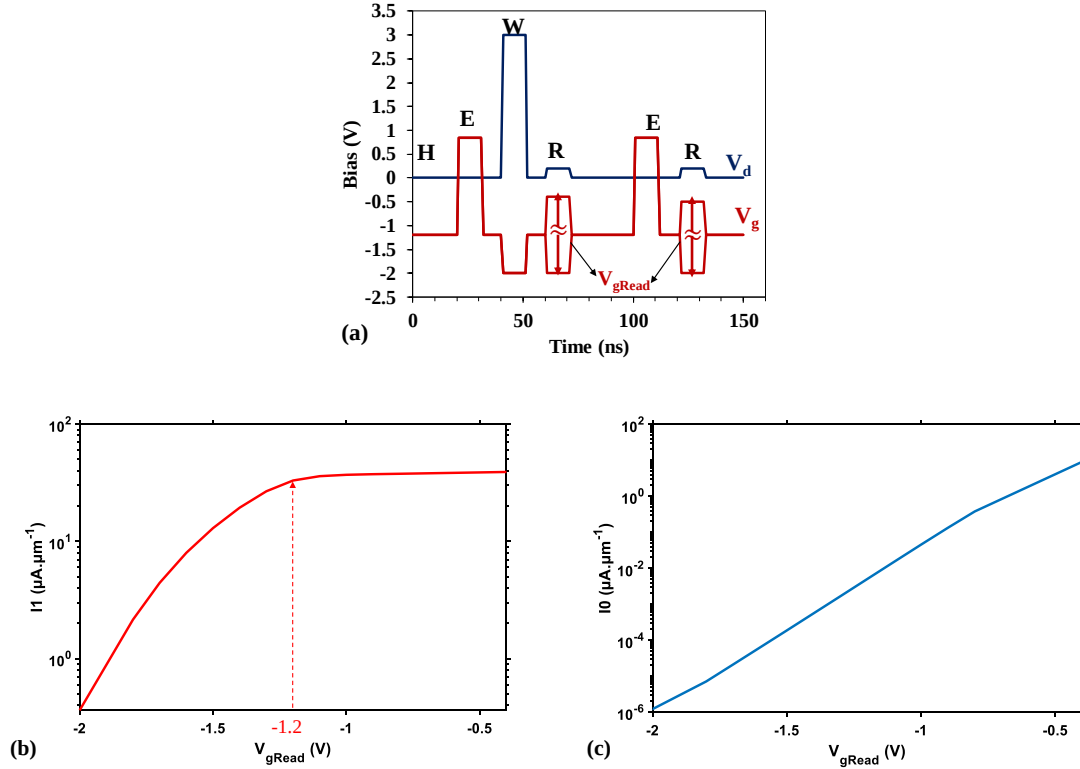


Figure 2.15 : (a) Chronogramme de tensions appliquées sur le drain V_d et la grille face avant V_g ; la tension de lecture de la grille V_{gRead} varie de -2 V à -0.4 V (figure à refaire). Courant lu en fonction de la tension de lecture V_{gRead} extrait des simulations TCAD pour (b) l'état '1' ($I1$) et (c) l'état '0' ($I0$).

Lorsque $V_{gRead} > -1.2$ V, le courant $I1$ reste constant, mais il décroît pour des valeurs de tension de grille de lecture V_{gRead} inférieures à -1.2 V. La figure 2.16 montre le profil de la densité des trous stockés dans la profondeur du body. Leur quantité reste quasiment constante, et lorsque le champ électrique vertical dû par la grille augmente ($V_{gRead} < -1.2$ V) l'écrantage est moins efficace. Dans cette condition, le bridge conduit moins le courant car, il est déserté en partie par ce champ électrique. Ceci est confirmé par les cartographies de la densité de courant durant la lecture de l'état '1' sur la figure 2.17. Nous notons effectivement une réduction de la densité de courant lorsque V_{gRead} tends vers des valeurs très négatives, le champ électrique vertical déserte de plus en plus le bridge et donc moins de courant y circule.

Sur la figure 2.15.c, nous avons la variation de $I0$ en fonction de V_{gRead} . Lorsque V_{gRead} est plus négative le courant de fuite des diodes parasites diminue, donc $I0$ diminue. En revanche, lorsque V_{gRead} augmente (c'est-à-dire tends vers des valeurs moins négatives), en plus de l'augmentation du courant de fuite des diode source-body et drain-body, le champ électrique vertical dû à la grille face avant a de moins en moins d'effet sur le bridge. Ainsi, le bridge est de moins en moins déserté et le courant $I0$ augmente et tends à converger vers la valeur de $I1$.

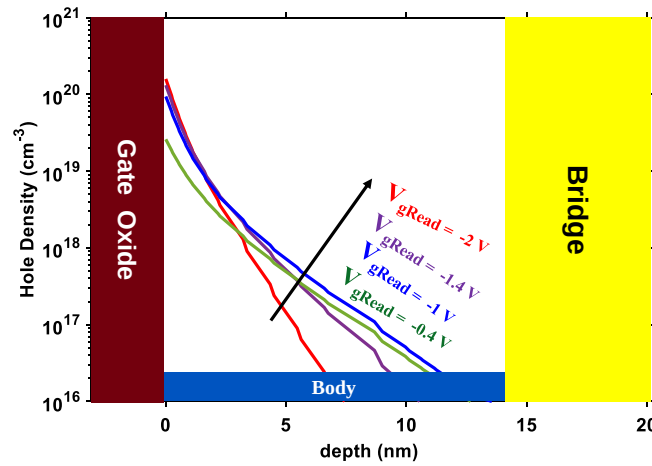


Figure 2.16 : Evolution de la densité de trous dans la profondeur du body pour différentes polarisations V_{gRead} extraites durant la lecture de l'état '1'.

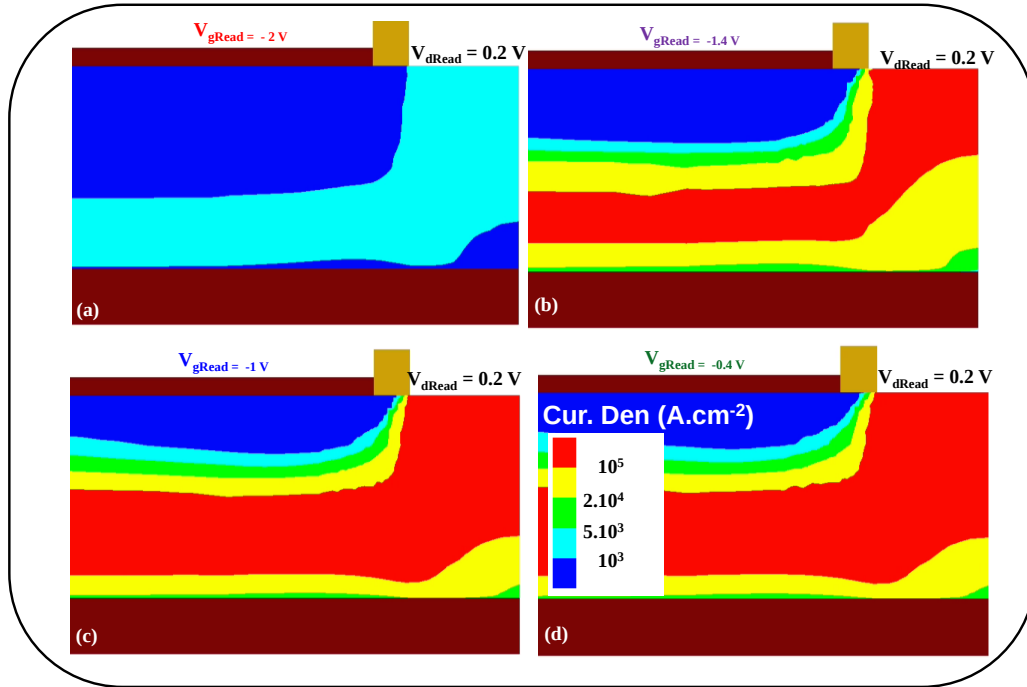


Figure 2.17 : Cartographie de la densité de courant durant la lecture de l'état '1' pour les tensions de grille de lecture de la figure 2.16.

En conclusion, le choix de la valeur de V_{gRead} impose un compromis entre la valeur de I_1 et I_0 ; si V_{gRead} est très négatif, I_1 est réduit car on a moins de trous pour écranter le champ électrique vertical induit par la grille. Si V_{gRead} se rapproche plus des valeurs positives, I_0 augmente à cause des fuites des diodes. Pour notre étude des figures 2.15.b et 2.15.c, le bon compromis est obtenu pour $V_{gRead} = -1.2$ V.

II.2.4 Variation des tensions de grille face avant pendant la lecture et la phase de maintien avec $V_{gRead}=V_{gHold}$

Pour cette étude, nous considérons que $V_{gRead} = V_{gHold}$ (figure 2.18.a) et nous allons étudier l'impact de cette polarisation sur $I1$ (figure 2.18.b) et $I0$ (figure 2.19.a).

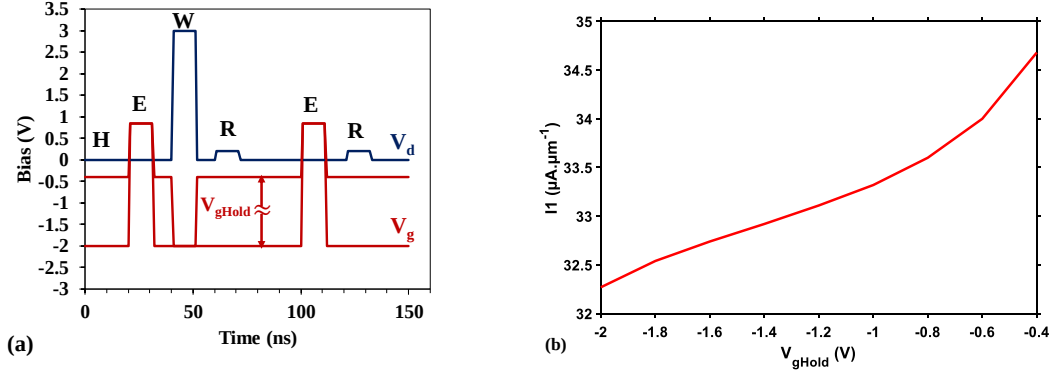


Figure 2.18 : (a) Chronogramme de tensions appliquées sur le drain V_d et la grille face avant V_g ; la tension de maintien de la grille V_{gHold} varie de -2 V à -0.4 V. (b) Courant lu de l'état '1' ($I1$) en fonction de la tension de maintien V_{gHold} extraits des simulation TCAD.

$I1$ reste quasiment constant quand V_{gHold} varie, ceci est dû au fait que nous combinons les deux effets avantageux observés sur les figures 2.13.b pour $V_{gHold} \leq -1.2$ V et 2.15.b pour $V_{gRead} > -1.2$ V.

Les variations de $I0$ sont identiques aux variations observées sur la figure 2.15.c. En effet, $I0$ augmente avec V_{gHold} parce que le champ électrique vertical dû à la grille diminue. Le bridge sera de moins en moins déserté, et il en suivra alors une conduction du courant dans le bridge. La figure 2.19.b montre le profil de la densité d'électrons dans la profondeur du film de silicium (body + bridge) à trois différentes valeurs de V_{gHold} . La densité d'électrons dans le bridge augmente lorsque V_{gHold} augmente, ce qui confirme une conduction dans le bridge pendant la lecture de l'état '0'. Elle est d'autant plus importante lorsque la tension de maintien de grille est faible (en valeur absolue).

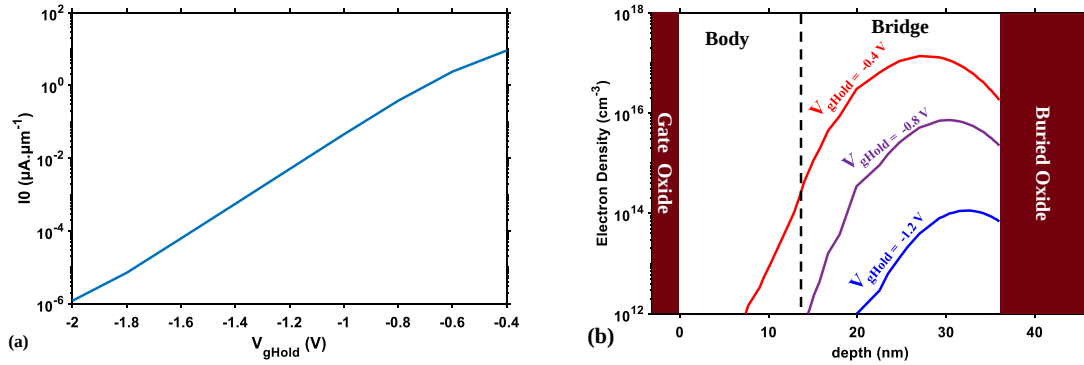


Figure 2.19 : (a) Courant lu de l'état '0' (I_0) en fonction de la tension de maintien V_{gHold} extrait des simulation TCAD. (b) évolutions de la densité d'électrons dans la profondeur du film de silicium (body + bridge) pour différente polarisation V_{gHold} extraites durant la lecture de l'état '0'.

De plus, si $V_{gRead} = V_{gHold}$ comme c'est notre cas, il est important d'étudier l'effet de cette tension sur le temps de rétention. Pour l'A2RAM, le pire des cas se produit durant une lecture prolongée de l'information que nous avons stockée [F.W. Tchene 17] : causées en partie par une génération parasite de la charge par GIDL durant une lecture prolongée de l'état '0', et pour l'état '1' par une fuite de la charge stockée via les diodes source/film de silicium et/ou drain/film de silicium. Les motifs de tensions pour évaluer le temps de rétention sont présentés sur les figures 2.20.a et 2.20.b pour l'état '1' et l'état '0' respectivement. Pour l'évaluation du temps de rétention de l'état '1', comme indiqué sur la figure 2.20.a, nous commençons par l'initialisation de l'état mémoire (E). Puis nous passons à la programmation classique de l'état '1', ensuite nous faisons une lecture continue de l'état '1' pendant quelques secondes. Sur la figure 2.20.b, nous programmons l'état '0' suivi également d'une lecture continue. Les deux chronogrammes de tensions des figures 2.20.a et 2.20.b sont appliqués sur la structure A2RAM de la figure 2.1.b. Dans les deux cas nous avons activé le mécanisme du courant de fuite de la grille (DirectTunneling section 1.4). Le courant de fuite de la grille étant un effet tunnel via l'oxyde de grille, il est important de tenir compte de l'empilement de l'oxyde de grille complet, c'est-à-dire dioxyde de silicium (SiO_2) + le dioxyde d'hafnium (HfO_2) [Navarro 14]. L'épaisseur de la couche de HfO_2 et sa permittivité ont été choisies de façon à avoir un EOT = 3.1 nm. Cette méthodologie de simulation du temps de rétention sera utilisée tout au long de ce travail de thèse.

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

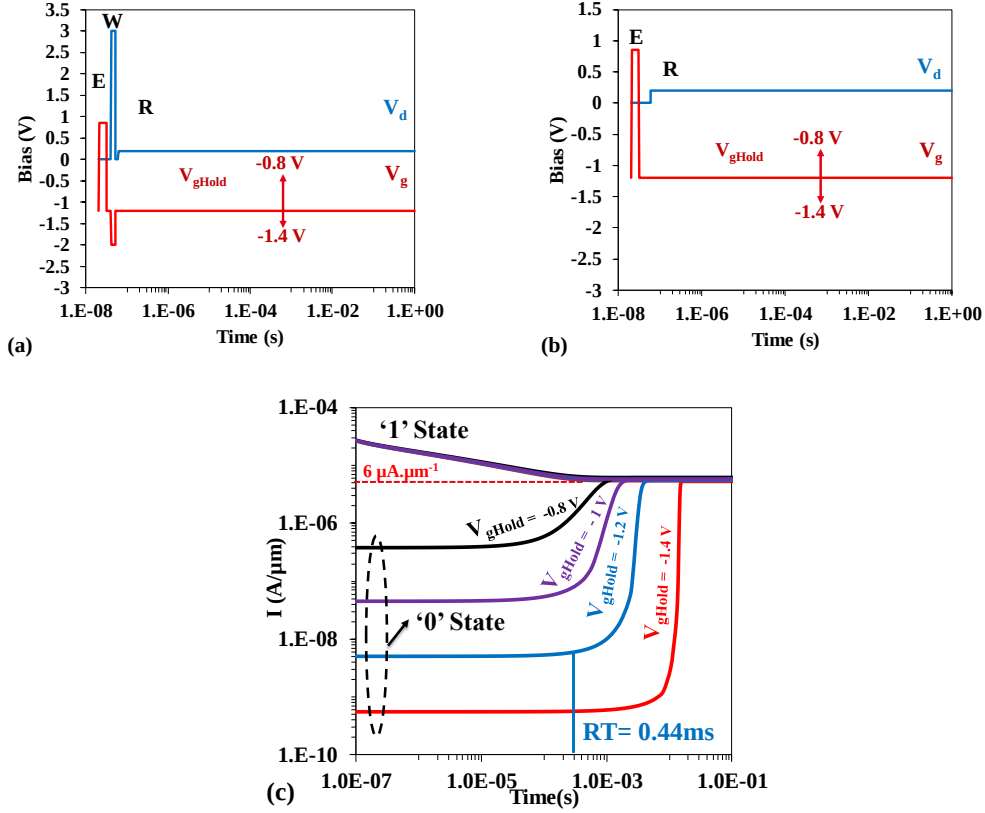


Figure 2.20 : Chronogramme de tensions appliquées sur le drain V_d et sur la grille face avant V_g pour une lecture continue des états (a) '1' (I_1) et (b) '0' (I_0) à différentes valeurs de la tension de maintien V_{gHold} (donc de lecture). (c) Evolution des courants lus I_1 et I_0 pour l'évaluation du temps de rétention extraite des simulations TCAD. La valeur minimale de I_1 ($6 \mu A \cdot \mu m^{-1}$) garantissant une fonctionnalité matricielle est indiquée.

La figure 2.20.c montre l'évolution des courants de drain pendant une lecture continue des états '1' et '0' à différentes valeurs de V_{gHold} . Notons que la valeur de I_1 diminue mais atteint rapidement sa valeur d'équilibre thermodynamique ($\approx 6 \mu A \cdot \mu m^{-1}$) quel que soit la valeur de la tension de maintien de la grille face avant V_{gHold} . L'évolution du courant lu de l'état '0' dépend fortement de la tension de maintien de la grille face avant, comme observé sur la figure 2.20.c. Elle reste constante pendant un certain temps, et augmente brusquement pour converger à la valeur de I_1 . On constate sur la figure 2.20.c que les courants saturent tous à la même valeur, donc le temps de rétention est limité par I_0 . Le temps de rétention lorsque $V_{gHold} = -1.2 V$ vaut 0.44 ms et va diminuer si V_{gHold} augmente.

En conclusion, nous avons observé que le choix de V_{gHold} est guidé par le compromis I_1 et le temps de rétention. L'évaluation du temps de rétention suggère que lorsque $V_{gHold} = V_{gRead}$, il faut que la tension de maintien soit inférieure ou égale à -1.2 V. Par ailleurs, de nos études des paragraphes 2.2 et 2.3, nous avons fait le choix de prendre $V_{gHold} = V_{gRead} = -1.2 V$. cette condition permet aussi de faciliter le travail du concepteur de circuit qui n'aura pas besoin d'utiliser plusieurs tensions d'alimentation pour le phase de maintien et de lecture.

II.2.5 Variation de la tension de drain pendant la lecture : V_{dRead}

Nous allons faire une étude similaire au cas précédent. Ici, nous fixons toutes les autres polarisations et faisons varier la tension de drain de lecture le V_{dRead} (figure 2.21.a).

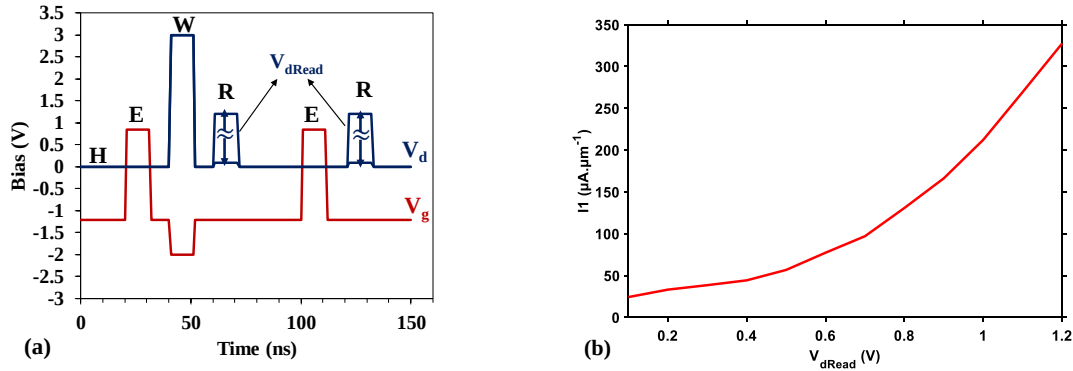


Figure 2.21 : (a) Chronogramme de tensions appliquées sur le drain V_d et la grille face avant V_g ; la tension de lecture du drain V_{dRead} varie de 0.1 V à 1.2 V. (b) Courant lu de l'état '1' (I_1) en fonction de la tension de lecture du drain V_{dRead} extraits des simulation TCAD.

Sur la figure 2.21.b, nous présentons la variation de I_1 en fonction de V_{dRead} . Pour $V_{dRead} \leq 0.5$ V, I_1 augmente linéairement avec V_{dRead} : durant la lecture, le courant circule seulement dans le bridge et le transport y est ohmique, avec une résistance de bridge constante, d'où la variation linéaire de I_1 observée sur la figure 2.21.b. Lorsque $V_{dRead} > 0.5$ V, le courant de fuite par GIDL commence à être prépondérant, d'où l'augmentation quasi exponentielle du courant I_1 . Cette observation est justifiée par les variations de I_0 en fonction de V_{dRead} de la figure 2.22.a.

I_0 augmente avec V_{dRead} , en raison du courant de fuite par GIDL qui induit une écriture parasite lorsqu'on a stocké l'état '0'. Pour confirmer cette observation, nous avons tracé les courbes des densités d'électrons et des trous le long du film de silicium durant la lecture de l'état '0' (figure 2.22.b). La densité de trous dans le body et la densité d'électrons dans le bridge augmentent avec V_{dRead} . Ceci confirme l'écriture parasite par effet tunnel bande à bande durant la lecture de l'état '0' si V_{dRead} est trop élevé.

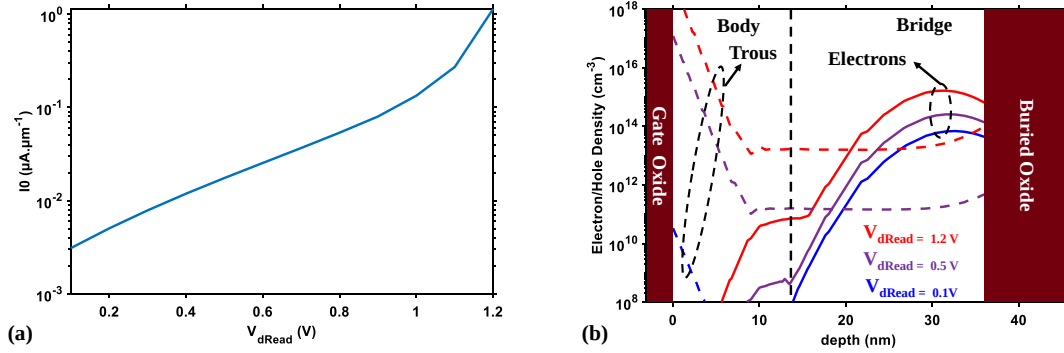


Figure 2.22 : (a) Courant lu de l'état '0' (I_0) en fonction de la tension de lecture du drain $V_{d\text{Read}}$ extraits des simulation TCAD. (c) Courbes représentant des évolutions de la densité d'électrons (traits continus) et des trous (traits discontinus) dans la profondeur du film de silicium (body + bridge) pour 3 différentes polarisations $V_{d\text{Read}}$ toutes extraits durant la lecture de l'état '0'.

En conclusion, nous devons maintenir $V_{d\text{Read}} \leq 500\text{ mV}$ pour éviter les écritures parasites via le courant de fuite par GIDL. Par ailleurs, la marge entre I_1 et I_0 diminue lorsque $V_{d\text{Read}}$ augmente, et si $V_{d\text{Read}}$ diminue, I_1 devient faible d'où une valeur idéale de $V_{d\text{Read}}$ est 200 mV.

II.2.6 Variation de la polarisation de grille arrière V_b

La structure que nous simulons n'a pas de plan de masse, mais un métal 'mid-gap' en face arrière. L'étude de la variation du V_b sur les performances mémoires nous permet d'évaluer l'impact du travail de sortie de la face arrière sur les performances mémoires. Nous utiliserons le motif de tension de la figure 2.23.a, en faisant varier V_b de -1 V à 1 V.

CHAPITRE II : DESCRIPTION DU FONCTIONNEMENT ELECTRIQUE D'UN POINT mémoire A2RAM VIA SIMULATIONS TCAD

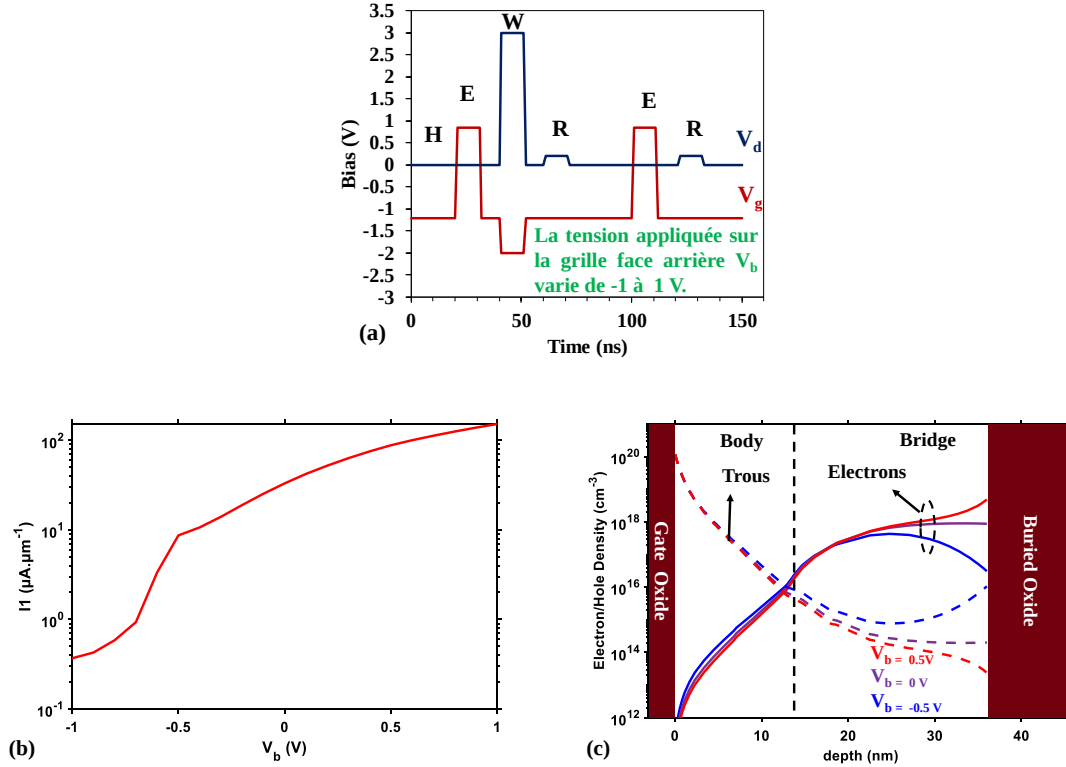


Figure 2.23 : (a) Chronogramme de tensions appliqué sur le drain V_d et la grille face avant V_g . La tension de la grille face arrière V_b varie de -1 V à 1 V. (b) Courant lu de l'état '1' (I_1) en fonction de la tension de la grille face arrière V_b extrait des simulations TCAD. (c) Evolution de la densité d'électrons (traits continus) et de trous (traits discontinus) dans la profondeur du film de silicium (body + bridge) pour 3 différentes polarisations V_b toutes extraites durant la lecture de l'état '1'.

Les extractions de I_1 sont représentées sur la figure 2.23.b en fonction de V_b . Une polarisation négative en face arrière ($V_b < 0$ V) favorise la déplétion du bridge, c'est pourquoi I_1 décroît. Une polarisation positive ($V_b > 0$ V) en face arrière favorise l'accumulation dans le bridge et I_1 augmente. Ceci est confirmé par le profil vertical de la densité de trous, et d'électrons dans le film de silicium durant la phase de lecture de l'état '1' (figure 2.23.c).

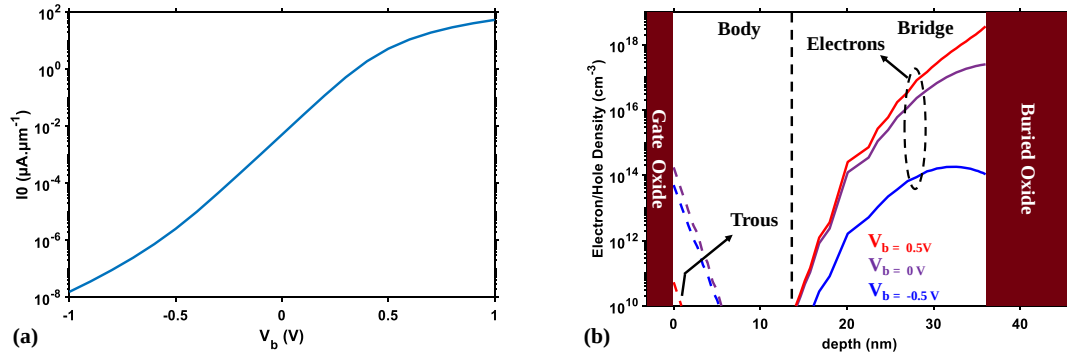


Figure 2.24 : (a) Courant lu de l'état '0' (I_0) en fonction de la tension de la grille face arrière V_b extrait des simulation TCAD. (b) Evolution de la densité d'électrons (traits continus) et de trous (traits discontinus) le long du film de silicium (body + bridge) pour 3 différentes polarisations V_b extraites durant la lecture de l'état '0'.

Sur la figure 2.24.a nous avons les variations de I_0 en fonction de V_b et sur la figure 2.24.b les profils de la densité de trous et d'électrons dans le film de silicium durant la lecture de l'état '0'. Les observations des figures 2.23.b et 2.23.c sont retrouvées sur la figure 2.24. I_0 augmente avec V_b car il favorise un régime d'accumulation en face arrière, sans toutefois impacter la densité de trous dans le body.

Pour garantir un bon compromis entre les valeurs de I_1 et de I_0 , les figures 2.23.c et 2.24.b semblent concorder sur le fait que la tension de grille face arrière V_b doit être égale à 0 V; cette valeur est dû au fait qu'on utilise en face arrière un métal 'mid-gap'. Car, un métal 'mid-gap' garantit une condition de bande plate en face arrière pour une polarisation sur la grille face arrière $V_b = 0$ V. C'est d'ailleurs pourquoi les densités d'électrons dans le bridge (juste au-dessus du BOX) sur les figures 2.23.c et 2.24.b sont horizontales lorsque $V_b = 0$ V.

En conclusion, il est intéressant d'appliquer une polarisation face arrière garantissant une condition de bande plate afin d'avoir un contrôle électrostatique du bridge seulement par la grille face avant. Dans notre cas, puisque nous avons déjà un métal 'mid-gap' en face arrière, la polarisation de la grille face arrière qui garantit une condition de bande plate est $V_b = 0$ V.

II.2.7 Variation de la tension de grille face avant pendant l'effacement :

$$V_{g\text{Erase}}$$

Nous allons étudier ici principalement l'impact de la tension de grille pendant l'effacement $V_{g\text{Erase}}$ sur la valeur du courant en lecture de l'état '0', en utilisant la séquence d'opération vue sur la figure 2.25.a.

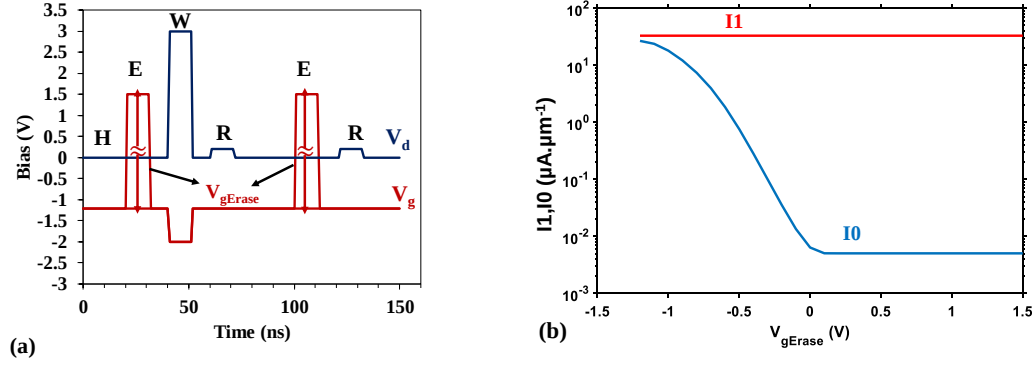


Figure 2.25 : (a) Chronogramme de tensions appliquées sur le drain V_d et la grille face avant V_g , la tension d'effacement V_{gErase} varie de -1.2 V à 1.5 V. (b) Courant lu de l'état '0' (I_0) en fonction de la tension d'effacement V_{gErase} extrait des simulations TCAD.

La figure 2.25.b montre I_0 en fonction de V_{gErase} obtenu par des simulations TCAD lorsqu'on applique le motif de tension de la figure 2.25.a. Naturellement, I_1 n'est pas affecté par V_{gErase} .

Pour $V_{gErase} \geq 0$ V l'efficacité de l'effacement reste identique, mais si $V_{gErase} < 0$ V, les trous ne sont pas totalement évacués du body, et la valeur du courant I_0 augmente progressivement jusqu'à I_1 . L'effacement est donc effectué lorsque $V_{gErase} \geq 0$ V.

II.3. Conclusion

Dans ce chapitre, nous avons fait une étude du comportement de la cellule A2RAM en analysant le rôle de chaque polarisation appliquée pendant chaque opération mémoire via l'outil TCAD Synopsys. Nous avons commencé par développer une méthodologie de simulation TCAD, et nous l'avons validée par des mesures électriques. Nous avons alors pu mettre en évidence l'importance du profil de dopage dans la profondeur du film de silicium sur les performances mémoires. Nous avons par la suite étudié les variations des performances mémoires : I_1 , I_0 et le temps de rétention lorsque l'on varie les polarisations lors des différentes opérations mémoires. Les variations des conditions d'écriture de l'état '1' ont montré que pour garantir l'écriture de l'état '1' en 10 ns par le courant de GIDL, il faut que la tension sur le drain V_d soit supérieure ou égale à 3 V, et sur la grille la tension doit être inférieure ou égale à -1.7 V. L'écriture de l'état '0' se faisant par couplage capacitif des grilles avant et arrière, nous considérons que l'effacement est effectif lorsque $V_{gErase} \geq 0$ V. Pour avoir un bon compromis entre I_1 , I_0 et le temps de rétention il est préférable d'avoir $V_{gRead} = V_{gHold} \leq V_{FB}$ (tension de bandes plates). Le choix de la valeur de la tension de lecture sur le drain est essentiellement limité par le courant de GIDL, c'est pourquoi il est préférable d'avoir $V_{dRead} \leq 200$ mV. Pour finir, l'étude des variations de la tension sur la grille face arrière révèle que les performances mémoires sont maximisées lorsque la condition de bandes plates est respectée dans le bridge.

CHAPITRE III : OPTIMISATION DE LA CELLULE A2RAM PAR UNE ETUDE DE SENSIBILITE TCAD

Dans ce chapitre, nous ferons une étude d'optimisation des performances mémoires de la cellule A2RAM via un outil de simulation TCAD (Technologie Computer-Aided Design). Nous commencerons par présenter un nouveau mécanisme de programmation de l'état '1'. Nous poursuivrons avec une étude de la sensibilité des performances du point mémoire A2RAM, pour des variations des paramètres technologiques. Cette étude nous permettra d'établir une méthodologie pour le choix des paramètres technologiques, et des structures optimisées de cellules A2RAM. Nous finirons par une étude des voies d'optimisation de ses performances mémoires.

III.1. Mise en évidence du nouveau mécanisme de programmation de l'état '1' dans cellule A2RAM

Pour rappel, l'opération mémoire est liée à la présence d'un excès ou d'un manque de porteurs majoritaires (trous) dans le body, qui module le flux de courant à travers le bridge. Une grande quantité de charge stockée garantie un courant de lecture élevé de l'état '1' (I1). Dans le Chapitre 2, nous avons vu que la programmation de l'état '1' est réalisée par la génération de charges via un effet tunnel bande à bande (B2B) au niveau de la zone de recouvrement drain-oxyde de grille face avant. Nous savons que l'avantage d'utiliser ce mécanisme est la faible consommation d'énergie et l'inconvénient est la lenteur de la vitesse d'écriture. C'est d'ailleurs pourquoi pour les simulations effectuées dans le Chapitre 2 pour un temps de programmation de 10 ns, il faut utiliser une polarisation de grille $V_{gW1} \leq -1.7$ V et de drain $V_{dW1} \geq 3$ V. Dans ce paragraphe, nous allons refaire la même étude que celle de la section 2.1 du Chapitre 2, mais ici nous allons aussi activer le mécanisme de génération de la charge par ionisation par impact. Ce mécanisme permet une nouvelle façon de programmer l'état '1' dans une A2RAM, avec une technique qui combine un temps d'écriture faible et un fonctionnement à basse tension.

III.1.1. Présentation du nouveau mécanisme de programmation de l'état '1'

Nous allons montrer qu'avec une polarisation de programmation appropriée appliquée sur la grille face avant (V_{gW1}) et sur le drain (V_{dW1}), l'écriture de l'état '1' dans une A2RAM peut être réalisée plus efficacement par ionisation par impact dans le bridge.

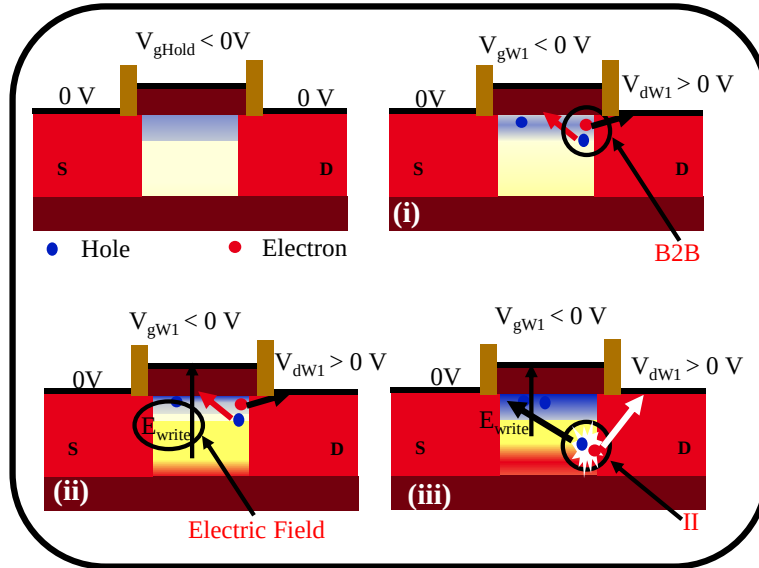


Figure 3.1 : Description du nouveau mécanisme de programmation de l'état '1' dans une A2RAM avec les polarisations appliquées sur la grille et le drain (la source est à masse), (i) Etat de maintien (body et bridge complètement désertés), (ii) début de l'écriture par le mécanisme d'effet tunnel B2B due à V_{gw1} et V_{dw1} , (iii) les trous générés par B2B sont stockés dans le body et le courant du bridge commence à circuler. (iv) L'activation de la génération II au niveau de la jonction bridge-drain qui complète la programmation et réduit le temps d'écriture à des faibles polarisations V_{dw1} et $|V_{gw1}|$.

Comme indiqué sur la figure 3.1, après une phase de maintien de l'état '0' où le body et bridge sont désertés, les tensions de programmation sont ensuite appliquées sur la grille face avant et le drain. Dans un premier temps (figure 3.1.i), le couple de tension (V_{gw1} , V_{dw1}) va fixer le taux de génération par effet tunnel bande à bande (B2B) dans la structure. Si on veut un B2B élevé, il faut maximiser ce couple de tensions V_{dw1} et $|V_{gw1}|$.

Durant la première (figure 3.1.i) et la seconde (figure 3.1.ii) phase, l'idée est de générer suffisamment de charges positives (trous dans le cas d'une A2RAM avec un bridge de type n) qui viendront faire écran au champ électrique vertical (E_{write}) induit par la grille, et permettre la conduction dans le bridge. Pour minimiser le temps d'écriture, il faut avoir une valeur du champ électrique vertical raisonnable afin de permettre au bridge de conduire le courant aussi tôt que possible. Par conséquent, ces phases demandent une valeur de $|V_{gw1}|$ faible, car une fois fixé le temps d'écriture :

- Si $|V_{gw1}| \downarrow \rightarrow E_{write} \downarrow \rightarrow$ besoin de moins de génération par effet tunnel bande à bande.
- Si $|V_{gw1}| \uparrow \rightarrow E_{write} \uparrow \rightarrow$ besoin de plus de génération par effet tunnel bande à bande.

La troisième phase (figure 3.1.iii) est marquée par la conduction du courant dans le bridge. Lorsque le bridge se met à conduire, au niveau de l'interface drain-bridge le fort champ électrique (due à la polarisation du drain) qui y règne accélérera les électrons ce qui leur permettra de générer des paires électrons-trous par ionisation par impact. Ainsi, le stockage sera complété par la génération de la charge par ionisation par impact au niveau du bridge.

L'efficacité de l'ionisation par impact est déterminée par la valeur de V_{dW1} . Plus elle sera grande, mieux on génèrera et stockera de la charge dans le body jusqu'à l'auto-polarisation de ce dernier.

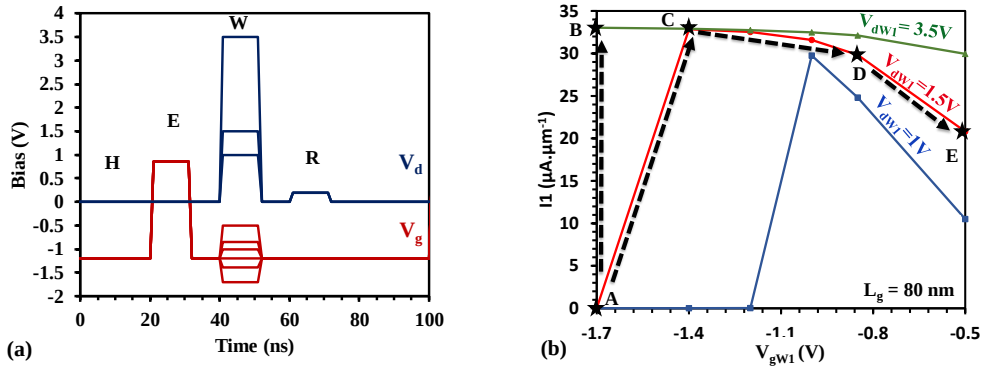


Figure 3.2 : (a) V_d et V_g en fonction du temps pour la séquence E-W-R avec une largeur d'impulsion de 10 ns et un front montant/descendant de 1 ns. (b) Courant de lecture de l'état I1 en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) extrait des simulations TCAD.

Pour mettre en évidence ce nouveau mécanisme, nous avons utilisé le motif de tension de la figure 3.2.a en faisant varier V_{gW1} et V_{dW1} . Pour chaque couple de tensions (V_{gW1} , V_{dW1}), nous avons extrait la valeur du courant lu de l'état '1' (I_1), et les résultats obtenus sont représentés sur la figure 3.2.b.

- De A à B, V_{dW1} augmente et V_{gW1} est constant : semblable aux variations de la figure 2.9, la génération par effet tunnel bande à bande est améliorée lorsque l'on augmente $V_{dW1} - V_{gW1}$, ce qui conduit à un courant I_1 supérieur. Cette observation est tout à fait cohérente avec la description du mécanisme de génération de la charge par effet tunnel bande à bande présenté dans le Chapitre 2.
- De A à C dans la figure 3.2.b, V_{dW1} est constant et V_{gW1} passe de -1.7 V à -1.4 V, donc le champ électrique entre grille et drain diminue. On peut noter que la valeur du courant en lecture de l'état '1' I_1 est supérieure lorsque $V_{gW1} = -1.4$ V ; ce qui revient à stocker plus de charge lorsque le champ électrique est faible. Cette observation n'est pas cohérente avec le cas typique (figure 2.9) d'une programmation uniquement avec génération de la charge par effet tunnel bande à bande car lorsque $|V_{gW1}|$ diminue, le champ électrique dans la zone de recouvrement oxyde de grille/drain diminue.

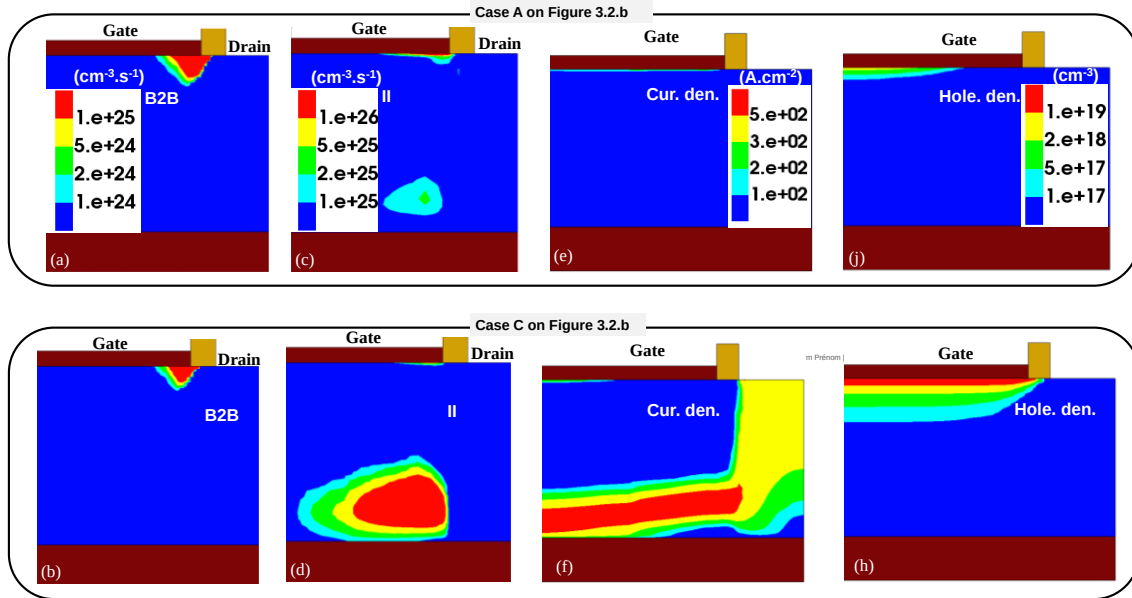


Figure 3.3 : a, c, e et j sont les cartographies pour le cas A ($V_{dW1} = 1.5$ V et $V_{gW1} = -1.7$ V) de la figure 3.2.b ; b, d, f et h sont les cartographies pour le cas C ($V_{dW1} = 1.5$ V et $V_{gW1} = -1.4$ V). (a, b) Profil du taux de génération B2B extrait à 10% de la phase d'écriture ; (c, d) profil du taux de génération II extrait à 10% de la phase d'écriture; (e, f) profil de la densité du courant extrait à 10% de la phase d'écriture ; (j, h) profils de densité des trous extraits durant la phase de maintien de l'état '1'; toutes les cartographies sont extraites des simulations TCAD.

Ceci est confirmé par les figures 3.3.(a, b), qui montrent le taux de génération B2B au cours de la programmation du '1' dans les cas A et C. La charge stockée dans le body pour le cas C ne s'explique donc pas par la génération par effet tunnel bande à bande. Même si c'est inattendu, il faut s'intéresser aux autres mécanismes de génération de la charge, comme l'ionisation par impact. Les figures 3.3.(c, d) montrent le taux de génération par ionisation par impact au milieu du créneau de programmation dans les cas A et C. On observe une forte génération au niveau de la jonction bridge/drain uniquement dans le cas C (plus faible $|V_{gW1}|$). En effet, puisque le champ électrique vertical due à la grille face avant est plus faible dans le cas C, le peu de charge générée par effet tunnel bande à bande est stocké dans le body et ceci permet de faire écran rapidement à ce champ électrique et donc de mettre le bridge en conduction. Cette observation est confirmée par les cartographies de la densité de courant durant l'écriture de l'état '1' des figures 3.3.(e, f) : du courant ne circule dans le bridge que dans le cas C. Etant donné que le drain est polarisé positivement, au niveau de jonction bridge-drain il existe une zone de charge d'espace. Les électrons qui la traversent sont accélérés et, par la présence d'un fort champ électrique, il y a un phénomène de génération par ionisation par impact. Les électrons générés sont alors recueillis au drain et les trous sont attirés dans le body sous l'effet de la polarisation de la grille, complétant ainsi le stockage de la charge. Nous avons confirmation d'un meilleur stockage dans le cas C, comme observé sur les cartographies de la densité de

trous stockés juste après l'écriture de l'état '1' (figures 3.3.(j, h)). Ces observations sont toutes à fait conformes avec la théorie du nouveau mécanisme que nous avons introduit en début de cette partie.

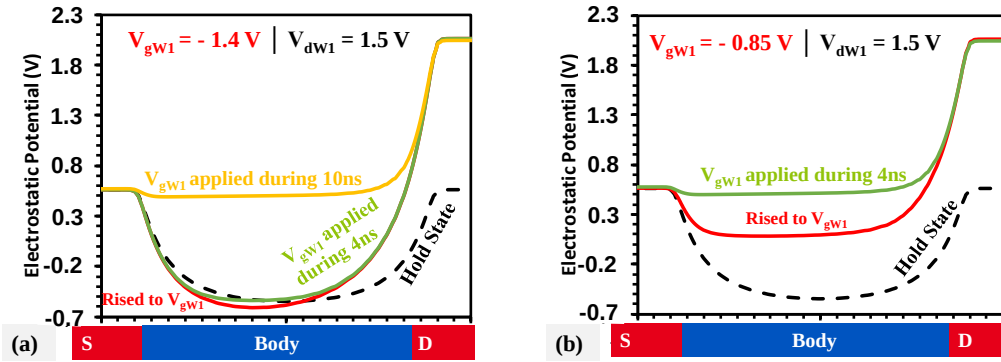


Figure 3.4 : Evolution du potentiel électrostatique de la phase maintien de l'état '0' à la phase d'écriture de l'état '1' de 0 à 10 ns. Les coupes des courbes ont été réalisées à 1 nm en-dessous de l'oxyde de grille face avant pour (a) le cas C, et (b) le cas D.

- De C à D : si nous gardons toujours V_{dw1} fixe et réduisons de plus en plus $|V_{gw1}|$ la valeur de I_1 diminue. Lorsque V_{gw1} se rapproche des valeurs positives, le body n'est plus en régime d'accumulation mais tends à un régime de désertion. Ainsi, son potentiel électrostatique diminue ce qui a pour conséquence de réduire la barrière de potentiel source-body. Dans cette condition on stocke moins de charge dans body, d'où la réduction de I_1 . Pour illustrer cette observation, nous avons extrait les coupes du potentiel électrostatique le long de la jonction source-body-drain dans les cas C (figures 3.4.a) et D (figure 3.4.b). En passant de l'état de maintien (courbes rouges sur les figures 3.4.(a, b)) au début de la phase d'écriture, la barrière de potentiel électrostatique est réduite dans le cas D comparée au cas C. C'est pourquoi moins de trous sont nécessaires dans le body dans le cas D pour atteindre l'autopolarisation du body (la jonction body-source passe en direct). Ceci explique également pourquoi l'autopolarisation du body est atteinte moins vite dans le cas C (évolution du potentiel électrostatique plus lente), car ici on a besoin de plus de charge dans le body pour atteindre l'autopolarisation du film de silicium.

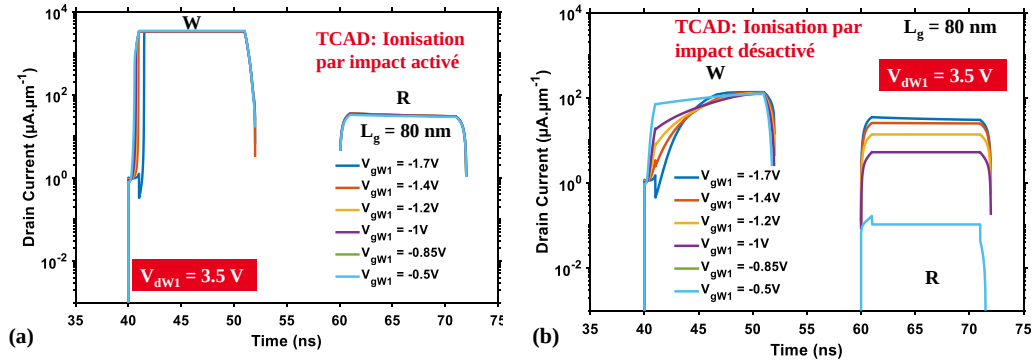


Figure 3.5 : Variation du courant de drain de l'A2RAM (figure 2.1.b) en fonction du temps lorsqu'on applique sur la cellule le chronogramme de tensions de la figures 3.2.a extrait de TCAD avec (a) l'ionisation par impact activée, (b) l'ionisation par impact désactivée.

- De D à E : pour de plus faibles $|V_{gW1}|$ que dans le cas D, l'écriture ne dépend que de l'ionisation par impact dans le bridge. Pour le démontrer, nous représentons les chronogrammes de courant durant la phase d'écriture et de lecture de l'état '1' à différent V_{gW1} pour $V_{dW1} = 3.5$ V (figure 3.5) : avec $V_{dW1} = 1.5$ V, l'ionisation par impact prédomine quel que soit V_{gW1} avec un temps d'écriture de 10 ns. Nous considérons deux cas : le mécanisme de génération par ionisation par impact est activé (figure 3.5.a) et désactivé (figure 3.5.b) dans nos simulations. Sur la figure 3.5.a, durant l'écriture les amplitudes du courant se confondent, car l'ionisation par impact dépend de V_{dW1} comme nous l'avions mentionnée dans nos explications de notre nouveau mécanisme. Durant la lecture de l'information (figure 3.5.a), les valeurs de courants se confondent presque tous, preuve que le stockage était efficace. Sur la figure 3.5.b, nous notons déjà que les amplitudes de courant en écriture n'arrivent plus à la même valeur comme sur la figure 3.5.a. Car, ici le seul mécanisme activé est la génération par B2B, d'où le faible courant. Bien que durant la phase d'écriture les courants semblent saturer, pour $V_{gW1} > -1.2$ V sur la figure 3.5.b, on peut noter que les courants en lecture restent faibles comparés aux cas de la figure 3.5.a, ceci signifie que la charge stockée est faible. Ainsi, pour des polarisations $V_{gW1} > -1.2$ V, l'efficacité du stockage dépendra fortement de la polarisation du drain V_{dW1} .
- De $V_{dW1} = 1.5$ V à 1 V : nous constatons qu'indépendamment de la valeur de V_{gW1} que la valeur du courant de l'état '1' lu I1 est réduite. Ceci est expliqué par le fait que les mécanismes de génération de la charge (effet tunnel bande à bande et ionisation par impact) sont de plus en plus faibles. On stocke alors de moins en moins efficacement la charge dans le body (pour un temps de 10 ns). Pour mettre en évidence cette observation, dans un premier temps, nous avons extrait plusieurs lignes de coupe du profil du taux de génération par effet tunnel bande à bande faites à 1 nm en dessous de l'oxyde grille face avant en début de la phase d'écriture (c'est-à-dire lorsque V_g passe à V_{gW1} et V_d passe à V_{dW1} sur la figure 3.2.a) : lorsque $V_{dW1} = 1.5$ V (figures 3.6.a) et $V_{dW1} = 1$ V (figures 3.6.b). Comme décrit dans le Chapitre 2, l'effet tunnel B2B est plus faible lorsque $|V_{dW1} - V_{gW1}|$ diminue, donc moins de charges stockées. Dans le cas particulier de $V_{dW1} = 1$ V, le taux de

génération par effet tunnel est très faible, et on ne peut pas stocker suffisamment de la charge en 10 ns pour $V_{gW1} > -1.2$ V, ceci explique en partie pourquoi la valeur du courant I1 est faible sur la figure 3.2.b.

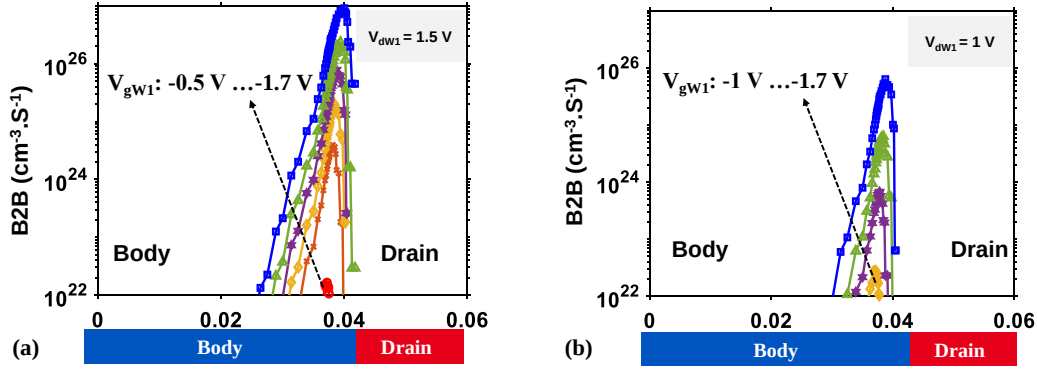


Figure 3.6 : (a, b) Taux de génération B2B le long de la zone de chevauchement de la grille et du drain. Les courbes ont été extraites en faisant une coupe horizontale à 1 nm en-dessous de l'oxyde de grille.

- Sur les figures 3.7.(a,b), toujours pour $V_{dW1} = 1.5$ V et $V_{dW1} = 1$ V, sont représentés les profils de la génération par ionisation par impact le long de source-bridge-drain à différent V_{gW1} . Toutes les coupes sont faites à 5 nm au-dessus du BOX en début de la phase d'écriture. En général, le taux de génération par ionisation par impact est réduit lorsqu'on passe de $V_{dW1} = 1.5$ V à $V_{dW1} = 1$ V. Ceci s'explique par le fait que le champ électrique transversal dans la zone de charge d'espace entre le bridge et drain où a lieu la génération par ionisation par impact dépend de la tension appliquée sur le drain. Par ailleurs, dans le cas particulier de $V_{dW1} = 1$ V, l'ionisation par impact est très élevée seulement pour $V_{gW1} \geq -1.4$ V. Au vu des résultats de la figure 3.6.b c'est le seul mécanisme responsable du stockage de la charge, d'où le fort courant I1 sur la figure 3.2.b.

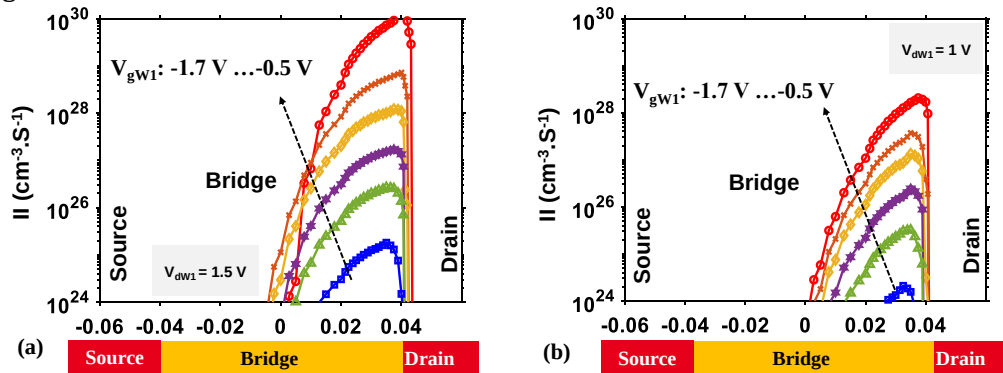


Figure 3.7 : Taux de génération d'ionisation par impact (II) le long de source-bridge-drain. Courbes extraites en faisant une coupe horizontale à 5 nm au-dessus du BOX. Toutes les courbes sont extraites en début de la phase d'écriture à l'état '1'.

En réduisant la valeur de V_{dW1} , puisque la génération par ionisation par impact devient faible, pour stocker suffisamment de la charge il faut augmenter le temps de programmation. C'est pourquoi nous avons fait les mêmes simulations que sur la figure 3.2, avec un temps de programmation à 300 ns. Les extractions des courants $I1$ sont montrées sur la figure 3.8 avec pour $V_{dW1} = 1.5$ V et $V_{dW1} = 1$ V. Nous pouvons constater que le stockage avec $V_{dW1} = 1$ V est tout aussi efficace que celui avec $V_{dW1} = 1.5$ V. De plus, nous pouvons programmer l'état '1' avec une tension de grille face $V_{gW1} < -1.4$ V pour $V_{dW1} = 1.5$ V et avec $V_{gW1} < -1$ V pour $V_{dW1} = 1$ V comparé à la figure 3.2.b.

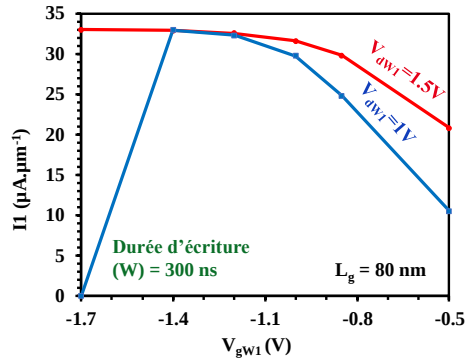


Figure 3.8 : Courant de lecture de l'état '1' $I1$ en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) avec une durée de temps d'écriture (W) de 300 ns extrait des simulations TCAD.

III.1.2. Impact des modèles de transport de la charge sur le nouveau mécanisme d'écriture de l'état '1'

Le mécanisme de programmation de l'A2RAM mis en évidence précédemment fait donc appel à l'ionisation par impact à la jonction bridge-drain. Afin de borner les polarisations à mettre en œuvre pour utiliser cette nouvelle méthode de programmation, mais aussi pour vérifier qu'il ne s'agit pas d'un 'artéfact' de simulation, nous avons effectué les mêmes simulations de la figure 3.2 mais cette fois avec le modèle de transport de la charge : l'Hydrodynamique (HD). En effet, comme décrit dans la section 1 du Chapitre 2, le modèle HD est connu pour sous-estimer la génération de porteurs par ionisation par impact du fait des variations de la température des porteurs. Alors que le transport par dérive-diffusion (DD) surestime la génération de porteurs par ionisation par impact, car il suppose que les porteurs et le réseau cristallin sont en équilibre thermodynamique. La variation du courant $I1$ en fonction de la tension de la grille pendant la programmation obtenue pour ces deux modèles de transport de la charge est présentée sur la figure 3.9. Les allures des courbes obtenues sont proches pour les deux modèles de transport de la charge mais pour des polarisations différentes : le courbes $V_{dW1}=2.5V$ en HD correspond à celle $V_{dW1}=2V$ en DD. Les différences s'expliquent par le plus faible taux de génération par ionisation par impact prévu en HD. Ceci sera démontré en effectuant les mêmes simulations, mais en désactivant les modèles de génération de charge un à un.

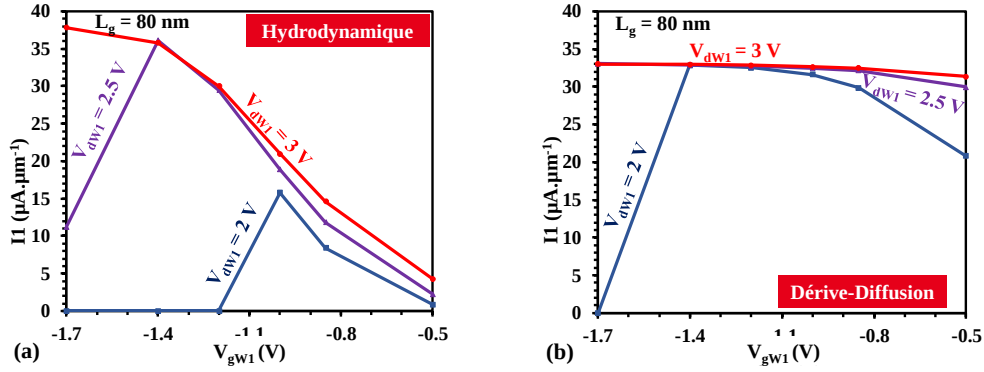


Figure 3.9 : Courant de lecture de l'état '1' I_1 en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) extrait des simulations TCAD : comparaison (a) HD et (b) DD.

Sur la figure 3.10, nous montrons les variations du courant I_1 lorsque le mécanisme d'effet tunnel bande à bande est désactivé pour des simulations effectuées avec les deux modèles de transport de la charge (HD et DD). En comparant avec la figure 3.9, on constate que les résultats sont les mêmes avec et sans génération bande à bande, sauf pour $V_{gW1} = -1.7$ V dans le cas de simulations HD. Ceci démontre bien que la programmation de l'A2RAM s'effectue dans les deux cas par ionisation par impact dans le bridge.

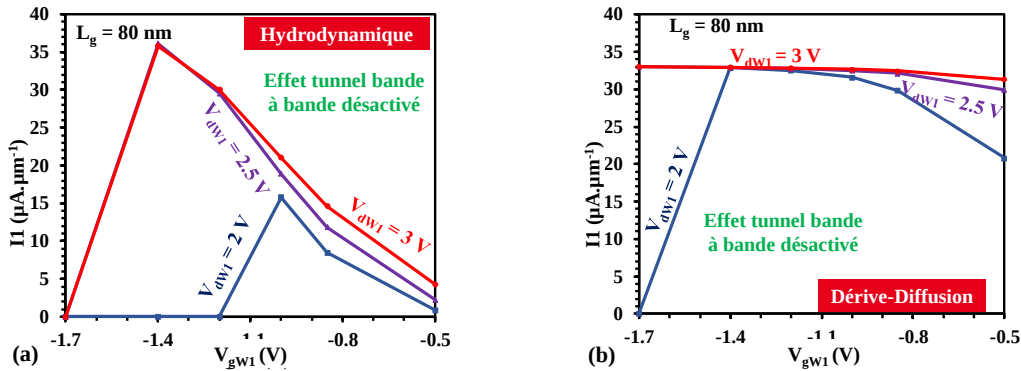


Figure 3.10 : Courant de lecture de l'état '1' I_1 en fonction des couples de tension d'écriture (V_{gW1} , V_{dW1}) lorsque le mécanisme de génération de la charge par effet tunnel bande à bande est désactivé des simulations TCAD : comparaison (a) HD et (b) DD.

Sur la figure 3.11, comme précédemment nous comparons la variation de I_1 avec V_{gW1} obtenue avec les deux modèles de transport, en désactivant la génération par ionisation par impact. Comme attendu, les résultats obtenus avec HD et DD sont presque identiques. Ensuite, on constate que la programmation de l'état '1' n'a lieu que pour de fortes polarisations de grille et de drain, ce qui est cohérent avec une programmation par GIDL uniquement. Ce résultat démontre également qu'une programmation rapide et à basse tension n'est possible que par ionisation par impact dans le bridge.

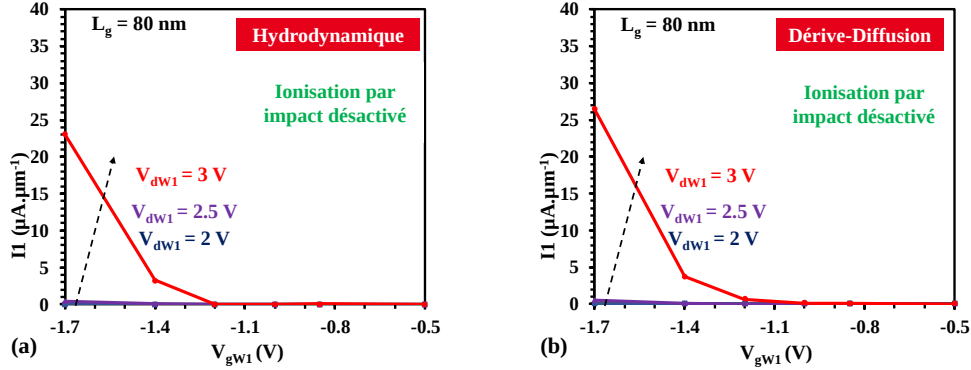


Figure 3.11 : Courant de lecture de l'état '1' I_1 en fonction des couples de tension d'écriture (V_{gw1} , V_{dw1}) lorsque le mécanisme de génération de la charge par ionisation par impact est désactivé des simulations TCAD : comparaison (a) TH et (b) DD.

Ce paragraphe nous permet de démontrer l'existence de la programmation dans l'A2RAM par ionisation par impact dans le bridge. Les simulations effectuées avec les deux modèles de transport de charge nous permettent aussi de borner les polarisations à appliquer pour une mise en évidence expérimentale, qui sera conduite au paragraphe suivant.

III.1.3. Tentative de mise en évidence expérimentale du nouveau mécanisme de programmation de l'état '1' de l'A2RAM

Le but ici est de mettre en évidence ce nouveau mécanisme de programmation expérimentalement, par caractérisation électrique des dispositifs A2RAM précédemment fabriqués au Leti [Rodriguez 12]. En l'occurrence, nous voulons retrouver cette remontée de courant du I_1 lorsque l'on réduit la valeur de V_{dw1} ($V_{dw1} \leq 1.5 \text{ V}$) comme présenté sur la figure 3.2.b qui est une signature claire du nouveau mécanisme. Il ne sera cependant pas possible de se mettre dans les mêmes conditions en mesure qu'en simulations qui ont été réalisées avec des largeurs de créneau de 10 ns et des temps de montée et de descente de 1 ns. En effet, comme expliqué dans la section 1 du Chapitre 2, il faut tenir compte des temps de chargement et déchargement des capacité parasites. De plus, l'analyseur B1500A impose des durées de temps d'acquisition en fonction de l'amplitude des courants à mesurer. C'est pour ces raisons que nous utiliserons une durée des pulses d'environ 300 ns, et pour les fronts montants et descendants nous restons toujours à 100 ns (figure 3.12).

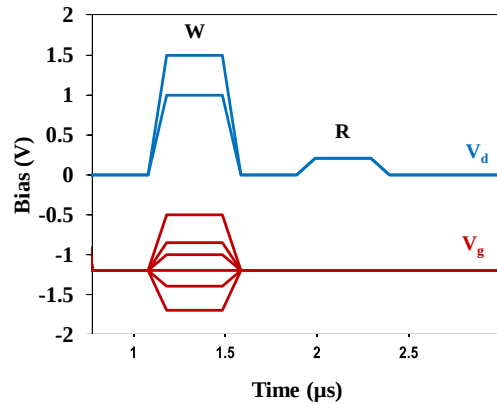


Figure 3.12 : V_d et V_g en fonction du temps pour la séquence W-R (la phase d'effacement n'est pas montrée) avec une largeur d'impulsion de 300 ns et un front montant/descendant de 100 ns.

Nous avons appliqué le chronogramme de tensions de la figure 3.12 sur notre échantillon et observé les variations du courant de drain. Les chronogrammes des courants sont représentés sur les figures 3.13.(a, b) pour $V_{dW1} = 1.5$ v et $V_{dW1} = 1$ V respectivement. Lorsque $V_{dW1} = 1.5$ V, les courants en écriture saturent tous, signe que le stockage est aussi maximum. Mais en lecture, les amplitudes de I_1 dépendent de V_{gW1} car comme nous l'avons montré, la quantité de la charge dans le body dépend de V_{gW1} (moins il est négatif, moins on stockera de la charge (figure 3.4)). Pour $V_{dW1} = 1$ V, les courant en écriture ne sature plus, mais lorsque $V_{gW1} < -1.2$ V, nous observons une augmentation progressive du courant qui signifie que le stockage de la charge est en cours mais aussi que la charge maximale n'est pas encore atteinte car les amplitudes de courant ne saturent pas. C'est pourquoi durant la lecture de l'information, I_1 est élevé seulement pour $V_{gW1} < -1.2$ V. Les allures des courbes de la figure 3.13.b sont similaires à celles de la figure 3.5.b lorsqu'on a comme unique mécanisme de génération de la charge l'effet tunnel bande à bande.

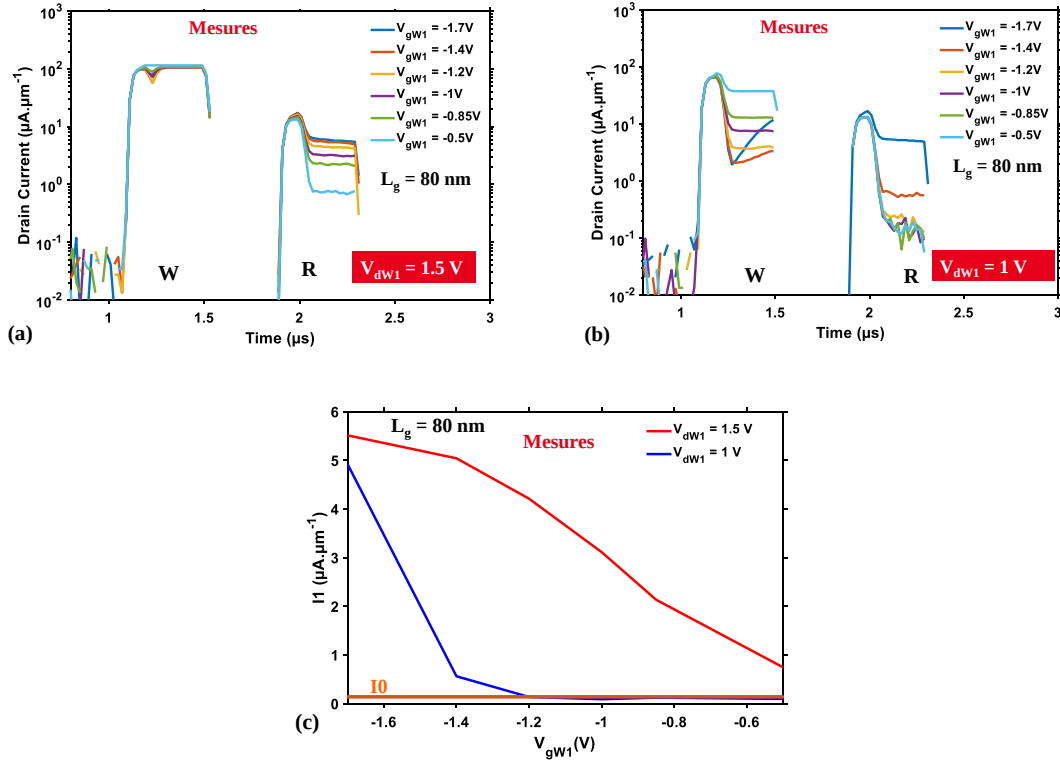


Figure 3.13 : Variation du courant de drain mesurée avec (a) $V_{dw1} = 1.5$ V et (b) $V_{dw1} = 1$ V. (c) Courant de lecture de l'état I1 en fonction des couples de tension d'écriture extrait des mesures électriques des courbes des figures 3.13.(a, b).

Pour finir, sur la figure 3.13.c sont reportées les extractions des courants I_1 faites sur les figures 3.13.a et 3.13.b; nous avons aussi indiqué la valeur du courant I_0 pour référence. Mais, que ça soit avec $V_{dw1} = 1.5$ V ou $V_{dw1} = 1$ V, nous n'observons pas les remontées des courbes (en référence à la figure 3.2.b). Il faut noter qu'ici, avec un temps d'écriture de 300 ns, la remontée n'est visible que pour une tension de drain en écriture inférieure à 1.5 V (figure 3.8). Il est assez difficile en mesure de faire une distinction entre le courant dû à l'effet tunnel bande à bande et celui de la génération par ionisation par impact comme nous l'avons fait précédemment en simulation. Par conséquent, à cause du temps d'écriture très long (300 ns) imposé par l'instrument de mesure, nous ne pouvons pas confirmer la présence du nouveau mécanisme sur ce lot de mémoire A2RAM.

III.2. Étude des variations des paramètres technologiques d'A2RAM sur ses performances

En partant de la géométrie de la structure A2RAM présentée à la figure 2.1, l'objectif de cette partie est d'optimiser cette architecture afin d'atteindre un meilleur compromis performances/densité d'intégration. Cette optimisation sera effectuée par le biais d'une analyse de sensibilité des performances du point mémoire A2RAM pour des variations de ses paramètres technologiques. Nous allons mettre en œuvre des simulations TCAD, en utilisant

la séquence de tensions de la figure 2.2 avec une largeur d'impulsion de 10 ns et un temps de montée/descente de 1 ns. Pour chaque variation étudiée, nous avons pris soin de choisir un couple de tensions (V_{gW1} , V_{dW1}) qui garantit un stockage maximum.

III.2.1. Réduction de la longueur de la grille (L_g)

A partir de la structure de référence sur la figure 2.1, nous évaluons ici l'effet de la longueur de grille (de 80 nm [Rodriguez 12] à 30 nm) sur les performances de l'A2RAM. La figure 3.14 montre la variation des courants de lecture des états '1' et '0' (I_1 et I_0) et du ratio I_1/I_0 en fonction de la longueur de la grille (L_g).

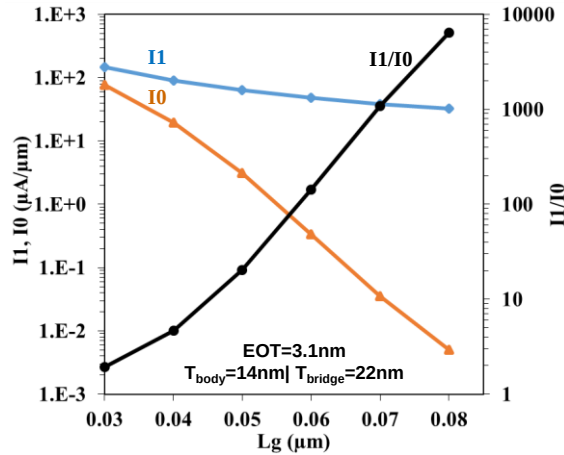


Figure 3.14 : Evolution de I_1 , I_0 (axe de gauche) et I_1/I_0 (axe de droite) en fonction de la longueur de grille de l'A2RAM obtenue par des simulations TCAD.

Lorsque L_g diminue, I_1 augmente car la longueur du bridge, et donc sa résistance diminue. I_0 augmente également car le contrôle électrostatique de la grille face avant sur le film de silicium (body + bridge) est dégradé par l'apparition des effets canaux courts (SCEs). Ainsi, autour de $L_g = 30$ nm, le rapport I_1/I_0 est très proche de l'unité (axe de droite de la figure 3.14), ce qui signifie que la grille face avant est trop courte pour assurer la fonctionnalité A2RAM.

Par ailleurs, nous avons vérifié la variation du temps de rétention n'est pas dégradé lorsqu'on réduit la longueur de la grille de 80 nm à 30 nm. Ainsi, nous avons évalué le temps de rétention suivant le même protocole que celui de la section 2.4 du Chapitre 2. Sur la figure 3.15, nous avons les variations du temps de rétention en fonction de la longueur de la grille. Pour $L_g = 70$ nm, le temps de rétention est légèrement supérieur à la structure de départ ($L_g = 80$ nm qui est de 0.44 ms), mais il décroît très rapidement pour des longueurs inférieures. Ainsi, bien que réduire L_g augmente le courant I_1 , les longueurs de grille inférieures à 70 nm sont limitées par l'augmentation très rapide de I_0 à cause des effets canaux courts (figure 3.14).

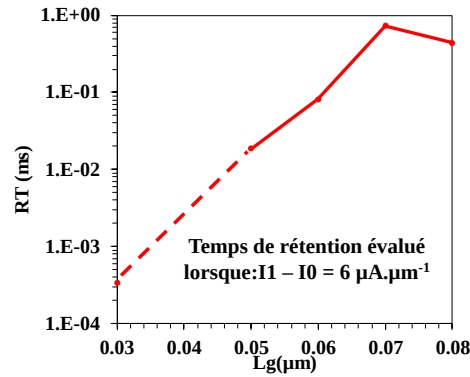


Figure 3.15 : Evolution du temps de rétention en fonction de la longueur de la grille de l'A2RAM.

Réduire la longueur de la grille (L_g) est donc efficace pour augmenter I_1 , mais la dégradation associée de I_0 est bien trop forte. Pour limiter cette dégradation, il faut améliorer le contrôle des effets canaux courts, par exemple en réduisant l'épaisseur équivalente d'oxyde de grille (EOT), l'épaisseur du body (T_{body}) et du bridge (T_{bridge}) et/ou le dopage du bridge (N_{bridge}).

III.2.2. Variation de l'épaisseur d'oxyde de la grille face avant (EOT)

Sur les figures 3.16.(a, b, et c), nous présentons les variations de I_1 et I_0 en fonction de l'épaisseur équivalente de l'oxyde grille EOT pour différentes longueurs de grille.

Comme on peut constater sur la figure 3.16.a, le courant I_1 reste constant lorsque l'on réduit l'EOT. Notons que pour le point de simulation avec $L_g = 80$ nm avec un $EOT = 1.2$ nm, on a une chute du courant I_1 parce que le champ électrique est beaucoup trop fort (EOT très mince et longueur très grande) ; il faudrait changer les conditions de lecture.

En revanche, le courant en lecture I_0 diminue (figure 3.16.b) lorsqu'on réduit l'EOT, ce constat reste vrai lorsqu'on réduit la longueur de la grille. C'est dû au bon contrôle électrostatique pour une faible épaisseur d'EOT, car les effets canaux courts sont réduits.

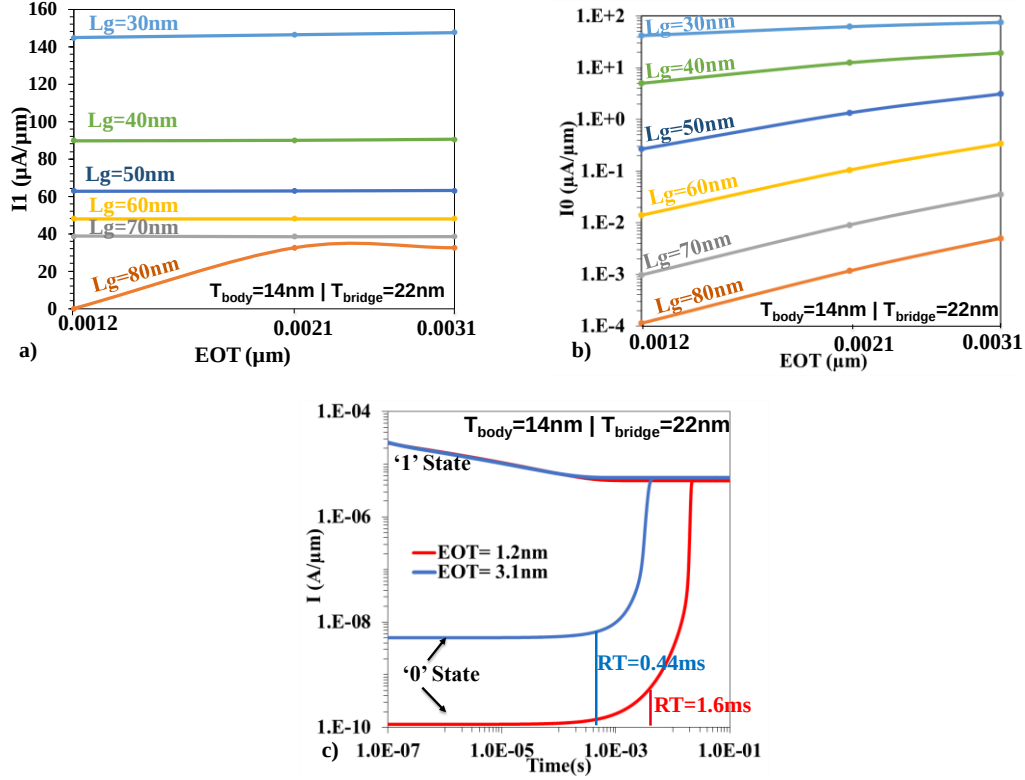


Figure 3.16 : Evolutions de (a) I_1 , et (b) I_0 en fonction de l'EOT pour différentes longueurs de grilles extrait de simulations TCAD. (c) Evaluation du temps de rétention : courant du drain en fonction du temps pour une lecture continue de l'état '1' et l'état '0' avec deux valeurs d'EOT.

Le temps de rétention est ensuite évalué (figure 3.16.c) pour deux valeurs d'EOT (3.1nm et 1.2nm) et une même longueur de grille ($L_g = 80nm$) utilisant le protocole de la section 2.4 du Chapitre 2. Le temps de rétention augmente quand l'EOT diminue (passe de 0.44 ms pour EOT = 3.1 nm à 1.6 ms pour EOT = 1.2 nm) car le contrôle électrostatique est amélioré. En conclusion, il est bénéfique de réduire l'EOT : nous avons peu d'effet sur I_1 , et une réduction de I_0 , donc une amélioration du temps de rétention.

III.2.3. Variation de l'épaisseur du body : T_{body}

Une seconde manière d'améliorer le contrôle électrostatique sur le point mémoire A2RAM est de réduire l'épaisseur du film de silicium, par exemple ici en réduisant celle du body (T_{body}). Les résultats de simulations (I_1 et I_0) sont présentés sur les figures 3.17.(a, b) en fonction de la longueur de la grille pour différentes valeurs de T_{body} .

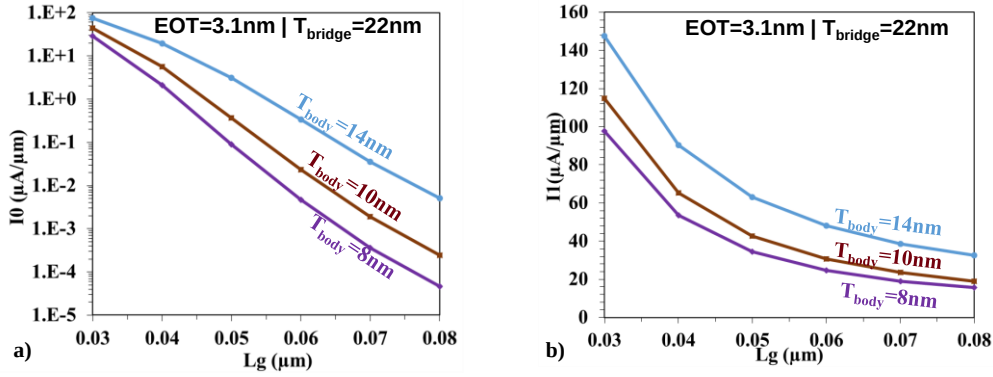


Figure 3.17 : Evolution de (a) I_0 et (b) I_1 en fonction de la longueur de la grille L_g , pour $T_{\text{body}} = 14\text{ nm}$ (référence), 10 nm et 8 nm ; extraits des simulations TCAD.

La réduction de T_{body} entraîne une diminution de I_0 (Figure 3.17.a). Ceci démontre une amélioration du contrôle électrostatique de la grille face avant (réduction du SCE). Cependant, comme on peut le voir sur la Figure 3.17.b, I_1 diminue également. En effet, la valeur de I_1 dépend, de l'efficacité de l'écrantage du champ électrique vertical de grille pendant la lecture, donc de la quantité de trous stockés dans le body. Pour vérifier cela, nous avons évalué la quantité de trous stockés pendant la phase de maintien de l'état '1' (après l'écriture de l'état '1') dans chaque dispositif, en intégrant la densité des trous dans le body. La figure 3.18 montre que le courant I_1 et la quantité de trous stockés augmentent avec T_{body} . Nous avons pris soin de vérifier que pour chaque dispositif la charge maximale dans le body était atteinte durant la phase d'écriture de l'état '1'. Par conséquent, réduire l'épaisseur du body réduit le volume où les porteurs majoritaires peuvent être stockés donc la quantité maximale de charges pouvant être stocké lors de la programmation.

La valeur optimale de T_{body} relève d'un compromis entre (i) amélioration du contrôle électrostatique pour réduire les fuites et I_0 , et (ii) volume de stockage de porteurs majoritaires lié à I_1 .

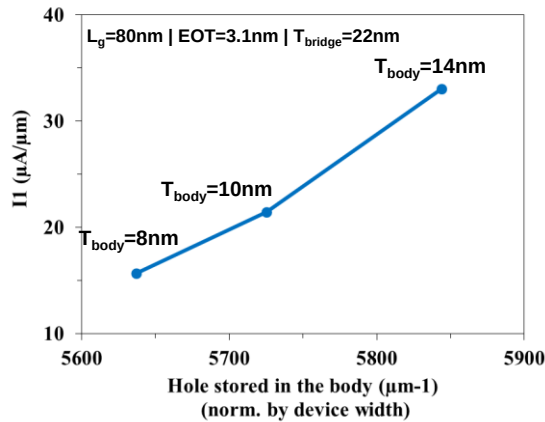


Figure 3.18 : I_1 en fonction de la quantité de trous stockés dans le body pour trois valeurs de T_{body} et pour une longueur de grille $L_g = 80\text{ nm}$.

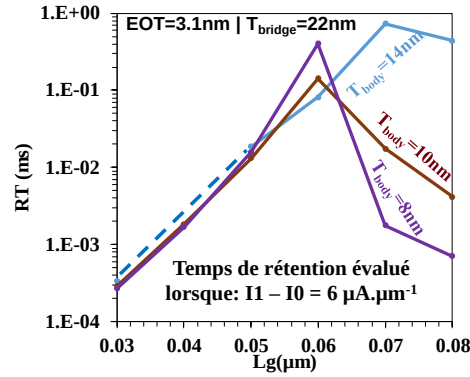


Figure 3.19 : Evolution du temps de rétention en fonction de la longueur de la grille de l'A2RAM, pour $T_{body} = 14\text{nm}$ (référence), 10 nm et 8 nm ; extraits des simulations TCAD.

Sur la figure 3.19, nous avons l'évolution du temps de rétention pour les deux épaisseurs du body (figure 3.17) en comparaison avec les structures de départ (figure 3.14). Pour des longueurs L_g inférieures à 60 nm, la réduction de T_{body} n'améliore pas le temps de rétention. Nous avons un meilleur temps de rétention seulement pour une longueur de grille de $L_g = 60$ nm.

III.2.4. Variation de l'épaisseur du bridge : T_{bridge}

Toujours dans l'objectif d'améliorer le courant de lecture de l'état '0', nous étudions ici l'effet de la réduction de l'épaisseur du bridge en conservant constants tous les autres paramètres. Le principe reste le même que dans la section 2.3 : nous avons tracé les variations de I_1 (figure 3.20.a) et I_0 (figure 3.20.b) en fonction de la longueur de la grille pour 3 épaisseurs de bridge (T_{bridge}) différentes : 22 (ref), 18 et 12 nm.

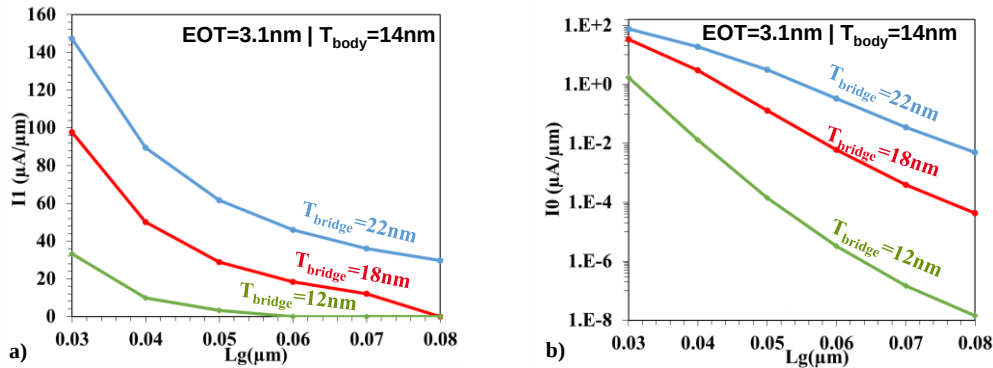


Figure 3.20 : Evolution de (a) I_1 et (b) I_0 en fonction de la longueur de la grille L_g , pour $T_{bridge} = 22\text{ nm}$ (référence), 18 nm et 12 nm ; extrait des simulations TCAD.

Sur la figure 3.20.a, nous observons une réduction du courant I_1 lorsqu'on réduit l'épaisseur du bridge (T_{bridge}), et une augmentation de I_1 lorsqu'on réduit la longueur de la grille L_g . En effet, le courant en lecture de l'état '1' I_1 est dû à la conduction dans le bridge, nous pouvons le modéliser par une loi ohmique donnée par l'équation 3.1.

$$I1 \approx \frac{V_{dRead} T_{bridge}}{\rho L_g}, \quad (3.1)$$

où ρ est résistivité du bridge, V_{dRead} la tension de drain de lecture. L'équation 3.1 montre que $I1$ est directement proportionnel à T_{bridge} , et inversement proportionnel à L_g , d'où les variations observées de $I1$ de la figure 3.20.a.

La réduction associée à $I0$ s'explique par deux phénomènes : la résistance du bridge augmente et le contrôle électrostatique de la grille face avant est amélioré puisque l'épaisseur totale du film de silicium est réduite. Mais lorsque $L_g < 40$ nm, les effets canaux courts dégradent le contrôle électrostatique du bridge par la grille face avant, d'où $I0 \approx I1$.

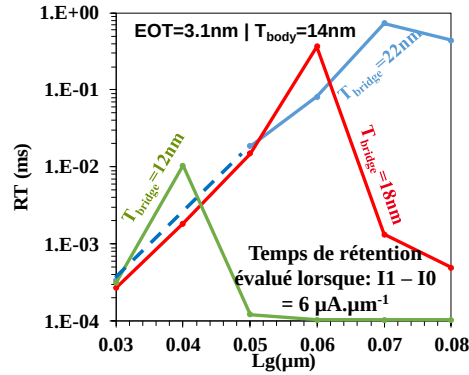


Figure 3.21 : Evolution du temps de rétention en fonction de la longueur de la grille de l'A2RAM, pour $T_{bridge} = 22$ nm (référence), 18 nm et 12 nm ; extraits des simulations TCAD.

Le temps de rétention est ensuite évalué pour deux structures A2RAM avec un body de même épaisseur (14 nm) et trois épaisseurs de bridge différentes (22, 18 et 12 nm). La figure 3.21 montre la variation du temps de rétention en fonction de la longueur de grille L_g . Comparé aux structures de départ (figure 3.14), le temps de rétention est amélioré pour $L_g = 60$ nm avec $T_{bridge} = 18$ nm, et $L_g = 40$ nm avec $T_{bridge} = 12$ nm.

Nous avons constaté que réduire les épaisseurs du body et du bridge améliore le contrôle par la grille, car le courant $I0$ est réduit. Mais, il s'associe aussi une réduction de $I1$ ce qui implique un compromis sur le choix des valeurs de T_{bridge} et T_{body} . Par ailleurs, lorsque l'on compare les temps de rétention obtenus sur les figures 3.19 et 3.21, pour améliorer le temps de rétention tout en gardant la valeur de $I1$ équivalente à la structure de départ (figure 2.1 avec $L_g \leq 40$ nm), il est préférable de réduire l'épaisseur du bridge (T_{bridge}).

III.2.5. Variation du dopage du bridge : N_{bridge}

Nous considérons toujours la structure initiale de la figure 2.1 (dopage bridge $N_{bridge} = 10^{18} \text{ cm}^{-3}$), mais nous faisons varier le dopage du bridge (N_{bridge}) de $3 \cdot 10^{18} \text{ cm}^{-3}$ à 10^{16} cm^{-3} . Comme précédemment, nous traçons $I1$ (Figure 3.22.a) et $I0$ (Figure 3.22.b) en fonction de la longueur de grille pour différents dopages du bridge.

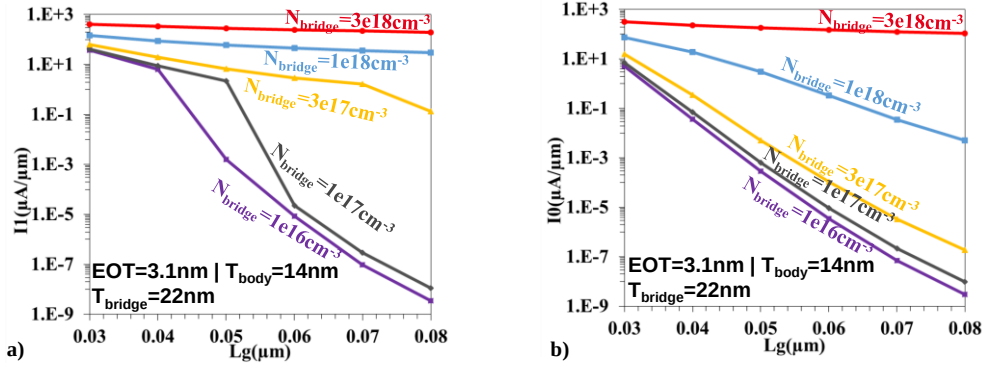


Figure 3.22 : Evolution de (a) I_1 et (b) I_0 en fonction de la longueur de la grille L_g , pour $N_{\text{bridge}} = 3 \cdot 10^{18}$, 10^{18} (référence), $3 \cdot 10^{17}$, 10^{17} , et 10^{16} cm^{-3} ; extraits des simulations TCAD.

L'effet du dopage du bridge sur I_1 (figure 3.22.a) est similaire à l'effet de l'épaisseur du bridge (section 2.4), car N_{bridge} a un effet sur sa résistivité (équation 3.1). Si N_{bridge} augmente, la résistance du bridge diminue et donc I_1 augmente, vice-versa la résistivité diminue lorsque N_{bridge} diminue d'où la réduction de I_1 lorsque $N_{\text{bridge}} < 10^{18} \text{ cm}^{-3}$.

Les variations de I_0 sur la figure 3.22.b montrent que pour $N_{\text{bridge}} < 10^{18} \text{ cm}^{-3}$, I_0 diminue car le champ électrique nécessaire pour maintenir le bridge en désertion est beaucoup plus faible. Bien évidemment, lorsqu'on réduit la longueur de la grille, le contrôle du film de silicium (body + bridge) est moins efficace, d'où l'augmentation du courant I_0 lorsque L_g est réduit.

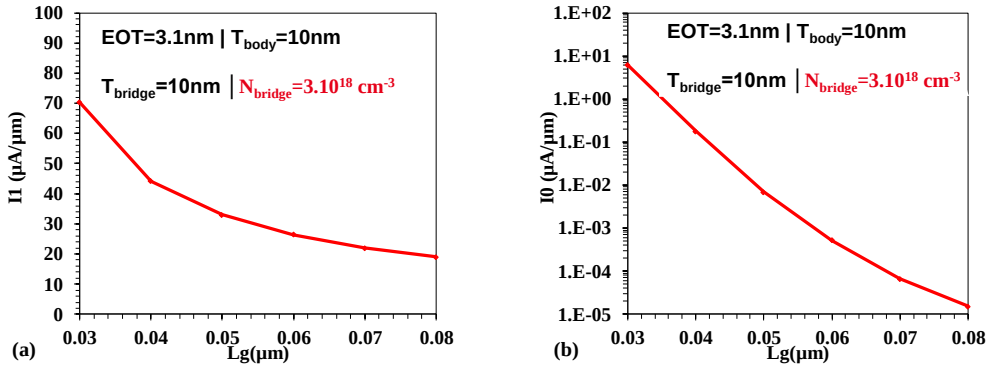


Figure 3.23 : Evolution de (a) I_1 et (b) I_0 en fonction de la longueur de la grille L_g , pour $N_{\text{bridge}} = 3 \cdot 10^{18}$ extraits des simulations TCAD.

Notons que sur les figures 3.22.a et 3.22.b, si N_{bridge} est très élevé ($\geq 3 \cdot 10^{18} \text{ cm}^{-3}$), le bridge reste toujours en conduction et maintient un court-circuit entre la source et le drain. L'A2RAM est toujours à l'état activé et n'est plus fonctionnelle ($I_1 = I_0$ pour $N_{\text{bridge}} = 3 \cdot 10^{18} \text{ cm}^{-3}$). De nos études, nous avons observé que pour un dopage de bridge $N_{\text{bridge}} \geq 3 \cdot 10^{18} \text{ cm}^{-3}$, il faut réduire l'épaisseur du film de silicium (body+bridge), ce qui a pour effet

CHAPITRE III : OPTIMISATION DE LA CELLULE A2RAM PAR UNE ETUDE DE SENSIBILITE TCAD

d'améliorer le contrôle du bridge par la grille face avant (le champ électrique vertical augmente). Sur les figures 3.23.(a, b), nous reportons les courants I1 et I0 en fonction de différentes longueurs de grille avec $T_{\text{body}} = T_{\text{bridge}} = 10 \text{ nm}$. Nous constatons qu'à présent il est possible de trouver des structures fonctionnelles. Notamment, pour des longueurs de grille $L_g \geq 40 \text{ nm}$ nous avons un courant I1 élevée avec un faible courant I0.

Cette étude montre une fois de plus que les performances de l'A2RAM sont très sensibles aux variations du dopage du bridge. Sur la figure 3.22, on pourrait conclure qu'un dopage de bridge supérieur à 10^{18} cm^{-3} ne garantit plus une fonctionnalité mémoire A2RAM, mais avec un bon jeu de paramètres d'épaisseurs du body et bridge, nous notons sur la figure 3.23 qu'il est possible de trouver des structures A2RAM fonctionnelles.

Réduction de :	I1	I0	Temps de rétention
L_g	↑ (bridge court & écriture améliorée par SCE)	↑ (SCE ↑, donc les courants de fuites ↑)	↓ (car SCE ↑, les courants de fuites ↑)
EOT	Très peu d'impact sur les longueurs de grille L_g courtes	↓ car amélioration du contrôle électrostatique	↑ car amélioration du contrôle électrostatique
T_{body}	↓ car le réservoir de stockage est réduit.	↓ car amélioration du contrôle électrostatique (le film de silicium est réduit).	↑ car amélioration du contrôle électrostatique
T_{bridge}	↓ car la résistivité du bridge augmente.	↓ car amélioration du contrôle électrostatique (le film de silicium est réduit).	~ très peu de courant de fuite, mais I1 est très faible.
N_{bridge}	↓ car la résistivité du bridge augmente.	↓ à cause du meilleur contrôle électrostatique favorisé par un faible dopage dans le bridge.	↑ grâce au meilleur contrôle électrostatique du bridge par la grille.

Tableau 3.1. Compromis liés à la miniaturisation de l'A2RAM

Le tableau 3.1 résume et quantifie l'effet des principaux paramètres technologiques sur I1, I0 et le temps de rétention. A partir de cette étude, nous proposons ci-dessous plusieurs structures optimisées, avec différents compromis entre I1 et I0. En effet, certaines applications nécessitent en priorité un fort I1 alors que d'autres privilégient un temps de rétention élevé.

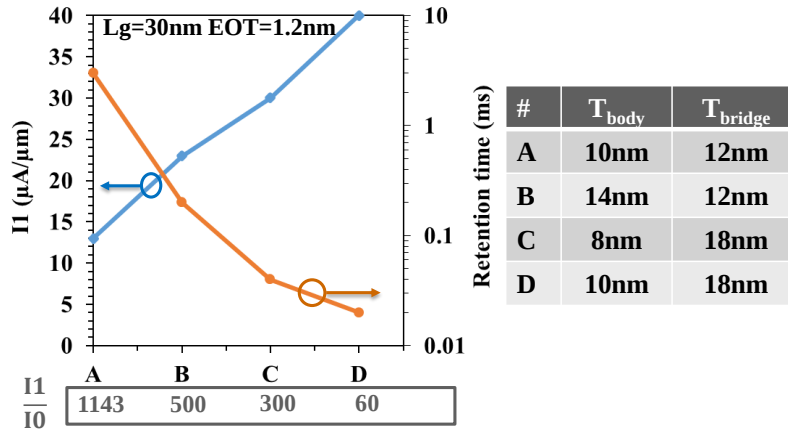


Figure 3.24 : Valeur du I_1 (sur l'axe de gauche) et temps de rétention (sur l'axe de droite) pour les structures A2RAM : A, B, C et D dont les épaisseurs du body et bridge sont données dans le tableau à droite.

La figure 3.24 présente des cellules A2RAM (A, B, C, et D) dont la longueur de grille est de 30 nm (pour assurer une densité d'intégration élevée). Les paramètres importants de ces structures sont reportés dans le tableau à droite de la figure 3.24. Ces structures ont été choisies tels que : $I_1/I_0 > 40$ et $I_1 - I_0 = 6 \mu\text{A}.\mu\text{m}^{-1}$ afin d'assurer un bon fonctionnement une fois sous matrice. Nous constatons sur la figure 3.24 que le meilleur temps de rétention est garanti pour la structure A qui a une épaisseur totale du film de silicium (body + bridge) la plus faible (22 nm).

III.3. D'autres voies d'optimisation de performances de l'A2RAM

III.3.1. Intégration d'une mono couche de bore pur dans une cellule A2RAM : la structure A2RAM B++

A la fin de la section 2, nous avons proposé des structures optimisées d'A2RAM. Nous avons constaté que le temps de rétention augmente lorsque l'épaisseur totale du film de silicium ($T_{\text{si}} = T_{\text{body}} + T_{\text{bridge}}$) diminue, mais il s'associe à une réduction de la valeur de I_1 . Pour augmenter I_1 en maintenant constante ou en réduisant T_{si} , il est aussi possible d'augmenter le dopage du bridge comme nous l'avions mentionnée en fin de la section 2.5 (figure 4.23). Cependant les valeurs de dopage mis en jeu étant très élevées au regard l'épaisseur totale du film silicium visé ($T_{\text{si}} \leq 20 \text{ nm}$), durant la phase de recuit (activation des dopants du bridge), il sera difficile de contrôler la diffusion de la charge majoritaire du bridge dans le body. Comme conséquence, au lieu d'avoir un body intrinsèque, nous nous retrouverons avec un transistor sans jonction. Mais, une solution serait d'utiliser une monocouche de bore pur dans le film de silicium. La technologie de monocouche de bore pur a démontré sa faisabilité dans diverses applications comme le souligne [Nanver 10].

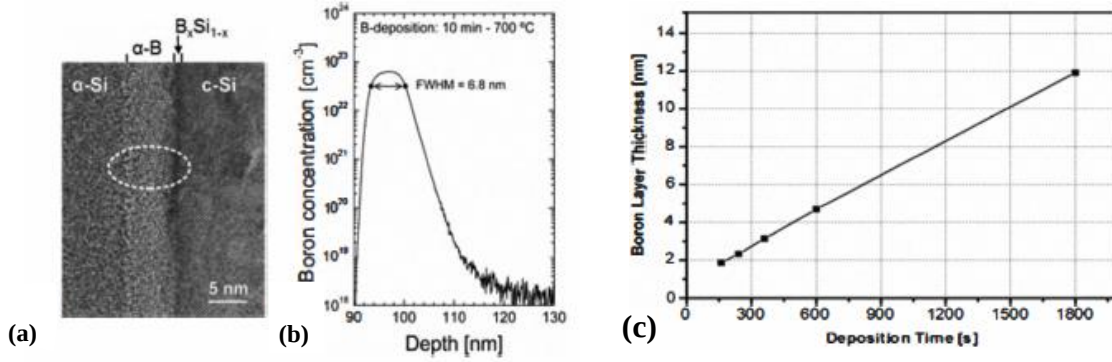


Figure 3.25 : (a) Image TEM à haute résolution et (b) Profil SIMS (du faisceau d'ions primaires à 1 keV) d'une couche B telle que déposée sur une surface de (100) de Si à 700°C après 10 min d'exposition au B_2H_6 [Nanver 10]. (c) Epaisseur de la couche de bore (mesurée par ellipsométrie) en fonction du temps.

Dans son travail, [Nanver 10] démontre qu'on peut obtenir une couche de bore pure sur du silicium par un dépôt chimique en phase vapeur (CVD : chemical-vapor deposition). Les figures 3.25.(a, b), reproduisent une image TEM de la structure et le profil de dopage du bore obtenus. La couche de bore pur se trouve entre une couche de silicium amorphe (α -Si) et une couche de silicium cristalline (c-Si). Dans ce même travail, ils atteignent une couche de bore pure de concentration $10^{23} cm^{-3}$ pour une épaisseur de bore déposée d'environ 7 nm, mais il est possible de faire croître des épaisseurs avoisinant 2 nm, comme le montre la figure 3.25.c où nous avons l'épaisseur de couche de bore en fonction du temps de dépôt CVD. Dans ce travail, nous démontrons par des simulations TCAD qu'une monocouche de bore pur en bas du film de silicium peut améliorer les performances mémoires de l'A2RAM.

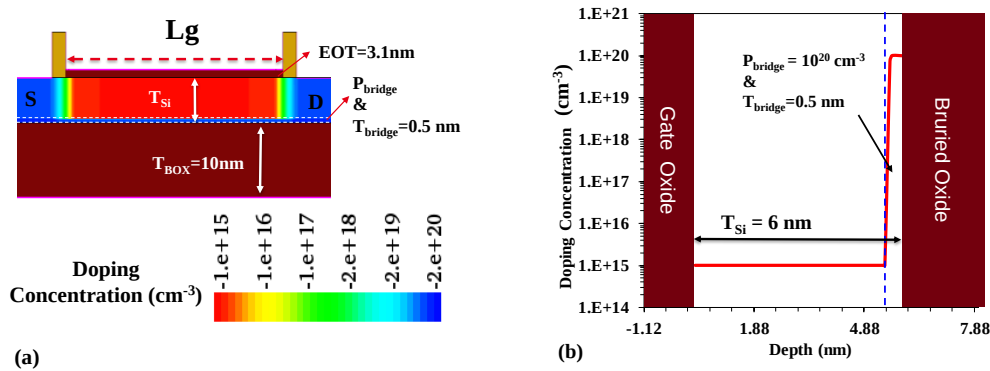


Figure 3.26 : (a) Cellule A2RAM B++ simulée (avec ses paramètres par défaut), EOT = 3.1 nm et une épaisseur de bridge T_{bridge} fixe de 0.5 nm. (b) Exemple d'un profil de dopage vertical d'une A2RAM avec une monocouche de bore pur dans le bridge.

L'architecture d'une A2RAM avec une monocouche de bore pur (A2RAM B++) est illustrée sur la figure 3.26.a. Cette structure ressemble à une cellule A2RAM de type p ; avec un bridge

mince ($T_{\text{bridge}} = 0.5 \text{ nm}$ et $N_{\text{bridge}} = 10^{20} \text{ cm}^{-3}$) qui va émuler la monocouche de bore pur. La figure 3.26.b montre le profil de dopage très abrupte le long du film de silicium.

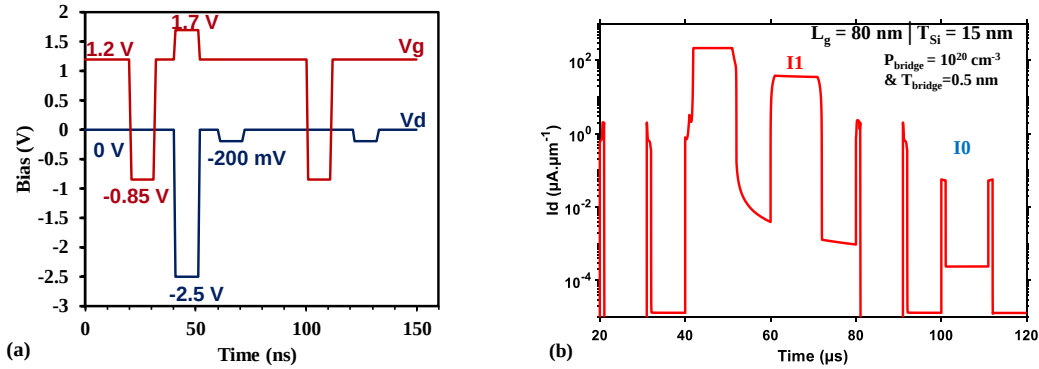


Figure 3.27 : (a) Motif utilisé pour la caractérisation mémoire de la structure A2RAM B++ ; similaire à celui de la figure 2.2, mais les polarités sont inversées. (b) Chronogramme de courant du drain correspondant.

En utilisant la méthodologie d'évaluation de performance de la section 1 du Chapitre 1, mais cette fois avec le chronogramme de la figure 3.27.a, les performances de l'A2RAM simulée avec une monocouche de bore comme bridge vont être évaluées. Vérifions, dans un premier temps, la fonctionnalité de la structure avec une A2RAM (figure 3.26.a) dont l'épaisseur totale de film de silicium est $t_{\text{Si}} = 15 \text{ nm}$. Sur la figure 3.27.a, nous avons le chronogramme de courant de drain obtenu lorsqu'on applique le motif de tensions sur la structure A2RAM B++. Nous observons que le courant en écriture sature démontrant ainsi que la charge maximale est atteinte dans le body. C'est pourquoi il est possible d'avoir un courant I1 élevé en lecture ($\approx 30 \mu\text{A} \cdot \mu\text{m}^{-1}$). L'effacement se fait toujours par couplage capacitif entre les grilles face avant et arrière, et ici, nous pouvons considérer que nous avons effectivement évacué la charge stockée dans le body, car $I0 \neq I1$.

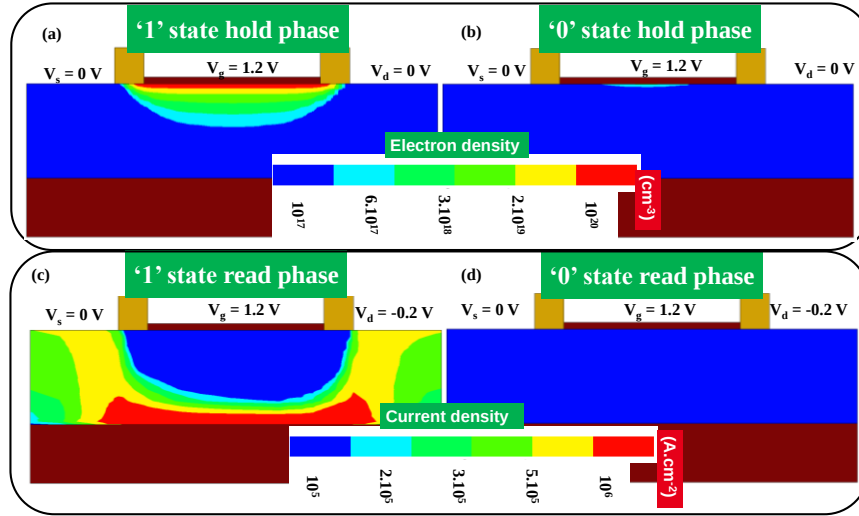


Figure 3.28 : Cartographies de la densité électrons pendant la phase de maintien après (a) la phase d'écriture de l'état '1', et (b) la phase d'écriture de l'état '0'. Cartographies de la densité de courant pendant la phase de lecture de (c) l'état '1' et (d) de l'état '0'. Toutes sont extraites de TCAD lors de la caractérisation de la mémoire (en utilisant le motif présenté sur la figure 3.27.a) de la figure 3.26 mais avec $T_{Si} = 15 \text{ nm}$.

Les figures 3.28.a et 3.28.b montrent les cartographies de la densité d'électrons lors de la phase de maintien de l'état '1' et de l'état '0' respectivement. Lors du maintien de l'état '1', nous notons la présence des porteurs majoritaires dans le body, ici les électrons, ceci n'est pas le cas pour l'état '0' où aucune charge n'est présente dans le body. Comme pour une A2RAM classique, ceci explique la forte densité de courant qui circule à travers la monocouche de bore (figures 3.28.c et 3.28.d) uniquement pendant la lecture de l'état '1'. La fonctionnalité de l'A2RAM B ++ est démontrée et son fonctionnement est donc similaire à celui d'une cellule standard A2RAM.

Après avoir validé l'effet mémoire dans une A2RAM B ++, nous avons fait des études de sensibilité des performances lorsqu'on varie T_{Si} et P_{bridge} . Partant d'une structure ayant une grille de longueur $L_g = 80 \text{ nm}$, nous faisons varier l'épaisseur totale de silicium T_{Si} ainsi que le dopage du bridge P_{bridge} . Nous reportons les variations de I_1 , I_0 et du ratio I_1/I_0 en fonction du dopage du bridge sur la figure 3.29.

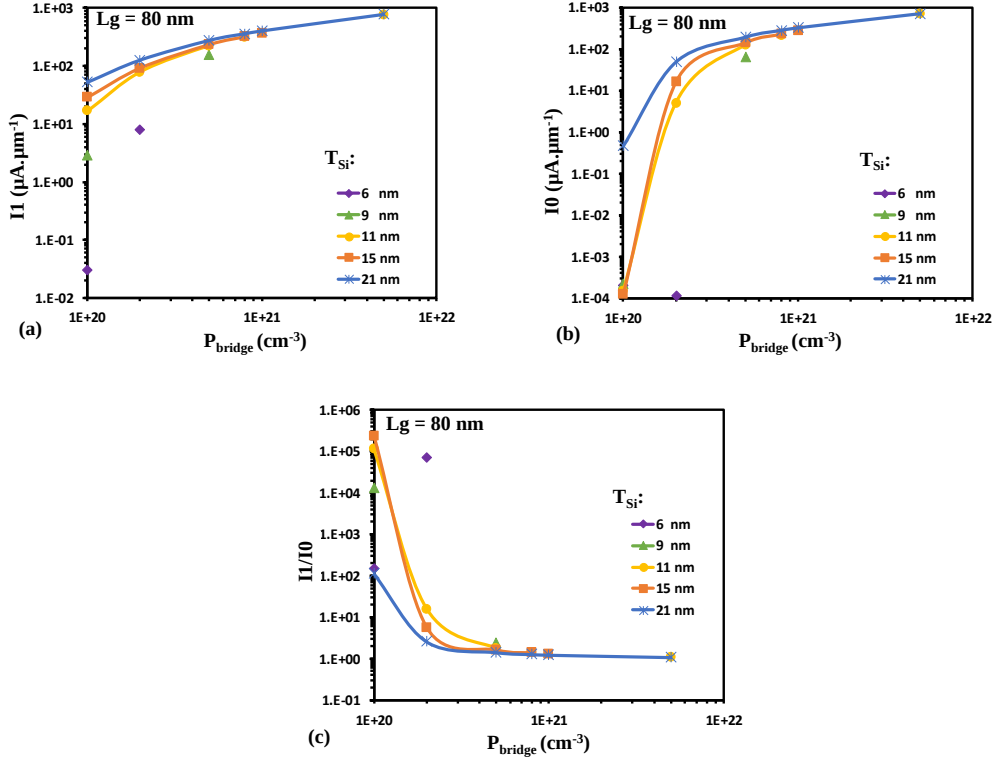


Figure 3.29 : (a) Courant lu d'état '1' I_1 , (b) courant lu d'état '0' I_0 et (c) rapport I_1/I_0 en fonction de P_{bridge} et T_{Si} , extraits de TCAD.

Sur les figures 3.29.a et 3.29.b, les variations de I_1 et I_0 sont totalement similaires à celles observées aux sections 2.3, 2.4 et 2.5 pour l'A2RAM de type n. Nous pouvons noter qu'une concentration $P_{bridge} > 2.10^{20} cm^{-3}$ ne permet pas d'avoir un effet mémoire dans une structure A2RAM B++ même si nous réduisons l'épaisseur du film de silicium jusqu'à 6 nm, car $I_0 = I_1$. Cette observation est confirmée par la figure 3.29.c, où nous avons l'évolution de la fenêtre de programmation I_1/I_0 . Il est préférable de viser $T_{Si} \leq 21$ nm et $P_{bridge} \leq 2.10^{20} cm^{-3}$ si nous voulons avoir des structures fonctionnelles ($I_1/I_0 > 1$).

A partir de cette succincte étude de sensibilité, nous allons optimiser l'architecture A2RAM B++. En ciblant toujours une forte densité d'intégration des cellules A2RAM B++, nous avons examiné des structures avec une longueur de grille L_g de 30 nm en faisant varier l'épaisseur du film entre 21 nm et 6 nm et le dopage du bridge de $5.10^{20} cm^{-3}$ à $5.10^{18} cm^{-3}$.

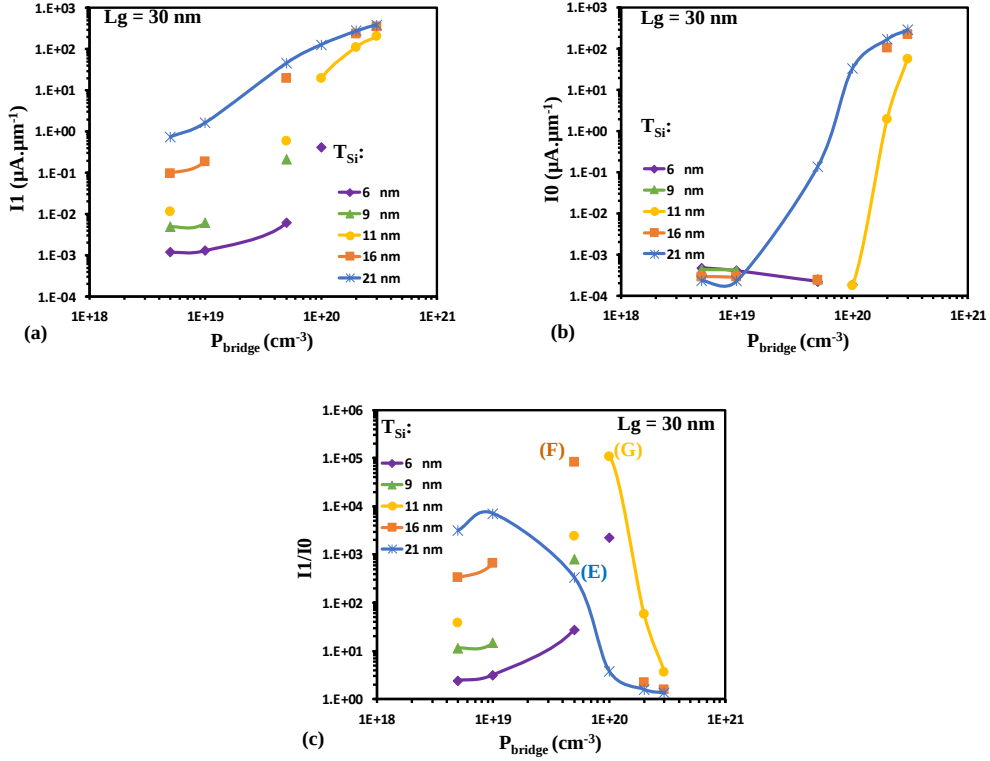


Figure 3.30 : (a) Courant lu d'état '1' I_1 , (b) courant lu d'état '0' I_0 et (c) rapport I_1/I_0 en fonction de P_{bridge} et T_{Si} , extraits de TCAD.

Sur les figures 3.30.a et 3.30.b, les variations de I_1 et I_0 sont données en fonction du dopage du bridge pour différentes épaisseurs du bridge à $L_g = 30$ nm. Comparées aux observations faites sur les figures 3.29.(a, b, c), il est possible ici d'avoir une fonctionnalité mémoire pour des dopages de bridge supérieurs à $2.10^{20} \text{ cm}^{-3}$, car $I_0 \neq I_1$. Par contre, si le dopage du bridge est inférieur à $5.10^{19} \text{ cm}^{-3}$, le courant I_1 converge de plus en plus vers I_0 . Ceci est dû à l'augmentation de résistivité du bridge lorsqu'on réduit son dopage comme nous l'avons étudié dans la section 2.5.

Ainsi, sur la figure 3.30.c, où nous avons l'évolution de la fenêtre de programmation I_1/I_0 en fonction du dopage bridge pour différentes épaisseurs du bridge, les structures A2RAM B++ qui garantissent $I_1/I_0 > 40$ et $I_1 - I_0 \geq 6 \mu\text{A}\cdot\mu\text{m}^{-1}$ sont définies pour $T_{\text{Si}} > 9$ nm et $10^{19} \text{ cm}^{-3} < P_{\text{bridge}} \leq 5.10^{20} \text{ cm}^{-3}$. En utilisant ces critères comme garants de la fonctionnalité de la cellule A2RAM B++ sous matrice, nous pouvons retenir les trois structures suivantes : E ($T_{\text{Si}} = 21$ nm et $P_{\text{bridge}} = 5.10^{20} \text{ cm}^{-3}$), F ($T_{\text{Si}} = 16$ nm et $P_{\text{bridge}} = 5.10^{20} \text{ cm}^{-3}$) et G ($T_{\text{Si}} = 11$ nm et $P_{\text{bridge}} = 10^{20} \text{ cm}^{-3}$) comme indiqué sur la figure 3.30.c. Pour finir, nous allons évaluer les temps de rétention de ces structures et les comparer à la structure A2RAM de type n ayant obtenu le meilleur temps de rétention (structure [A], de la figure 3.24).

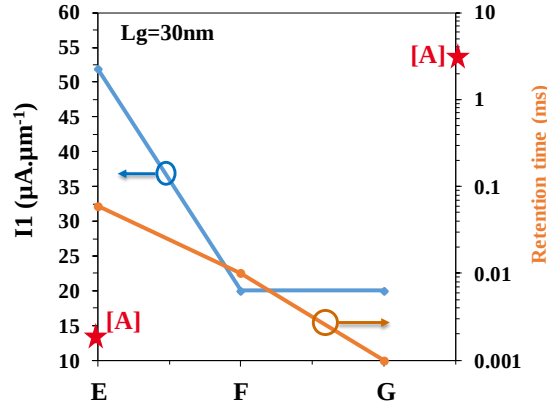


Figure 3.31 : Le compromis entre le courant I_1 (axe de gauche) et le temps de rétention (axe de droite) pour les cas E, F et G, par rapport à la structure optimisée ayant le meilleur de temps de rétention ([A] de la figure 3.24).

La figure 3.31 compare I_1 et le temps de rétention des structure E, F et G à la structure [A] optimisée de la section 2. Avec une A2RAM B ++, le temps de rétention et I_1 augmentent lorsque T_{Si} augmente, mais la valeur du temps de rétention reste très faible pour envisager une fonctionnalité sous matrice comparée à une A2RAM standard. Mais, ce travail nous a permis de mettre évidence la possibilité d'utiliser la technologie de bore pure pour améliorer d'autres performances de l'A2RAM.

III.3.2. Hétérostructures et performances mémoires de l'A2RAM : cas du SiGe

Dans une 1T-DRAM, la quantité de charge pouvant être stockée dans le film de silicium est limitée par l'auto-polarisation du film de silicium : dans le cas de l'A2RAM, c'est la jonction source/body qui passe en direct. Une solution pour augmenter la barrière de potentiel entre la source et body pourrait être l'intégration des matériaux à faible gap dans le body [Bawedin 11]. [Ertosun 08] et [Shin 11] relèvent notamment la possibilité d'intégrer des matériaux III-V dans le film de silicium pour améliorer le stockage de la charge. Comme on peut le voir sur la figure 3.32.a, [Ertosun 08] montre la différence entre le diagramme de bandes d'une structure avec un body en silicium et une autre avec du SiGe intégré dans une partie du body. L'avantage est le puit de potentiel crée par la présence du SiGe. Ce puit de potentiel sert de réservoir pour le stockage des trous, ce qui augmentera l'efficacité du stockage. De même, [Shin 11] intègre le SiGe cette fois dans la totalité du body comme sur la figure 3.32.b, nous avons la structure 2D, et la comparaison des diagrammes à bandes d'énergie avec et sans SiGe dans le body. On peut noter que la différence entre les bandes de valence entre la source/drain et le body augmente avec du SiGe.

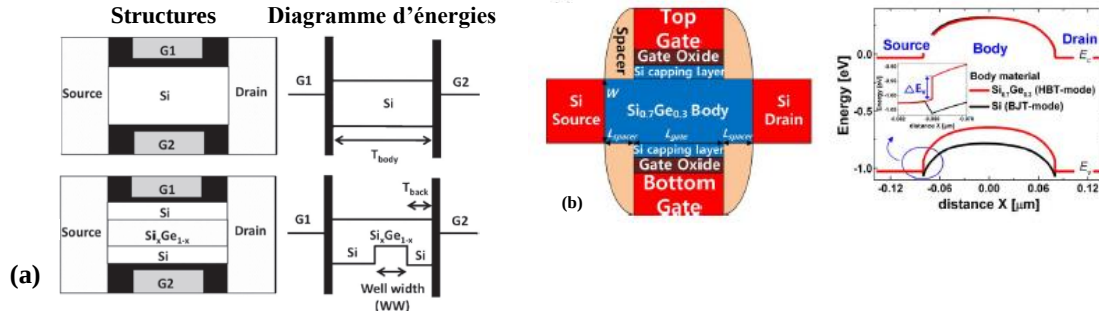


Figure 3.32 : (a) Schémas d'une 1T-DRAM à double grille totalement silicium et de la 1T-DRAM à double grille avec puit quantique de SiGe dans le body et leurs diagrammes de bande d'énergies correspondants. (b) Vue de dessus en coupe de la cellule 1T-DRAM et son diagramme de bande d'énergies le long de la jonction source-body-drain à l'équilibre thermodynamique avec du SiGe ou seulement du Si dans le body.

S'inspirant de ces différents travaux, le but de ce paragraphe est d'évaluer si l'introduction d'un matériau III-V dans l'architecture A2RAM notamment dans le body peut améliorer ses performances. Il conviendra ensuite de déterminer la concentration de germanium optimale à utiliser. Cette étude ayant servi de base à la fabrication de dispositifs A2RAM au CEA-Leti, nous nous sommes focalisés uniquement sur l'intégration du SiGe. Car c'est le seul matériau semiconducteur avec un gap plus faible que le silicium disponible dans les salles blanches du CEA-Leti, et compatible avec les procédés standards de fabrication CMOS.

Dans un premier temps, nous allons utiliser la structure de la figure 2.1. Pour déterminer dans quelle région du film de silicium il est le plus avantageux d'utiliser le SiGe, nous avons remplacé le silicium par du silicium-germanium dans le body uniquement (figure 3.33.a) puis dans le bridge uniquement (figure 3.33.b). Pour une première simulation, considérons une fraction molaire du germanium $x_{Ge} = 20\%$. Les performances mémoires sont étudiées en utilisant le motif de la figure 2.2 (avec une durée des impulsions de 10 ns et des temps de montée et de descente 1 ns).

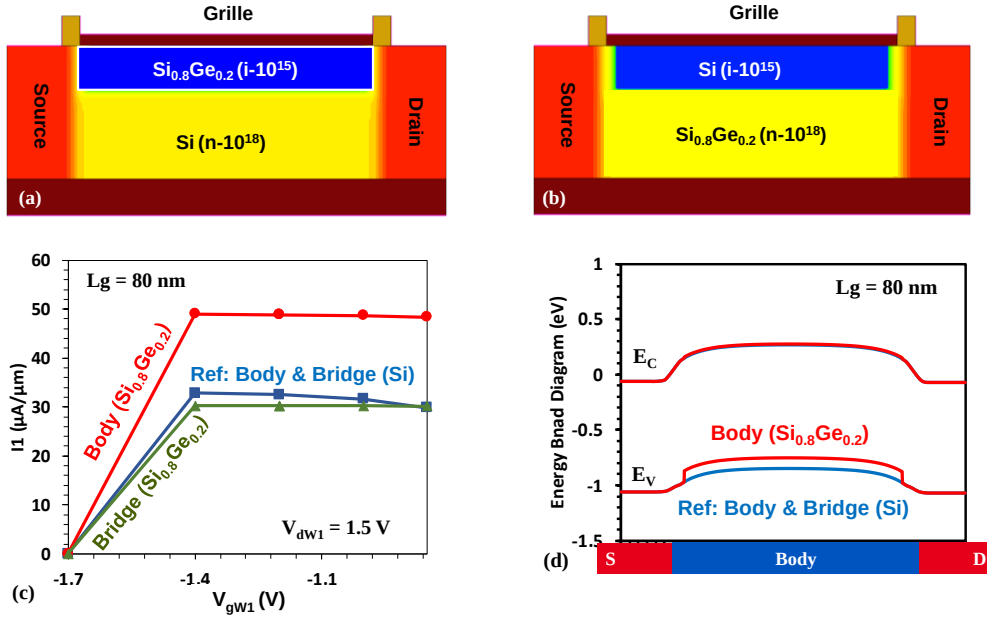


Figure 3.33 : (a) I_1 en fonction de V_{gW1} pour $V_{dW1} = 1.5$ V, pour une A2RAM en silicium (courbe bleue), pour une A2RAM avec body SiGe et bridge silicium (courbe rouge) et pour une A2RAM avec body silicium et bridge SiGe (courbe verte). (d) Diagramme de bande d'énergie le long de la source-body-drain pour une A2RAM avec body SiGe et l'A2RAM de référence en silicium.

La figure 3.33.c, montre les variations de I_1 en fonction de la tension de grille appliquée pendant la programmation de l'état '1' pour les 3 structures : le SiGe dans le body (courbe rouge), SiGe dans le bridge (courbe verte), et le cas de référence de l'A2RAM silicium (courbe bleu).

- Body Si/Bridge SiGe: lorsque le SiGe est intégré dans le bridge, comparé à la structure de référence nous observons une légère dégradation sur la valeur de I_1 . Ayant le même matériau dans le body que la structure de référence, cette différence n'est pas liée à l'efficacité du stockage, mais à une différence au niveau de la lecture. Une simulation de type R carré qui permet d'évaluer la résistivité d'un matériau, nous a permis de conclure que cette différence est due à la résistivité élevée du bridge lorsque nous y intégrons du SiGe.
- Body SiGe/Bridge Si : une forte valeur de I_1 est obtenue lorsqu'on intègre le SiGe dans le body. Nous pouvons expliquer ce résultat par le fait que le confinement des trous dans le body est plus efficace avec du SiGe, grâce à l'augmentation de la barrière de potentiel des trous entre la source/drain et le body. Sur la figure 3.33.d, nous avons fait une coupe à 1 nm en dessous de l'oxyde de la grille face avant et extrait les diagrammes de bandes d'énergie entre la source-body-drain. Nous pouvons noter que la barrière d'énergie des trous dans le body est supérieure lorsqu'on y intègre du SiGe ; d'où un meilleur stockage de trous.

Les résultats obtenus sur la figure 3.33 nous permettent de conclure qu'il est intéressant d'intégrer du SiGe dans le body d'une A2RAM. Cependant cette étude ne fournit pas la concentration optimale du Germanium qu'il faut dans le body pour garantir un meilleur stockage. Ainsi, nous avons repris la structure de la figure 3.33.a et refait les simulations de la figure 3.33.c en variant la concentration x_{Ge} du germanium dans le body. Nous avons limité cette étude à une concentration de x_{Ge} maximale de 0.5.

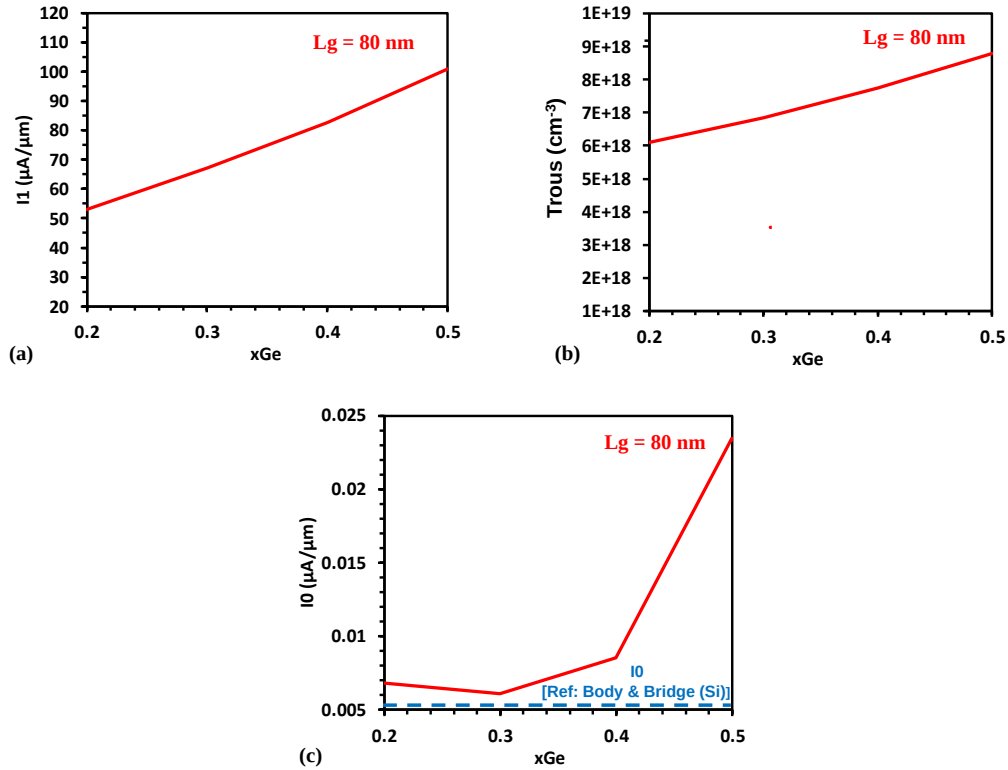


Figure 3.34 : Evolution des courants I_1 et I_0 durant une lecture continue pour un Overlap de (a) 0 nm, (b) 2 nm, (b) 6 nm, (c) 10 nm, en variant x_{Ge} dans le body.

- Sur la figure 3.34.a, nous avons les variations de I_1 en fonction de la concentration x_{Ge} du germanium dans le body. I_1 augmente linéairement avec x_{Ge} , car augmenter x_{Ge} réduit l'énergie de bande interdite dans le body, par conséquent la barrière d'énergie de valence entre la source et le body augmente, d'où un meilleur stockage. Nous avons intégré la concentration de trous stockés dans tout le volume du body pour chaque structure évaluée sur la figure 3.34.a. La figure 3.34.b montre l'évolution de la concentration de trous stockés dans le body en fonction de x_{Ge} ; nous observons que la concentration de trous augmente lorsqu'on augmente x_{Ge} , d'où le fort courant I_1 .
- Sur la figure 3.34.c, les variations de I_0 sont présentées en fonction de x_{Ge} . Nous avons aussi indiqué sur cette figure la position du courant I_0 obtenu sur la cellule de référence en silicium. Pour $x_{Ge} \leq 0.3$, I_0 reste quasiment constant, ce qui revient à dire qu'intégrer une concentration de germanium de 0.2 ou 0.3 n'a pas d'impact significatif sur les fuites de l'A2RAM. Mais au-delà de 0.3, I_0 augmente brusquement ;

cette augmentation est due au courant de fuite des diodes source/body et drain/body qui sont de moins en moins contrôlées par la grille face avant. Par conséquent, si nous voulons garder un bon contrôle électrostatique tout en utilisant un matériau à faible gap dans le body il est préférable d'avoir une concentration $x_{Ge} \leq 0.3$, et pour maximiser la valeur de I1 il faut que $x_{Ge} = 0.3$.

Nous avons vérifié que l'intégration du SiGe dans le body peut améliorer ses performances mémoires. Dans la suite, nous allons évaluer les performances de l'A2RAM en fonction de la position du SiGe dans le body. Comme indiqué sur la figure 3.35.a, nous faisons varier la position où on doit intégrer le SiGe dans le body. Partant d'une position alignée avec les espaceurs (overlap = 0.0 nm), la position (overlap) du SiGe sera de plus en plus recentrée dans le body. Le but est de vérifier s'il peut y avoir une amélioration sur le courant de fuite des diodes parasites source/body et drain/body de l'A2RAM, sans toutefois dégrader la valeur de I1.

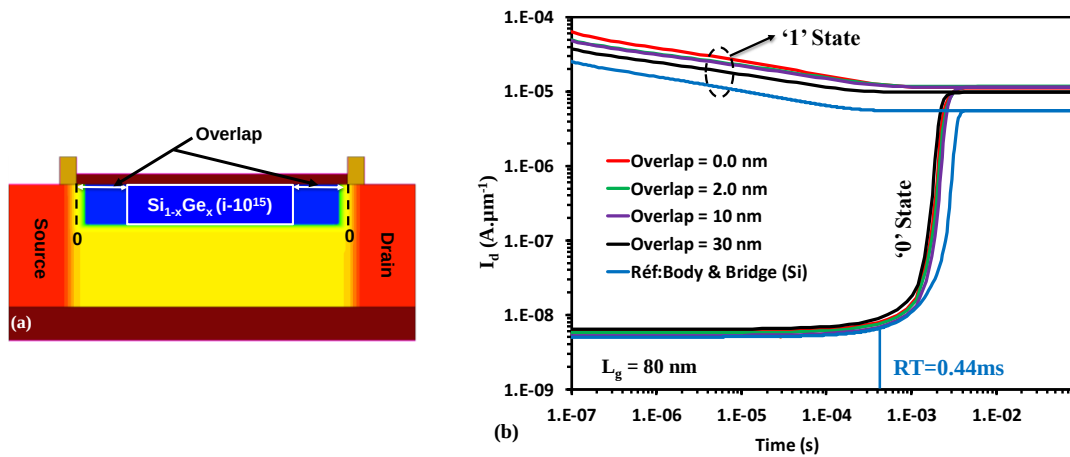


Figure 3.35 : Evolution du des courants I1 et I0 durant une lecture prolongée pour un Overlap de (a) 0 nm, (b) 2 nm, (b) 6 nm, (c) 10 nm, en variant x_{Ge} dans le body.

Sur la figure 3.35.b, nous présentons l'évolution des courants de drain durant une lecture prolongée des états '1' et '0' évalués en utilisant le protocole de la section 2.4 du Chapitre 2. Pour une concentration $x_{Ge} = 0.3$ du germanium et différentes positions dans le body, nous comparons le maintien de l'information ('1' et '0') avec celui de la cellule de référence.

Pour les variations de I1, nous observons que recentrer la position de la couche de SiGe dans le body dégrade la valeur de I1, car le réservoir de trous est réduit. C'est pourquoi, pour un overlap supérieur à 10 nm, I1 se rapproche de plus en plus de la valeur de la structure avec seulement du silicium dans le body.

Les variations de I0 montrent une légère amélioration : une réduction de courant de fuite de l'A2RAM. Les structures avec SiGe dans le body ont l'avantage d'avoir un courant I1 élevé, bien que le courant I0 soit légèrement supérieur à celui de la structure en silicium. C'est pourquoi l'évaluation du temps de rétention des toutes ces structures s'équivalent (0.44ms).

III.4. Conclusion

Dans ce chapitre, nous avons cherché à améliorer les performances de la cellule mémoire A2RAM en optimisant son architecture via l'outil de simulation commercial TCAD : Synopsys. Une étude approfondie de l'écriture de l'état '1', nous a permis de mettre en évidence un nouveau mécanisme d'écriture de l'état '1' qui a fait l'objet d'un dépôt de brevet. Grâce à cette technique, il est possible d'envisager l'écriture de l'état '1' en 10 ns avec une tension sur le drain supérieure ou égale à 1 V et une tension sur la grille inférieure ou égale à -0.5 V. Notons que les limitations du test sous pointe n'ont pas permis de mettre en évidence ce nouveau mécanisme expérimentalement. A partir d'une analyse de sensibilité des performances de l'A2RAM à ses principaux paramètres technologiques durant les différentes opérations mémoires, nous avons proposé des structures A2RAM optimisées qui garantissent une fonctionnalité sous matrice et une densité d'intégration élevée. De plus, nous avons proposé une structure A2RAM optimisée composée de monocouche de bore pur (A2RAM B ++), qui améliore le compromis performance/densité d'intégration. Enfin, l'utilisation d'un matériau semiconducteur à faible gap a été évaluée : l'intégration du SiGe dans le body de l'A2RAM peut améliorer ses performances.

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

Dans ce chapitre, nous allons aborder la question du profil de dopage dans une structure A2RAM. Comme nous l'avions mentionné dans la section 1 du Chapitre 2, l'une des raisons expliquant la différence entre mesure expérimentale et simulation serait la différence des profils de dopage le long du film de silicium (body + bridge). De plus, l'étude de sensibilité de l'A2RAM présentée dans la section 2 du Chapitre 3 démontre que le profil de dopage du bridge a un impact majeur sur les performances mémoires. La caractérisation du bridge, c'est-à-dire connaître son épaisseur et son niveau de dopage est par conséquent nécessaire. Les méthodes de caractérisations physiques telles que la spectroscopie de masse d'ions secondaires (SIMS : Secondary-Ion Mass Spectroscopy) ou dark holographie ne sont pas utilisables ici car, le film de silicium est trop mince et le niveau de dopage trop faible (environ 10^{18} cm^{-3}). Pour la dark holographie, pour le moment, la technique a été démontrée pour des dopages allant de $4 \cdot 10^{17} \text{ cm}^{-3}$ à 10^{19} cm^{-3} , mais avec des épaisseurs du silicium supérieures à une centaine de nanomètres [Cooper 16]. Par conséquent, une méthode basée sur la mesure électrique est l'option la plus appropriée. Pour cette étude, nous avons utilisé la caractéristique électrique décrivant les variations de la capacité de la grille face avant en fonction de la polarisation de la grille face avant (C_{gg} - V_g). La méthodologie d'extraction du profil de dopage du bridge sera développée par simulation TCAD et sera validée à l'aide de données expérimentales. Nous finirons en présentant des suggestions pour améliorer les performances mémoires.

IV.1. Technique d'extraction des paramètres technologiques de l'A2RAM

Nous allons dans un premier temps démontrer la faisabilité de la technique d'extraction du profil de dopage dans le film de silicium de l'A2RAM par simulation TCAD. Pour cela, nous allons exploiter la caractéristique électrique décrivant les variations de la capacité de la grille en fonction de la polarisation de la grille face avant, afin d'extraire la valeur du dopage du bridge et les épaisseurs des différentes parties constituant le film de silicium de l'A2RAM.

IV.1.1. Étude de la caractéristique C_{gg} - V_g de la cellule A2RAM

La structure A2RAM simulée est celle de la figure 2.1 (du Chapitre 2). Ses principaux paramètres technologiques sont : l'épaisseur du body $T_{\text{body}} = 14 \text{ nm}$ (de type p, $N_{\text{body}} = 10^{15} \text{ cm}^{-3}$), l'épaisseur du bridge $T_{\text{bridge}} = 22 \text{ nm}$ (de type n, $N_{\text{bridge}} = 10^{18} \text{ cm}^{-3}$), l'épaisseur d'oxyde équivalente de grille face avant $EOT = 3.1 \text{ nm}$, et épaisseur d'oxyde enterré (BOX) $T_{\text{BOX}} = 10 \text{ nm}$.

Nous allons prendre pour $L_g = 1 \mu\text{m}$ pour nos simulations, afin d'éviter que la caractéristique électrique C_{gg} - V_g soit biaisée par les effets de canaux courts et donc ainsi ne considérer que les effets 1D. La méthodologie de simulation pour l'extraction de la capacité de la grille est présentée à la figure 4.1. La grille face avant est connectée à une source de tension continue (DC) V_g mixée à une source petit signal alternatif (AC) et les source et drain sont à la masse. La fréquence du signal alternatif est fixée à 1 MHz pour s'aligner avec le choix fait pour les mesures expérimentales. La source et le drain sont court-circuités et connectés à la masse. La grille face arrière est connectée sur une source de tension continue (DC) V_b indépendante, et sera utilisée pour modifier le régime de l'interface arrière entre le bridge et l'oxyde enterré.

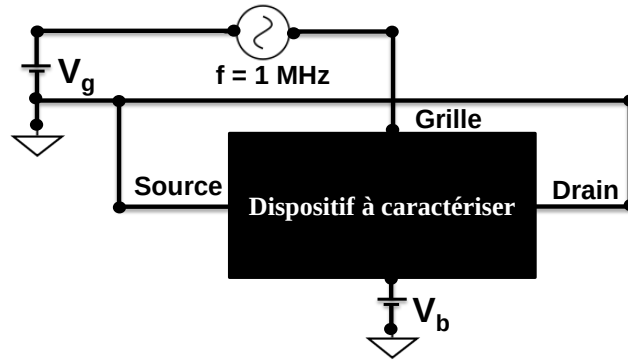


Figure 4.1 : Protocole utilisé pour la caractérisation électrique de la capacité de la grille face avant de l'A2RAM.

Afin d'analyser le comportement capacitif de l'A2RAM et notamment identifier l'impact du bridge, nous traçons la variation de la capacité grille en fonction de la polarisation de grille pour l'A2RAM décrite ci-dessus mais aussi pour la même structure sans bridge : c'est-à-dire un transistor SOI "standard" de même épaisseur que l'A2RAM (figure 4.2.a) avec un film de silicium non dopé (figure 4.2.b).

Sur la figure 4.2.c, nous reportons les caractéristiques des variations de la capacité de la grille face avant en fonction de la polarisation de la grille face avant ($C_{gg}-V_g$). On remarque d'abord que pour des polarisations élevées sur la grille face avant $V_g \geq 0.6$ V, les caractéristiques de l'A2RAM et du transistor SOI sont superposées. A ces polarisations, les 2 deux dispositifs sont en régime de forte inversion et on peut donc conclure que ce régime n'est pas impacté par le bridge de l'A2RAM. En effet, la couche d'inversion face avant écran toutes variations de charge qui pourrait se produire dans le body. Pour des polarisations de grille plus faibles, la face avant est donc en régime de déplétion et on remarque que les courbes $C_{gg}-V_g$ des dispositifs ne sont plus superposées. Pour le transistor SOI, le film de silicium est rapidement complètement déplété : pour $V_g < 0.3$ V, la capacité est constante et reste à sa valeur minimale. A contrario, pour l'A2RAM, la capacité diminue entre 0.3 V et - 0.6 V jusqu'à atteindre une valeur minimale correspondant à la désertion complète du film de silicium incluant donc la désertion du bridge. Ce comportement s'explique par la conduction du bridge, qui commence à être affectée par le petit signal AC autour de $V_g = -0.6$ V.

De cette description, il est alors possible de définir pour l'A2RAM deux tensions de seuil dues à la formation de ces deux canaux de conduction. Pour mettre en évidence cette observation, nous avons évalué les dérivées premières des caractéristiques $C_{gg}-V_g$ de la figure 4.2.c. Les caractéristiques obtenues sont présentées sur la figure 4.2.d. Pour l'A2RAM, lorsque la tension appliquée sur la grille face avant augmente des tensions négatives vers des positives, nous pouvons en déduire les informations suivantes :

- Le premier maximum correspond au début de la conduction dans le bridge. La tension de grille à laquelle apparait ce maximum définit la tension de seuil du bridge à V_{thb} .

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

- Le second pic correspond à la formation " classique " du canal d'inversion face avant et définit donc la tension de seuil face avant V_{thf} . La figure 4.2.d confirme que cette tension de seuil est la même pour l'A2RAM et le transistor SOI.

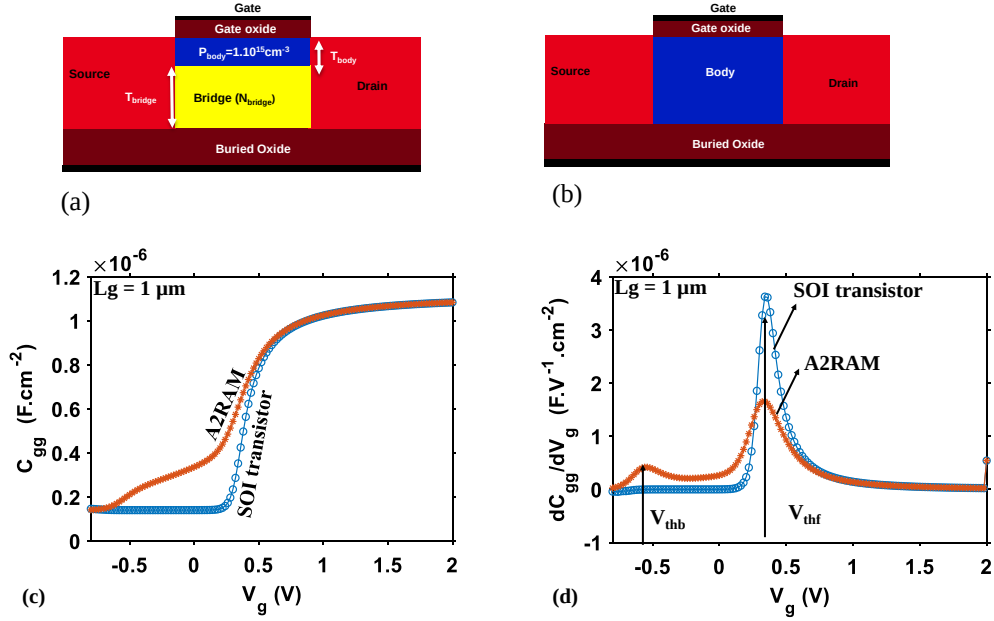


Figure 4.2 : (a) A2RAM et (b) Transistor sur substrat SOI avec la même épaisseur de silicium, caractérisés avec le protocole de la figure 4.1. (c) Caractéristique C_{gg} - V_g de l'A2RAM et du transistor sur substrat SOI. (d) Dérivées premières des courbes C_{gg} - V_g .

Pour confirmer les observations faites sur la figure 4.2, la figure 4.3 montre les variations du profil vertical de la densité d'électrons dans la cellule A2RAM à différentes polarisations de la grille face avant V_g en comparaison des concentrations en porteurs issues du profil de dopage le long du film de silicium (body + bridge). Avant l'activation du bridge ($V_g < V_{thb}$), la structure entière est en régime de désertion, c'est pourquoi, sur la figure 4.2.c, les C_{gg} - V_g du transistor SOI standard et de l'A2RAM sont similaires. Après l'activation du bridge ($V_g \geq V_{thb}$), le profil de la densité d'électrons augmente brusquement dans la partie inférieure du film de silicium (le bridge) mais sa valeur reste inférieure à la valeur nominale du dopage dans le bridge. Le body peut encore être considéré comme étant en régime de désertion. À $V_g = 0$ V, la valeur de la densité électrons augmente, et dans le bridge, elle est presque égale à la valeur du dopage. Cependant, le body commence à entrer en régime de faible inversion en face avant. Enfin, après l'activation du canal face avant ($V_g \geq V_{thf}$), le profil de la densité d'électrons met évidence la présence d'une forte charge d'inversion sous l'oxyde de la grille face avant beaucoup plus élevée que le dopage dans le bridge ; c'est pourquoi, sur la figure 4.2.c, les C_{gg} - V_g du transistor sur substrat SOI et de l'A2RAM sont à nouveau identiques.

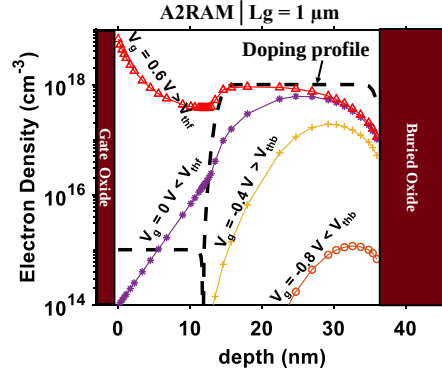


Figure 4.3: Comparaison du profil de dopage avec le profil de densité électrons dans le film de silicium (body + bridge) à différentes polarisations de la grille face avant V_g lors de la simulation de petits signaux présentée figure 4.2.

IV.1.2. Extraction du dopage du bridge

L'extraction du profil de dopage par l'intermédiaire de mesures de C_{gg} - V_g a déjà été réalisée avec plusieurs techniques sur des transistors sur substrat massif. La méthode choisie pour notre cas a été adaptée et repose sur trois équations.

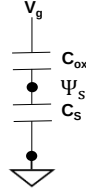


Figure 4.4 : Schéma très simplifié du diviseur capacitif utilisé pour évaluer l'équation 4.1 ci-dessous.

L'équation 4.1 est établie pour un MOSFET sur substrat massif où C_{gg} correspond à C_{ox} en série avec C_s (figure 4.4, où Ψ_s est le potentiel de surface du canal avant à l'interface silicium-oxyde).

$$C_{gg}^{-1} = C_{ox}^{-1} + C_s^{-1} \quad (4.1)$$

où C_{gg} , C_{ox} et C_s sont la capacité de la grille totale, la capacité d'oxyde de grille et la capacité de semi-conducteur, respectivement.

$$Y(V_g) = \frac{1}{C_{gg}^3} \frac{dC_{gg}}{dV_g} = \frac{1}{C_s^3} \frac{dC_s}{d\Psi_s} \quad (4.2)$$

La seconde équation 4.2 est communément appelée 'Maserjian formula', elle a été démontrée dans [Maserjian 74] en utilisant l'équation 4.1 pour extraire la densité d'état d'interface du canal face avant dans les transistors des technologies MOS. À partir de l'équation 4.2, la

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

densité totale de charges dans le semi-conducteur (N_{net}) peut être évaluée en utilisant l'équation 4.3 appelée 'Doping function'.

$$N_{\text{net}} = \left(q\epsilon_s Y(V_g) \right)^{-1} \quad (4.3)$$

avec q la charge élémentaire et ϵ_s la permittivité du silicium.

Ainsi, dans l'hypothèse de régime de désertion totale, l'équation 4.1 peut être utilisée pour évaluer la profondeur de la zone de charge d'espace dans le semi-conducteur et l'équation 4.3 pour déterminer la charge totale dans le semi-conducteur. Dans [Gelder 71], cette technique est utilisée pour un MOSFET sur substrat massif uniformément dopé afin d'extraire le dopage en fonction de l'épaisseur de la zone de charge d'espace. De la même manière, [Chattopadhyay 03] a également utilisé cette technique toujours pour des transistors sur substrat massif, mais avec une hétérojonction fortement et uniformément dopée. Enfin, [Cristoloveanu 86] a utilisé la même technique en régime de désertion pour extraire le dopage en fonction de la profondeur de la zone de charge d'espace dans un transistor sur substrat SOI à film de silicium épais (transistor partiellement déserté PDSOI) et uniformément dopé.

Cette technique était utilisée que pour les transistors MOS épais et / ou uniformément dopés. Par conséquent, notre objectif est d'adapter cette méthode à une couche de SOI mince et faiblement dopée comme la structure A2RAM et d'explorer ses limites de précision. L'idée est d'extraire l'épaisseur du body avec l'équation 4.1 et d'utiliser l'équation 4.3 pour déterminer la valeur de dopage dans le bridge (N_{bridge}). Pour que la densité de charges extraite (N_{net}) soit égale à N_{bridge} , il est important que le body et uniquement lui soit en régime de désertion.

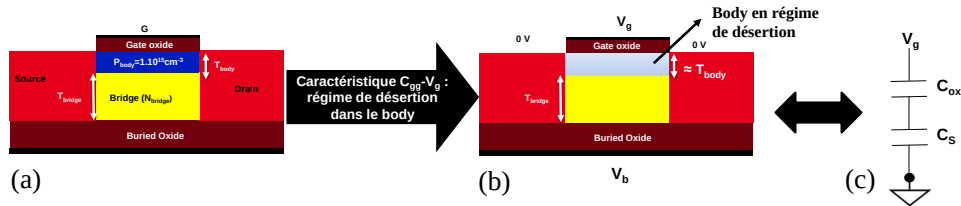


Figure 4.5 : (a) Structure A2RAM initiale ; (b) Structure très simplifiée de l'A2RAM durant les simulations petits signaux, et (c) schéma électrique équivalent de la structure lorsque $V_{\text{thb}} < V_g < V_{\text{thf}}$.

Grâce à la simulation petit signal de la figure 4.2, nous pouvons déterminer cette condition en étudiant l'évolution de la charge (densité d'électrons) dans la structure A2RAM. De la description faite sur la figure 4.3, nous pouvons conclure que l'extraction du profil de dopage doit être effectuée dans la plage de polarisations de la grille face avant V_g telle que : $V_{\text{thb}} < V_g < V_{\text{thf}}$. Car dans cette plage de polarisation la structure de l'A2RAM (figure 4.5.a) correspond idéalement à celle de la figure 4.5.c, c'est-à-dire le bridge en conduction et le body totalement déserté. Son schéma électrique équivalent est présenté sur la figure 4.5.c ; ce qui nous permet d'utiliser l'équation 4.3. La figure 4.6 montre l'évolution de la valeur absolue de la charge $|N_{\text{net}}|$ évaluée à partir de l'équation 4.3 ('Doping function') en fonction de V_g . L'intervalle de polarisations de la grille face avant compris entre V_{thb} et V_{thf} correspond à la situation décrite

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

de la figure 4.5. La 'Doping function' présente un large pic avec la valeur maximale à $V_g = 0$ V égale à 10^{18} cm^{-3} . Cette valeur correspond à la valeur du dopage définie dans la structure TCAD simulée. Par conséquent, le résultat obtenu, nous permet de conclure que cette technique d'extraction du dopage du bridge fonctionne.

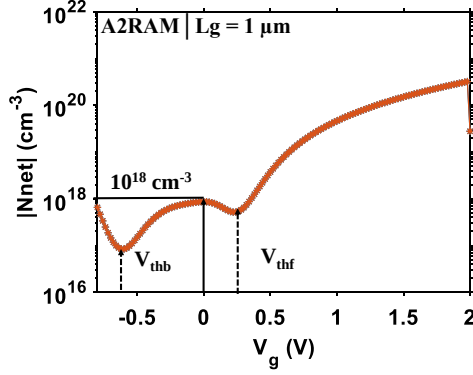


Figure 4.6 : 'Doping function' évaluée à partir de la courbe C_{gg} - V_g de l'A2RAM de la figure 4.2.c.

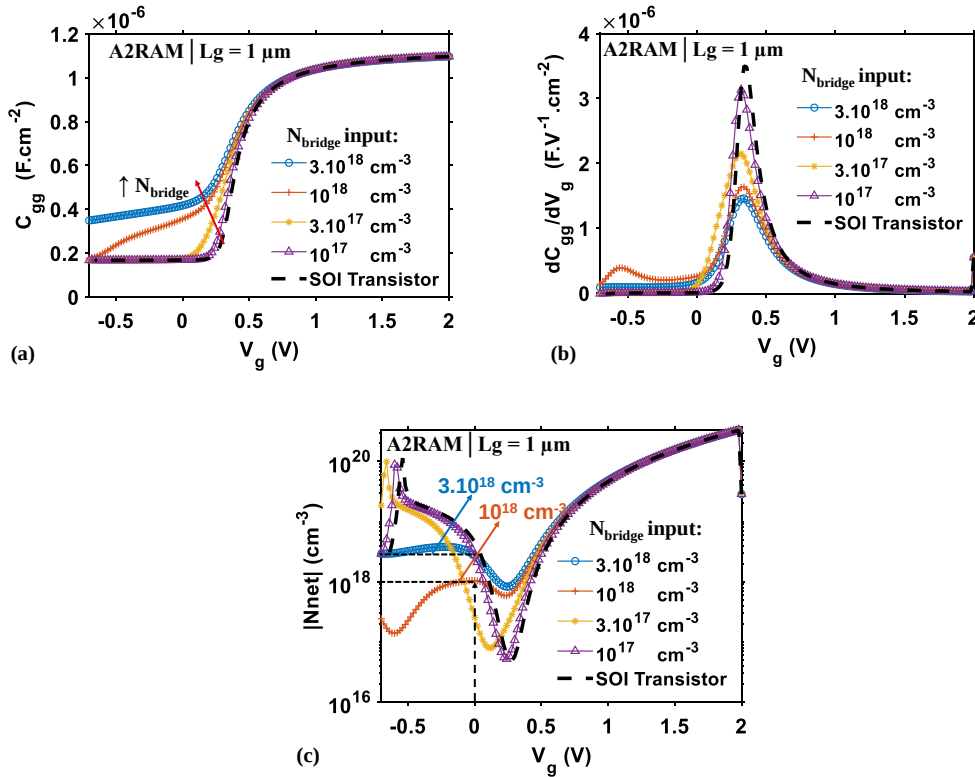


Figure 4.7 : Comparaison des courbes C_{gg} - V_g et de leurs dérivées premières d'un transistor SOI avec des structures A2RAM avec différents dopages de bridge (a, b). (c) 'Doping functions' correspondantes.

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

Pour bien définir les conditions d'utilisation de cette méthode, nous la testons sur des structures A2RAM générées en TCAD où le dopage du bridge variera. Tout d'abord, des variations de dopage du bridge de 10^{17} cm^{-3} à $3.10^{18} \text{ cm}^{-3}$ sont étudiées sur la figure 4.7 avec pour référence le transistor SOI (donc sans bridge).

La figure 4.7.a montre les courbes $C_{gg}-V_g$ et on y constate que le bridge n'a un impact que pour des dopages suffisants, ici supérieur à $3.10^{17} \text{ cm}^{-3}$. Ceci est confirmé par la dérivée première des caractéristiques $C_{gg}-V_g$ sur la figure 4.7.b. Nous pouvons effectivement voir le pic correspondant à la conduction du bridge uniquement pour $N_{\text{bridge}} > 3.10^{17} \text{ cm}^{-3}$. On constate également que ce pic n'est pas visible pour $N_{\text{bridge}} = 3.10^{18} \text{ cm}^{-3}$ parce qu'il apparaît à une tension de grille très négative. En effet, l'augmentation de N_{bridge} produit une diminution de V_{thb} car le champ électrique vertical nécessaire à la déplétion du bridge est plus fort. Le raisonnement opposé tient lorsqu'on réduit le dopage du bridge ($N_{\text{bridge}} < 10^{18} \text{ cm}^{-3}$) : V_{thb} augmente et tend de plus en plus à des valeurs positives. Ainsi, pour $N_{\text{bridge}} \leq 3.10^{17} \text{ cm}^{-3}$, les valeurs de V_{thb} et V_{thf} se rapprochent jusqu'à égalité. Cela implique que les conductions dans le bridge et dans le canal face avant s'activent simultanément. Dans cette condition, la tension seuil du bridge V_{thb} de l'A2RAM concide avec celle d'un transistor sur substrate SOI, et l'extraction de N_{bridge} n'est plus possible. Pour toutes les caractéristiques $C_{gg}-V_g$ de la figure 4.7.a, nous avons évalué leurs 'Doping function' et les courbes obtenues sont représentées sur la figure 4.7.c. Sans surprise, pour un dopage du bridge $N_{\text{bridge}} \leq 3.10^{17} \text{ cm}^{-3}$, il n'est pas possible d'extraire une valeur du dopage du bridge car comme observé sur les figures 4.7a et 4.7.b, le bridge conduit en même temps que l'interface face avant. Néanmoins, avec des valeurs plus élevées de N_{bridge} ($> 3.10^{17} \text{ cm}^{-3}$), le dopage du bridge peut être évalué avec précision (figure 4.7.c).

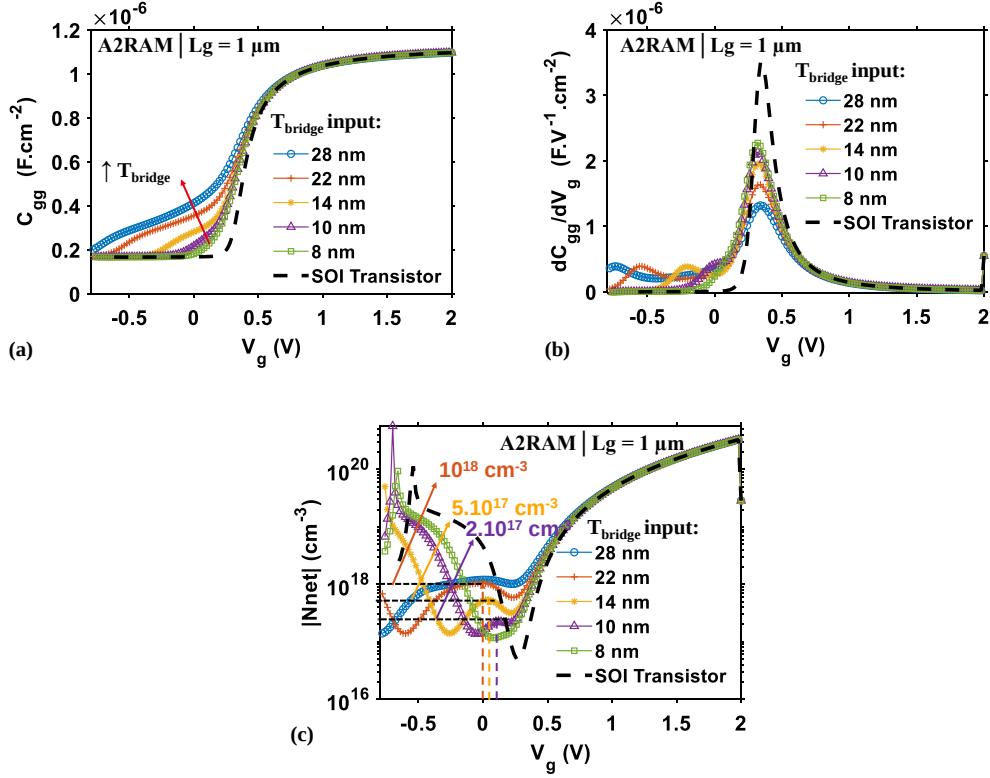


Figure 4.8 : Comparaison des courbes C_{gg} - V_g et de leurs dérivées premières d'un transistor SOI et des structures A2RAM (a, b) tel que définies à la figure 4.2.b mais avec $T_{bridge} = 8, 10, 14, 22, \text{ et } 28 \text{ nm}$, et (c) 'Doping function' obtenues des C_{gg} - V_g des structures A2RAM.

De même, nous avons effectué une analyse de sensibilité similaire pour évaluer la validité de notre méthode vis-à-vis des variations de l'épaisseur de bridge T_{bridge} entre 8 et 28 nm pour un body d'épaisseur constante (Figure 4.8).

La figure 4.8.a, montre les courbes C_{gg} - V_g pour différentes valeurs de T_{bridge} définies sous TCAD toujours en comparaison avec la structure d'un transistor sur substrat SOI. Nous constatons que l'effet du bridge est visible lorsque son épaisseur est supérieure à 8 nm. Ainsi, sur les courbes de la figure 4.8.b qui montre les dérivées premières des caractéristiques C_{gg} - V_g , le pic qui correspond à la conduction du bridge n'est visible que pour $T_{bridge} > 8 \text{ nm}$. De plus, toujours sur la figure 3.8.b, nous constatons qu'augmenter T_{bridge} diminue V_{thb} , c'est pourquoi le pic qui lui correspond se translate vers la gauche car le champ électrique vertical nécessaire à la déplétion du bridge est plus fort. En revanche, réduire la valeur de T_{bridge} va augmenter V_{thb} , car, la charge totale dans le volume du bridge est réduite et dans cette condition, il est plus facile pour la grille face avant de déserrer le bridge. Ce qui explique pourquoi le régime de désertion dans le bridge intervient pour des polarisations de grille face avant moins négatives. V_{thb} sera de plus en plus proche de la valeur de V_{thf} car le couplage de la grille face avant et de la grille face arrière se renforce. C'est pourquoi pour $T_{bridge} \leq 8 \text{ nm}$, il n'est plus possible d'observer un impact distinct du bridge sur la courbe C_{gg} - V_g . Au final, la figure 4.8.c montre les extractions de la 'Doping function' des différentes structures A2RAM dont les courbes C_{gg} - V_g sont présentées sur la figure 4.8.a. Une bonne précision peut être

garantie pour des épaisseurs de bridge dont les valeurs sont supérieures à 22 nm, comme il en ressort des extractions faites sur la figure 4.8.c. Pour un $T_{\text{bridge}} < 22$ nm, le dopage du bridge extrait est très faible par rapport à la valeur d'entrée définie en TCAD en raison du régime de désertion partielle présent dans le bridge.

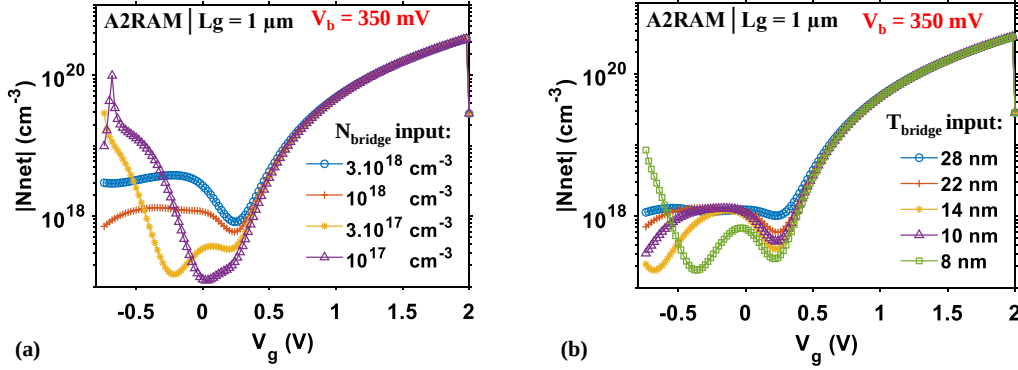


Figure 4.9 : 'Doping function' obtenues à partir des simulations petits signaux pour une tension de grille face arrière $V_b = 350$ mV des structures A2RAM comme définies sur la figures 4.2.b mais avec (a) $N_{\text{bridge}} = 3.10^{18}, 10^{18}, 3.10^{17}$ et 10^{17} cm^{-3} , et (b) $T_{\text{bridge}} = 8, 10, 14, 22$, et 28 nm.

Cependant, nous avons remarqué que sur la figure 4.3, lorsque le bridge est en conduction (pour $V_g > V_{\text{thb}}$), la densité des électrons à l'interface arrière reste inférieure à la valeur du dopage. Cela signifie que la condition de bande plate dans le bridge n'est pas respectée, et que donc l'interface arrière est toujours en régime de désertion. Pour $N_{\text{bridge}} > 3.10^{17} \text{ cm}^{-3}$, l'épaisseur de la zone la désertion du bridge reste à quelques nanomètres au-dessus du BOX, et est bien inférieure à l'épaisseur du bridge. Ceci explique pourquoi l'équation 4.3 est valable pour l'extraction de N_{bridge} (figure 4.8.c). Ainsi, une condition nécessaire pour pouvoir extraire le dopage du bridge avec l'équation 4.3 est que le bridge soit en régime de bande plate. Pour pouvoir respecter cette condition de bande plate lorsqu'on fait varier le dopage et/ou l'épaisseur du bridge, nous pouvons appliquer un signal continue sur la grille face arrière. L'étude de la variation du V_b sur la méthodologie d'extraction du bridge, nous permet d'étudier l'impact du travail de sortie de la face arrière sur la précision de l'extraction. En fait, si la grille arrière est polarisée de façon à maintenir un régime de bandes plates dans le bridge ici à $V_b = 350$ mV (au-delà de cette valeur, le bridge passe en régime de forte accumulation pour l'épaisseur du BOX choisie), il est possible d'extraire une valeur précise de dopage même pour un dopage faible ($< 10^{18} \text{ cm}^{-3}$), comme indiqué sur la figure 4.9.a. Cette observation nous a amené à vérifier aussi si nous pouvions améliorer la précision de l'extraction du dopage du bridge lorsque l'épaisseur du bridge T_{bridge} est réduite. La figure 4.9.b confirme que polariser l'interface arrière améliore la résolution de l'extraction.

Les tableaux 4.1 et 4.2, résument les valeurs du dopage extraites avec notre méthodologie dans les cas où $V_b = 0$ V et $V_b = 350$ mV. Nous constatons à présent que la plage de dopage pouvant être extraite passe de 10^{18} cm^{-3} avec $V_b = 0$ V à $3.10^{17} \text{ cm}^{-3}$ lorsque $V_b = 350$ mV.

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

Valeurs de N_{bridge} (cm^{-3}) définies dans la simulation	Valeurs extraites à $V_b = 0\text{ V}$	Valeurs extraites à $V_b = 350\text{ mV}$
3.10^{18}	3.10^{18}	3.10^{18}
10^{18}	10^{18}	10^{18}
3.10^{17}	NaN	3.10^{17}
10^{17}	NaN	1.5×10^{17}
Tableau 4.1 : Comparaison des valeurs extraites du dopage du bridge N_{bridge} pour deux valeurs de polarisation de la grille face arrière lorsqu'on varie le dopage du bridge en entrée dans TCAD.		

Lorsqu'on varie l'épaisseur du bridge, l'épaisseur minimale garantissant une bonne précision sur l'extraction de N_{bridge} passe de 22 nm avec $V_b = 0\text{ V}$ à 14 nm lorsque $V_b = 350\text{ mV}$.

T_{bridge} (nm) : valeurs définies dans la simulation pour $N_{bridge} = 10^{18}\text{ cm}^{-3}$	N_{bridge} (cm^{-3}) : valeurs extraites à $V_b = 0\text{ V}$	N_{bridge} (cm^{-3}) : valeurs extraites à $V_b = 350\text{ mV}$
28	10^{18}	10^{18}
22	10^{18}	10^{18}
14	5.10^{17}	10^{18}
10	7.10^{17}	10^{18}
8	NaN	7.10^{17}
Tableau 4.2 : Comparaison des valeurs extraites du dopage du bridge N_{bridge} pour deux valeurs de polarisation de la grille face arrière lorsqu'on varie l'épaisseur du bridge T_{bridge} en entrée dans TCAD.		

Par ailleurs, nos simulations considèrent une électrode en métal directement sous la BOX avec un travail de sortie 4.5 eV. Dans les dispositifs A2RAM réels comme introduits en début de ce chapitre, on a un plan de masse en silicium dopé de type p qui joue le rôle de grille face arrière. Nous pouvons en déduire que la présence d'un plan de masse au lieu d'une électrode en métal à l'interface arrière modifiera la condition de la bande plate. Pour déterminer l'impact du plan de masse sur notre méthodologie d'extraction, nous avons effectué une simulation d'une structure A2RAM comprenant un plan de masse de type p (dopage 3.10^{18} cm^{-3}). Sur les figures 4.10. (a, b) (courbes continue bleue et pointillée marron), en raison de la modification des conditions de bande plate à l'interface arrière, la présence du plan de masse induit la désertion dans le bridge, les courbes $C_{gg}-V_g$ ne sont pas alignées. Nous devons appliquer 500 mV sur le plan de masse pour superposer sa caractéristique $C_{gg}-V_g$ avec la structure de départ dont la grille arrière est en métal. Dans cette condition, la valeur du dopage extraite est la même pour les deux structures (figures 4.10. c).

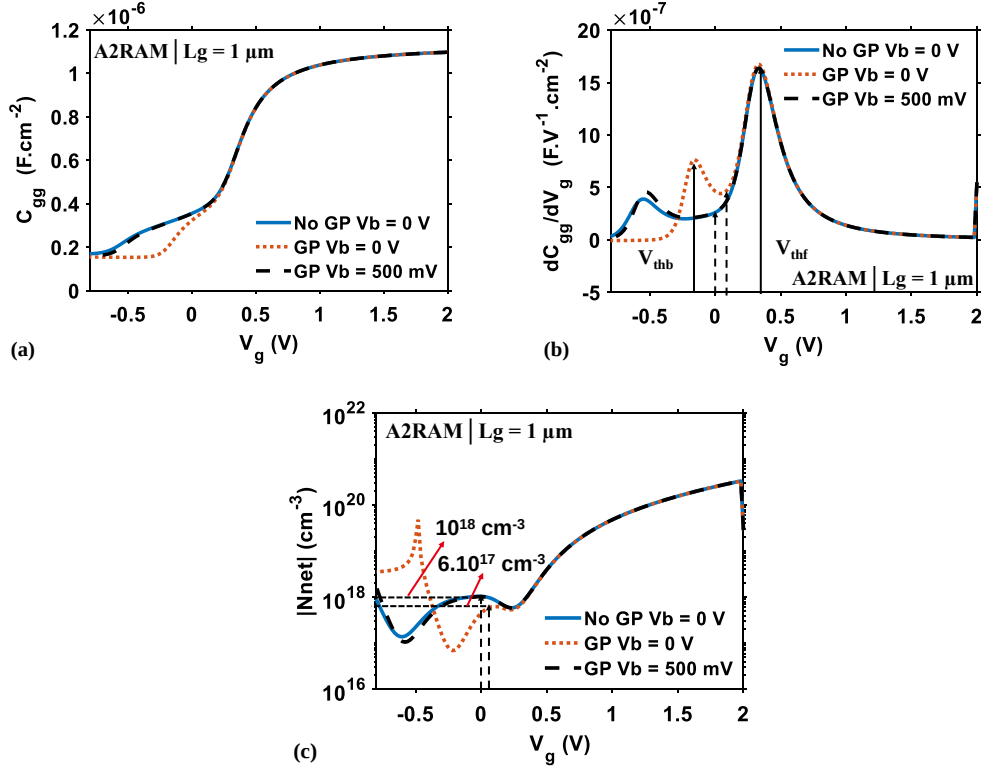


Figure 4.10 : (a) Caractéristiques C_{gg} - V_g , (b) dérivées premières des caractéristiques C_{gg} - V_g , et (c) 'Doping functions' évaluées à partir des caractéristiques C_{gg} - V_g , de l'A2RAM avec un métal sous l'oxyde enterré (BOX) et une A2RAM avec un plan de masse de type p (GP).

En conclusion, nous avons démontré par simulation TCAD que notre méthodologie d'extraction de dopage est valable pour $N_{bridge} > 10^{17} \text{ cm}^{-3}$ et $T_{bridge} > 10 \text{ nm}$ pour $T_{Si} = 36 \text{ nm}$ si la condition de bande plate à l'interface arrière est garantie par une polarisation V_b de la grille face arrière adéquate.

IV.1.3. Extraction de l'épaisseur du body et du bridge

Nous visons l'extraction de l'épaisseur du body (T_{body}), sachant qu'à une tension de grille comprise entre V_{thb} et V_{thf} le bridge est formé, et le body est en régime de désertion. Dans cette condition, on peut utiliser l'équation 4.1. L'épaisseur de la zone de charge d'espace correspond à l'épaisseur du body dopé de type p, on a donc :

$$\text{depletion depth} = T_{body} = \epsilon_{Si}(C_{gg}^{-1} - C_{ox}^{-1}) \quad (4.4)$$

Comme nous connaissons avec précision la valeur de l'épaisseur totale du silicium T_{Si} , nous pouvons facilement en déduire l'épaisseur du bridge T_{bridge} ($T_{bridge} = T_{Si} - T_{body}$). La figure 4.11. a montre la variation de la zone de charge d'espace en fonction de la polarisation de la grille face avant V_g pour des valeurs de dopage du bridge allant de 10^{17} cm^{-3} à $3 \cdot 10^{18} \text{ cm}^{-3}$. L'épaisseur de la zone de charge d'espace correspondante à l'épaisseur électrique du body doit être déterminée à la même tension de grille face avant que celle utilisée pour l'extraction du dopage du bridge (dans nos cas, à $V_g = 0 \text{ V}$). Nous pouvons noter que l'extraction d'une

épaisseur du body n'est possible que pour les mêmes plages de dopage du bridge (figure 4.7). Lorsque N_{bridge} est élevé ($> 3.10^{17} \text{ cm}^{-3}$), il est possible d'extraire l'épaisseur de zone de charge d'espace car le bridge n'est pas totalement déserté par l'effet du champ électrique verticale (issu de la grille face avant). Si N_{bridge} est $< 3.10^{17} \text{ cm}^{-3}$, le film de silicium est complètement déserté, le bridge y compris. Ainsi, la capacité C_{gg} dépend aussi de la capacité du BOX et l'équation 4.4 ne peut plus être utilisée. Ceci explique pourquoi sur la figure 4.11. a, $T_{\text{body}} > T_{\text{si}}$ pour $N_{\text{bridge}} < 3.10^{17} \text{ cm}^{-3}$. Cependant, l'extraction de T_{body} est plus précise pour un dopage élevé : pour $N_{\text{bridge}} = 3.10^{18} \text{ cm}^{-3}$, nous obtenons $T_{\text{body}} = 15.7 \text{ nm}$, à peine 2 nm plus épais que les 14 nm attendus.

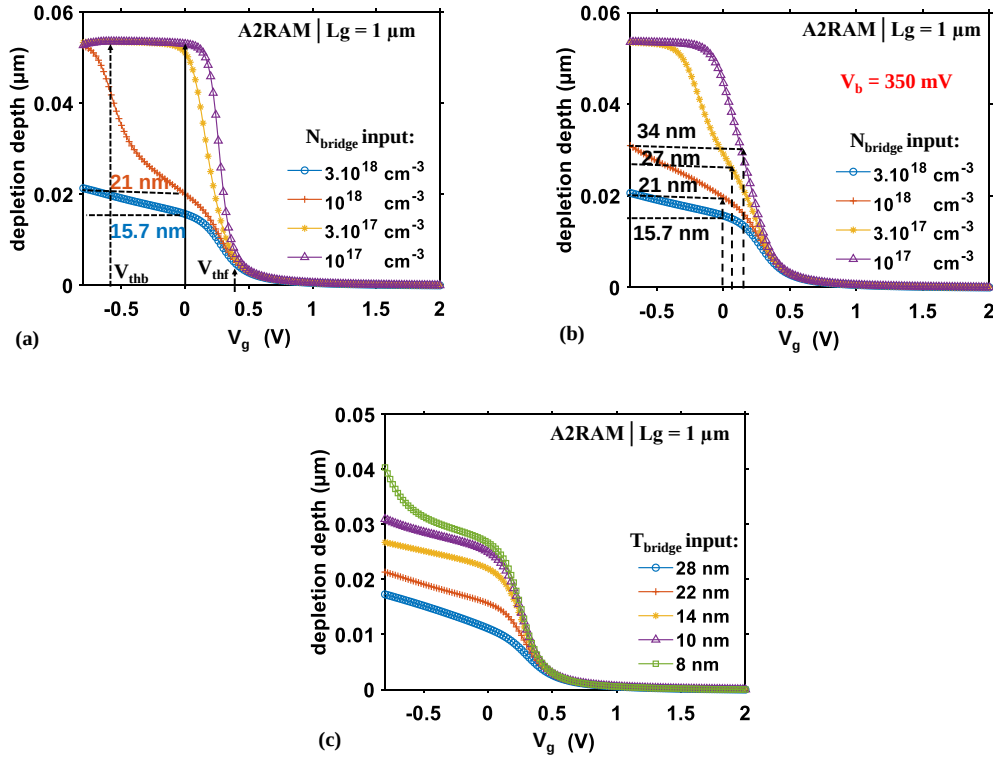


Figure 4. 11 : Epaisseur de la zone de charge d'espace évaluée à partir de l'équation 4. 4 des caractéristiques $C_{\text{gg}}-V_g$ des structures A2RAM comme définies sur la figures 4.2.b avec différents dopages de bridge pour des polarisations en face arrière (a) $V_b = 0 \text{ V}$, et (b) $V_b = 350 \text{ mV}$. (c) Epaisseur de la zone de charge d'espace évaluée à partir de l'équation 4.4 des caractéristiques $C_{\text{gg}}-V_g$ obtenues à partir de la structure de la figure 4. 2. b avec $N_{\text{bridge}} = 3.10^{18} \text{ cm}^{-3}$ et $T_{\text{bridge}} = 8, 10, 14, 22, \text{ et } 28 \text{ nm}$.

Afin d'examiner l'effet de la condition de bande plate sur la précision de T_{bridge} pour $N_{\text{bridge}} < 3.10^{18} \text{ cm}^{-3}$, la figure 4.11. b montre l'extraction de T_{body} à $V_b = 350 \text{ mV}$. Mais, même dans cette condition, la précision de l'extraction n'est pas améliorée. En conclusion, une bonne précision pour l'extraction de T_{body} est garantie si et seulement si $N_{\text{bridge}} > 10^{18} \text{ cm}^{-3}$.

Pour finir, considérons un dopage du bridge de $3.10^{18} \text{ cm}^{-3}$; nous allons déterminer maintenant une plage de valeurs d'épaisseurs du bridge T_{bridge} où l'équation 4 peut encore

être utilisée. La figure 4.11.c montre les résultats obtenus pour différentes valeurs d'épaisseur du bridge (28, 22, 14, 10 et 8 nm) lorsque le dopage du bridge N_{bridge} vaut $3 \cdot 10^{18} \text{ cm}^{-3}$. Les épaisseurs du bridge extraites de la figure 4.11.c sont reportées sur la figure 4.12 et comparées aux valeurs définies dans la simulation TCAD. L'erreur ne dépasse pas une différence de 2 nm (maximum). Par conséquent, la précision d'extraction du T_{body} dépend du dopage du bridge : plus N_{bridge} est élevé, plus la précision est grande.

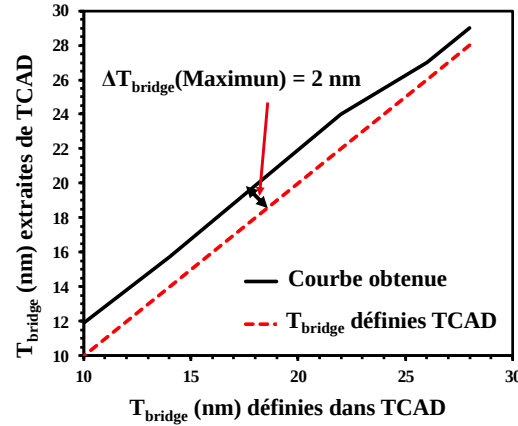


Figure 4.12 : Comparaison des valeurs de l'épaisseur de bridge extraites des simulations TCAD de la figure 4.11.c ; l'écart maximum entre les valeurs extraites et celles définies au préalable dans les simulations TCAD est de 2 nm.

IV.2. Validation expérimentale de la technique de caractérisation du profil de dopage de la cellule A2RAM

IV.2.1. Description des échantillons mesurés

Les échantillons de cellules A2RAM qui ont faits objet de notre étude ont été fabriqués en au Leti. Les étapes clés du processus de fabrication sont données à la figure 4.13. Après un amincissement de la couche de silicium qui se trouve au-dessus de l'oxyde enterré (SOI, d'épaisseur d'environ 6 nm), s'en suit l'étape d'implantation du plan de masse (GP) qui est dopé de type p. Des données que nous avons à notre disposition, la recette d'implantation utilisée prévoit après activation une concentration de Phosphore d'environ 10^{18} cm^{-3} . Nous passons ensuite à la phase d'implantation du dopage du bridge en pleine plaque. Durant cette phase, deux recettes d'implantation ont été utilisées, la différence réside au niveau de l'énergie d'implantation, qui est 1 KeV et 7 KeV. Le comportement attendu prédit par des simulations TCAD est que, pour une énergie faible, les atomes d'Arsenic seront plus localisés à l'interface supérieure du SOI. Pour une énergie d'implantation plus élevée, les atomes seront plus localisés à l'interface arrière du SOI. Après cette étape, nous avons une phase de recuit pour activer les dopants, la valeur du dopage prédit par simulation TCAD après activation est d'environ 10^{18} cm^{-3} . Ensuite, une épitaxie sélective est effectuée pour former le film de silicium complet ; à la fin de cette étape l'épaisseur totale est d'environ 36 nm. Il en suit une étape de dépôt de la grille face avant, puis la gravure premier espaceur, une

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

implantation et activation des zones LDD (Lightly Doped Drain) de source et drain, gravure du second espaceur, et enfin, l'implantation HDD (Highly Doped Drain) suivie de l'activation des dopants source et drain. Nous avons à notre disposition trois plaques, dont la différence principale était l'énergie d'implantation du dopage du bridge, comme reporté dans le tableau 4.3. Sans les mesures électriques, il est difficile de prédire l'impact que le processus de fabrication apportera sur les performances mémoires de l'A2RAM.

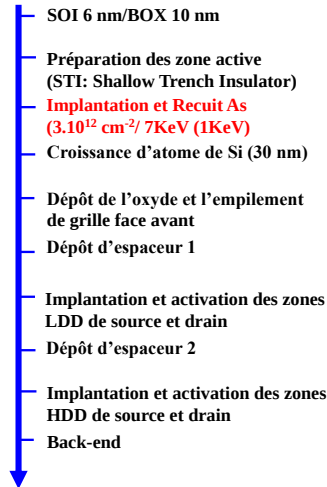


Figure 4.13 : principales étapes du procédé de fabrication de l'A2RAM basé sur le procédé standard de fabrication de transistor SOI du Leti.

	Wafer #07	Wafer #08	Wafer #09
Type de dopage du bridge	n		
Recette d'implantation du bridge	1 keV/2.10 ¹² cm ⁻³	7keV/2.10 ¹² cm ⁻³	
Epaisseur totale du film de silicium T _{Si} (T _{body} + T _{bridge})	36 nm		
Tableau 4.3 : Différences entre les plaques de lot de mémoires A2RAM			

IV.2.2. Extraction de profils de dopage sur des échantillons A2RAM par caractérisation électrique

Dans un premier temps, nous allons utiliser notre technique d'extraction de profil de dopage sur un des échantillons (issu du Wafer #09 tableau 4.3) fabriqués au Leti. Les caractéristiques expérimentales $C_{\text{gg}}-V_{\text{g}}$ à deux V_{b} sont montrées sur la figure 4.14. a, elles sont comparables aux caractéristiques $C_{\text{gg}}-V_{\text{g}}$ obtenues à partir de simulations TCAD. Sur la figure 4.14. b, le double pic met en évidence la conduction du bridge et à l'interface avant. La valeur de dopage du bridge extraite à $V_{\text{b}} = 0 \text{ V}$ est de $4.10^{17} \text{ cm}^{-3}$ (figure 4.14. c). D'après l'analyse TCAD de la section 1.2 de ce chapitre, puisque la valeur extraite est inférieure à 10^{18} cm^{-3} , pour garantir une extraction précise, il est nécessaire d'appliquer une tension sur la face arrière de l'échantillons $V_{\text{b}} > 0 \text{ V}$ aussi proche que possible de la condition de bande plate à

l'interface arrière. D'après le processus de fabrication, nous savons que la concentration du plan de masse (GP) de type P est d'environ 10^{18} cm^{-3} . Ainsi, à partir des résultats des simulations illustrées sur les figures 4.11 et 4.12, nous avons utilisé la caractéristique C_{gg} - V_g mesurée avec $V_b = 500 \text{ mV}$. En conséquence, sur les figures 4.14. a et 4.14. b, nous pouvons remarquer un décalage des courbes dû au changement de la condition de bande plate en face arrière. Ainsi, sur la figure 4.14. c à $V_b = 500 \text{ mV}$, la valeur du dopage du bridge N_{bridge} extraite correspondante (10^{18} cm^{-3}) est supérieure à la précédente ($4 \cdot 10^{17} \text{ cm}^{-3}$) et se situe dans la plage de la valeur attendue. Cependant, la valeur de l'épaisseur du body T_{body} extrait (figure 4.14. d) que ça soit à $V_b = 0 \text{ V}$ (27 nm) ou à $V_b = 500 \text{ mV}$ (24 nm) est surestimée car la valeur du dopage du bridge N_{bridge} est faible (comme nous l'avons constaté et conclu à partir des résultats de la 4.11. b).

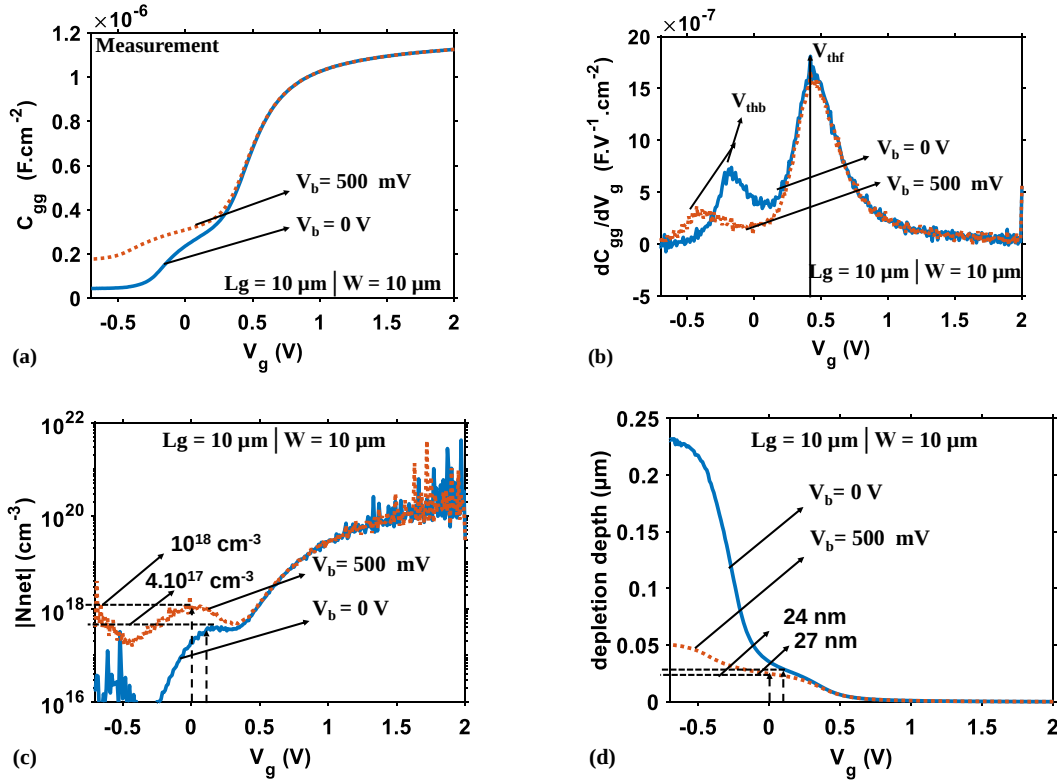


Figure 4.14 : (a) Caractéristiques C_{gg} - V_g de l'échantillons mesuré, (b) dérivée première des caractéristique C_{gg} - V_g , (c) 'Doping function', and (d) zone de charge d'espace en fonction V_g .

IV.2.3. Analyse des performances mémoires A2RAM

Nous avons réalisé des mesures électriques d'opérations mémoires en utilisant le même protocole présenté dans la section 1 du Chapitre 2. Le but principal était de pouvoir mettre en évidence le nouveau mécanisme de programmation l'état '1' que nous avons présenté dans la section 1 du Chapitre 3. Ayant remarqué des variations de résultats obtenus sur deux différentes puces d'une même plaque, nous avons décidé de faire une étude 'statique' en refaisant les mesures sur 10 puces de chaque plaque. Sur la figure 4.15, nous avons l'image

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

d'une plaque mesurée avec les positions des différentes puces (dies). Pour ces caractérisations mémoires, nous avons utilisé le motif des tensions de la figure 2.2 (Chapitre 2), en faisant varier les tensions de programmation de l'état '1'. Sur chaque puce mesurée (avec longueur de grille $L_g = 80 \text{ nm}$ et largeur $W = 10 \text{ }\mu\text{m}$), nous avons extrait les valeurs de I_1 , et I_0 . Par ailleurs, pour pouvoir analyser les résultats obtenus, nous avons utilisé notre méthodologie d'extraction du profil de dopage sur les 3 plaques que nous avons à notre disposition (tableau 4.3), et sur les mêmes puces de la figure 4.15. Pour l'extraction du profil de dopage, nous avons considéré différents échantillons avec des longueurs de grille $L_g = 10 \text{ }\mu\text{m}$ et largeur $W = 10 \text{ }\mu\text{m}$, et l'extraction de la caractéristique a été faite à deux polarisations de la grille face arrière $V_b = 0 \text{ V}$ et $V_b = 500 \text{ mV}$.

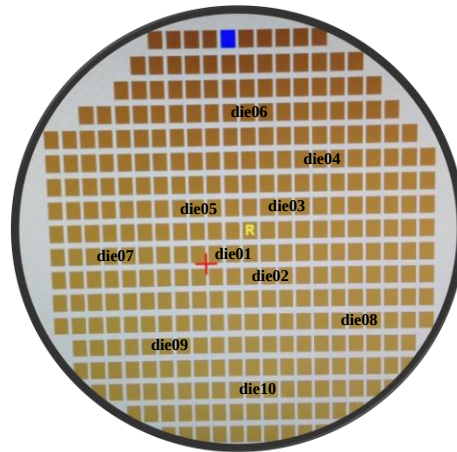


Figure 4. 15 : Cartographie d'une plaque de lot A2RAM avec indication des différentes puces (dies) mesurées.

Le tableau 4.4 illustre les extractions de T_{body} , T_{bridge} et N_{bridge} obtenues avec $V_b = 0 \text{ V}$ et $V_b = 500 \text{ mV}$ effectuées sur la plaque 08 (Wafer #08). Sur la figure 4.16, nous avons les valeurs de I_1 et I_0 pour les mêmes puces du tableau 4.4. Nous pouvons noter que les valeurs de I_1 et I_0 varient en fonction de la puce mesurée bien qu'elles aient subi le même processus de fabrication. Par ailleurs, les profils de dopage sont cohérents avec les valeurs de I_1 et I_0 extraites. Lorsque I_1 et I_0 sont élevés, nous pouvons noter que dans le tableau 4.4, ils correspondent aux puces qui ont un dopage de bridge N_{bridge} élevé et / ou une épaisseur de bridge T_{bridge} élevée. C'est le cas par exemple des puces 01 et 09, où nous constatons que le dopage du bridge pour la puce 01 reste largement supérieur à celui de la puce 09 quel que soit la valeur de V_b appliquée.

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

	Die 01	Die 02	Die 06	Die 09
$V_b=0V$	$T_{body}= 26 \text{ nm}$	$T_{body}= 25 \text{ nm}$	$T_{body}= 26.5 \text{ nm}$	$T_{body}= 46.6 \text{ nm}$
	$T_{bridge}= 10 \text{ nm}$	$T_{bridge}= 11 \text{ nm}$	$T_{bridge}= 9.5 \text{ nm}$	$T_{bridge}= \text{NaN}$
	$N_{bridge}= 4.10^{17} \text{ cm}^{-3}$	$N_{bridge}= 5.10^{17} \text{ cm}^{-3}$	$N_{bridge}= 3.6 \times 10^{17} \text{ cm}^{-3}$	$N_{bridge}= 6.10^{16} \text{ cm}^{-3}$
$V_b=0.5V$	$T_{body}= 24.4 \text{ nm}$	$T_{body}= 23.5 \text{ nm}$	$T_{body}= 24.3 \text{ nm}$	$T_{body}= 25 \text{ nm}$
	$T_{bridge}= 11.6 \text{ nm}$	$T_{bridge}= 12.5 \text{ nm}$	$T_{bridge}= 11.7 \text{ nm}$	$T_{bridge}= 11 \text{ nm}$
	$N_{bridge}= 1.1 \times 10^{18} \text{ cm}^{-3}$	$N_{bridge}= 1.10^{18} \text{ cm}^{-3}$	$N_{bridge}= 1.3 \times 10^{18} \text{ cm}^{-3}$	$N_{bridge}= 5.3 \times 10^{17} \text{ cm}^{-3}$

Tableau 4.4: Extraction du profil de dopage sur quatre puces de la plaque #08 définie dans le tableau 4.3.

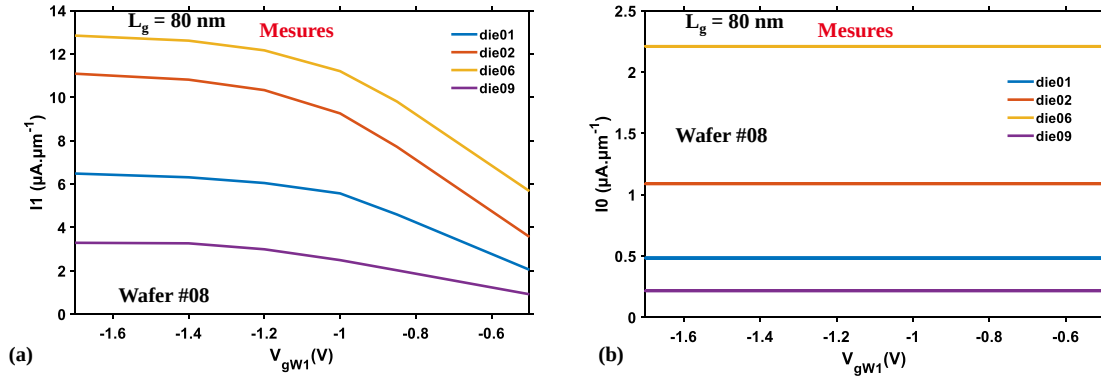


Figure 4. 16 : Valeurs de I_1 ($V_{dW1} = 1.5 \text{ V}$) et I_0 ($V_{gErase} = 0.85 \text{ V}$) extraites des mesures électriques d'opérations mémoires de la cellule A2RAM sur différentes puces pour le wafer #08.

Le tableau 4.5 illustre les extractions de T_{body} , T_{bridge} et N_{bridge} obtenues avec $V_b = 0 \text{ V}$ et $V_b = 500 \text{ mV}$ effectuées sur la plaque 09 (Wafer #09). La figure 4.17 montre les extractions de I_1 et I_0 sur les mêmes puces. Les variabilités observées sur la plaque 08 sont les mêmes que sur la plaque 09. Par ailleurs, sur les plaques #08 et #09, nous pouvons remarquer que même si tous les dispositifs ont le même processus de fabrication, les profils de dopage sont différents sur chaque échantillon. D'où la variabilité sur les résultats obtenus en caractérisation mémoire des figures 4.16 et 4.17.

	Die 01	Die 02	Die 06	Die 09
$V_b=0V$	$T_{body}= 27 \text{ nm}$	$T_{body}= 25.3 \text{ nm}$	$T_{body}= 29 \text{ nm}$	$T_{body}= 34.2 \text{ nm}$
	$T_{bridge}= 8.7 \text{ nm}$	$T_{bridge}= 10.7 \text{ nm}$	$T_{bridge}= 7 \text{ nm}$	$T_{bridge}= \text{NaN}$
	$N_{bridge}= 3.1 \times 10^{17} \text{ cm}^{-3}$	$N_{bridge}= 4.2 \times 10^{17} \text{ cm}^{-3}$	$N_{bridge}= 2.84 \times 10^{17} \text{ cm}^{-3}$	$N_{bridge}= 1.3 \times 10^{17} \text{ cm}^{-3}$
$V_b=0.5V$	$T_{body}= 24.6 \text{ nm}$	$T_{body}= 23.5 \text{ nm}$	$T_{body}= 24.3 \text{ nm}$	$T_{body}= 26 \text{ nm}$
	$T_{bridge}= 11.4 \text{ nm}$	$T_{bridge}= 12.5 \text{ nm}$	$T_{bridge}= 11.7 \text{ nm}$	$T_{bridge}= 10 \text{ nm}$
	$N_{bridge}= 1.1 \times 10^{18} \text{ cm}^{-3}$	$N_{bridge}= 1.10^{18} \text{ cm}^{-3}$	$N_{bridge}= 1.3 \times 10^{18} \text{ cm}^{-3}$	$N_{bridge}= 5.7 \times 10^{17} \text{ cm}^{-3}$

Tableau 4.5 : Extraction du profil de dopage sur quatre puces de la plaque #09 définie dans le tableau 4.3.

CHAPITRE IV : EVALUATION DU PROFIL DE DOPAGE PAR CARACTERISATION ELECTRIQUE DANS UNE CELLULE A2RAM

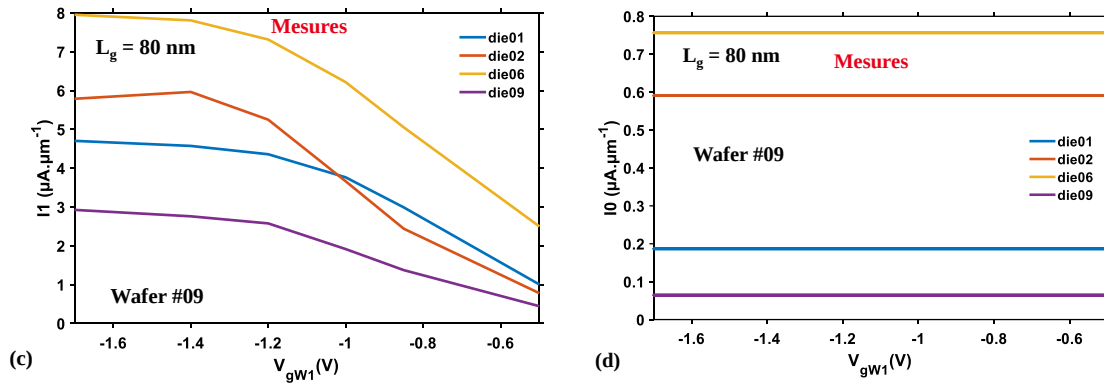


Figure 4.17 : Valeurs de I_1 ($V_{dW1} = 1.5$ V) et I_0 ($V_{gErase} = 0.85$ V) extraites des mesures électriques d'opérations mémoires de la cellule A2RAM sur différentes puces pour le wafer #09.

Dans le tableau 4.6 qui correspond à la plaque #07, nous remarquons que le dopage du bridge extrait à $V_b = 0$ V ou $V_b = 500$ mV reste toujours inférieur à 10^{18} cm^{-3} , par conséquent les épaisseurs extraites sont certainement surestimées. De plus, lorsque $V_b = 0$ V, les profils de dopage obtenus sur la plaque #07 restent différents de ceux des plaques #08 (tableau 4.4) et #09 (tableau 4.5) ; ceci pourrait expliquer pourquoi nous n'avons pas pu mettre en évidence un effet mémoire sur cette plaque.

	Die 01	Die 02	Die 06	Die 09
$V_b = 0V$	$T_{body} = 38 \text{ nm}$	$T_{body} = 34.1 \text{ nm}$		$T_{body} = 49 \text{ nm}$
	$T_{bridge} = \text{NaN}$	$T_{bridge} = \text{NaN}$		$T_{bridge} = \text{NaN}$
	$N_{bridge} = 9.10^{16} \text{ cm}^{-3}$	$N_{bridge} = 1.33 \times 10^{17} \text{ cm}^{-3}$		$N_{bridge} = 3.353 \times 10^{16} \text{ cm}^{-3}$
$V_b = 0.5V$	$T_{body} = 25.5 \text{ nm}$	$T_{body} = 24.4 \text{ nm}$		$T_{body} = 26 \text{ nm}$
	$T_{bridge} = 10.5 \text{ nm}$	$T_{bridge} = 11.6 \text{ nm}$		$T_{bridge} = 10 \text{ nm}$
	$N_{bridge} = 5.54 \times 10^{17} \text{ cm}^{-3}$	$N_{bridge} = 7.24 \times 10^{17} \text{ cm}^{-3}$		$N_{bridge} = 8.10^{17} \text{ cm}^{-3}$

Tableau 4.6 : Extraction du profil de dopage sur quatre puces de la plaques #07 définie dans le tableau 4.3.

De plus, cette technique d'extraction suggère que si l'on veut augmenter les performances de la mémoire, il faut garantir la condition de bande plate dans le bridge (c'est-à-dire appliquer une tension $V_b \neq 0$ V). Pour argumenter davantage cette observation, nous avons refait une caractérisation mémoire en tenant compte des résultats donnés par cette technique d'extraction. Sur la figure 4.18.a, nous avons les chronogrammes de tensions qui sont appliqués sur le drain V_d (en bleu) et celui appliqué sur la grille face avant V_g (en rouge). La figure 4.18.b montre les courants de drain en fonction du temps extraits pendant la séquence de polarisations (figure 4.18.a) à deux V_b différents sur l'échantillon de la puce 01 (die01) du tableau 4.4 avec $L_g = 100$ nm et $W = 10$ μm . On peut remarquer que I_1 est multiplié par 5 lorsque $V_b = 500$ mV alors que I_0 (courant de lecture de l'état '0') reste toujours à une très faible valeur.

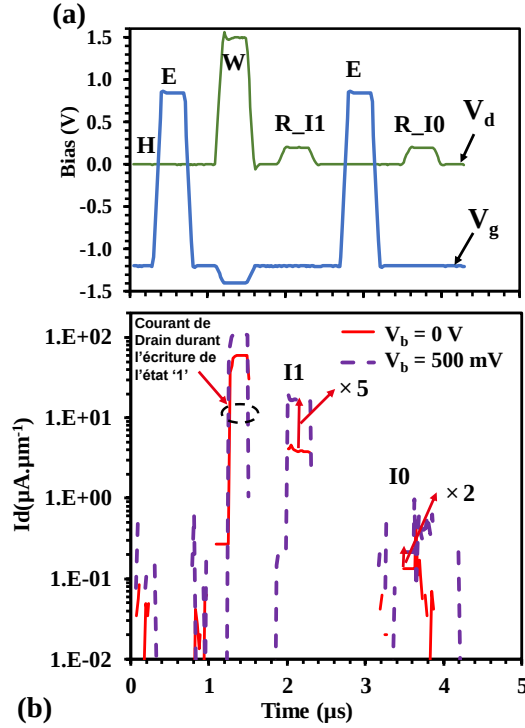


Figure 4. 18 : (a) Séquence de tensions utilisée pour la caractérisation électrique du die01 du tableau 4.4 avec une longueur de grille $L_g = 100$ nm et une largeur $W = 10 \mu m$. (b) Courant de drain extrait du die01 en fonction du temps lorsque nous appliquons la séquence de polarisation de la figure 4. 18. a.

IV.3. Conclusion

Dans ce chapitre, nous avons développé une méthode de caractérisation électrique pour extraire le profil de dopage dans les dispositifs à semiconducteurs sur substrat SOI à film mince avec pour application particulière la cellule mémoire A2RAM. Cette technique fournit des résultats précis pour un dopage du bridge supérieur à $10^{18} cm^{-3}$ et une épaisseur du bridge supérieure à 20 nm. Pour les faibles valeurs de N_{bridge} et de T_{bridge} , la précision peut être améliorée en polarisant la grille face arrière. La technique d'extraction a été utilisée sur les échantillons fabriqués au Leti, et les résultats concordent avec ceux de la simulation. Grâce aux données expérimentales obtenues, nous avons pu interpréter les résultats de la caractérisation mémoire obtenus sur ces échantillons.

CHAPITRE V : MODELISATION DE LA CELLULE MEMOIRE A2RAM

Dans ce chapitre, nous présentons la modélisation compacte des opérations mémoires d'une cellule A2RAM obtenue par la combinaison d'un circuit équivalent et d'un modèle du comportement statique de l'A2RAM. L'élément fondamental est la description analytique de la tension de seuil du bridge essentiel pour la définition de la fenêtre mémoire. Le modèle compact complet de l'A2RAM est ensuite implémenté en Verilog-A pour permettre son utilisation dans un simulateur SPICE et validé par TCAD. Enfin, le modèle est utilisé pour étudier le comportement d'une matrice 2x2 A2RAM par simulation SPICE.

V.1. Modèle compact

V.1.1. Qu'est-ce qu'un modèle compact ?

Le modèle compact est un ensemble d'équations mathématiques qui décrit le comportement d'un dispositif à semiconducteur. Il permet d'évaluer les performances d'un ensemble de dispositifs électroniques à travers des simulateurs de circuits de type SPICE [SPICE]

Dans le cas précis de la modélisation compacte, les équations mathématiques doivent être explicites afin de garantir robustesse et rapidité des simulations de circuit. Pour certains dispositifs électroniques, il est parfois nécessaire de décomposer en sous dispositifs électroniques élémentaires dont le modèle compact existe déjà ou peut être obtenu facilement pour construire le modèle compact d'un nouveau dispositif électronique : on parle dans ce cas de macro-modélisation.

Par ailleurs, l'enjeu du développement d'un modèle compact est de pouvoir décrire le comportement du dispositif électronique dans tous ses régimes de fonctionnement tout en garantissant une robustesse du modèle pour notamment différentes conditions de polarisations et de températures. Les modèles compacts sont indispensables aux concepteurs de circuit car ils leur permettent de simuler des circuits de plusieurs milliers de composants, ce qui est impossible avec des simulations TCAD.

Le Verilog-A est le langage de référence pour le développement du modèle compact. En effet, les modèles compacts implémentés en Verilog-A sont compilés et utilisés directement par les simulateurs de circuit commerciaux comme ELDO (outils de référence de l'industrie) [ELDO]. L'implémentation d'un modèle se fait habituellement en deux parties : une première qui décrit son comportement statique et une seconde qui décrit le comportement dynamique ou transitoire. Par exemple pour le cas d'un transistor MOSFET, le comportement statique nous donnera les dépendances des courants en fonction des polarisations appliquée, et le comportement dynamique nous décrira les dépendances de la charge (capacités) en fonction des polarisations.

V.1.2. Modèle compact de l'A2RAM : Macro-modélisation

Dès études bibliographiques et des chapitres 2 et 3, nous avons vu que les mécanismes physiques impliqués dans l'A2RAM sont complexes ; la modélisation compacte de l'A2RAM basée sur des équations analytiques décrivant chaque opération mémoire n'est pas triviale. Pour cette raison, nous allons utiliser une approche pragmatique qui considère que la fenêtre mémoire est due à un décalage de la tension de seuil induit par la présence de la charge dans le body. Pour cela, nous allons adapter la solution de la modélisation proposée dans [Martinie 17], pour une autre cellule 1T-DRAM appelée Z²-FET [Wan 13]. Considérant cette stratégie,

un circuit équivalent et un modèle compact DC sont combinés pour modéliser les opérations mémoires de l'A2RAM (figure 5.1). Nous présenterons tour à tour la modélisation compacte DC de l'A2RAM et ensuite la modélisation des opérations mémoires A2RAM, tous deux implémentés en Verilog-A pour être utilisés dans des outils de CAO tels que ELDO.

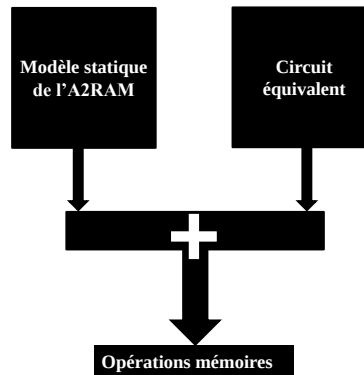


Figure 5.1 : Schéma décrivant la stratégie de modélisation compacte de l'A2RAM.

V.2. Modélisation compacte du comportement DC de l'A2RAM

V.2.1. Stratégie de modélisation

Le défi majeur pour la modélisation compacte quasi statique de l'A2RAM est la description de l'effet du bridge. Dans la section 1.1 du Chapitre 4 sur la figure 4.2.d, nous avons remarqué que la dérivée première de la capacité de grille face avant (dC_{gg}/dV_g) de l'A2RAM présente deux pics mettant en évidence l'activation de la conduction dans le bridge et du canal face avant.

Le pic observé pour les tensions de grille les plus élevées est aligné avec l'unique pic du transistor SOI et correspond à la tension de seuil du canal d'inversion face avant. Le pic vu pour les plus faibles polarisations de grille marque l'activation de la conduction dans le bridge de l'A2RAM. Par conséquent, deux tensions de seuil seront définies pour décrire le comportement de l'A2RAM: le seuil de conduction dans le bridge (V_{thb}) et le seuil du canal face avant dans une A2RAM (V_{thf}).

Comme les caractéristiques $C_{gg}-V_g$ (figure 4.2.c du Chapitre 4) se superposent lorsque $V_g \geq V_{thf}$, la conduction dans le canal face avant dans une structure A2RAM est la même que celle d'un transistor sur substrat SOI : on peut dire que le courant de la couche d'inversion est le même. En plus, l'A2RAM a un canal de conduction supplémentaire qui vient du bridge. Ainsi, le courant d'inversion dans une A2RAM peut être modélisé par le courant d'inversion d'un transistor sur substrat SOI, et la conduction due à la présence du bridge peut être modélisée avec une résistance variable. Notre stratégie pour la macro-modélisation du comportement DC consiste à considérer la cellule A2RAM comme une combinaison d'un transistor sur substrat SOI en parallèle avec une résistance variable (figure 5.2).

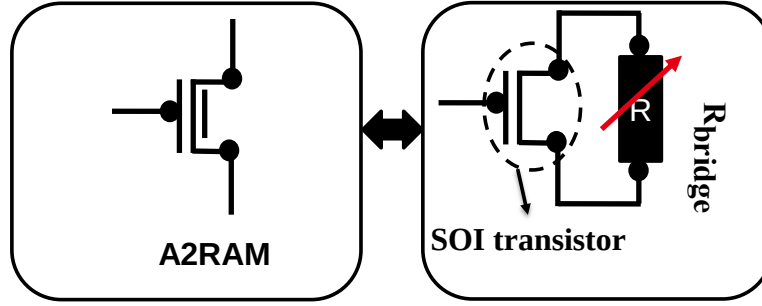


Figure 5.2 : Principe de modélisation compacte DC de l'A2RAM : la cellule est considérée comme une combinaison d'un transistor sur substrat SOI en parallèle avec une résistance variable.

V.2.2. Modélisation de la résistance du bridge de l'A2RAM

Dans un premier temps, nous allons faire l'étude de la caractéristique courant de drain I_d en fonction de la tension de la grille face avant de l'A2RAM (I_d - V_g) en la comparant avec celle d'un transistor sur substrat SOI standard (figure 4.2.b dans la section 1.1 du Chapitre 4). Le but étant de comprendre comment le transistor sur substrat SOI et la résistance variable interagiront pour la modélisation du comportement DC de l'A2RAM. Pour simplifier l'analyse, nous considérons des dispositifs à canal long ($L_g = 1 \mu\text{m}$) pour éviter les effets canaux courts. Les caractéristiques DC de l'A2RAM et du transistor sur substrat SOI sont obtenues à partir des simulations 2D TCAD.

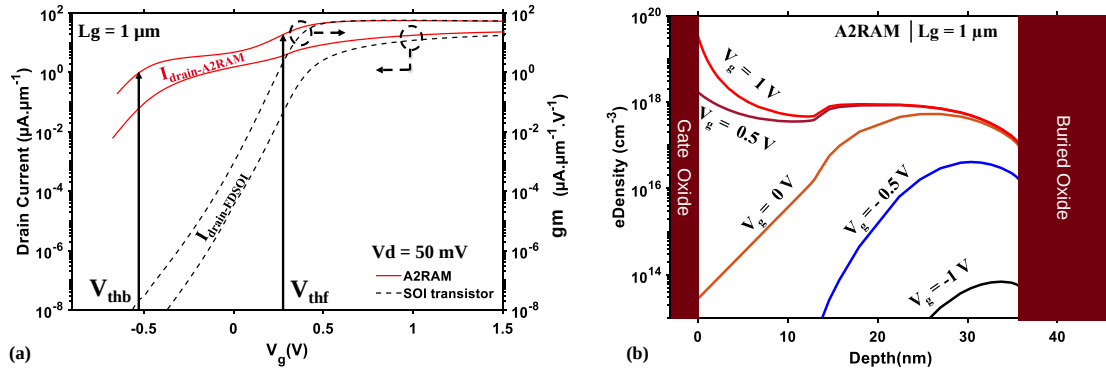


Figure 5.3 : (a) Comparaison de I_d - V_g (axe gauche) et g_m - V_g (axe droit) entre la cellule A2RAM et un transistor sur substrat SOI standard (figure 4.2.b). (b) Densité d'électron dans la profondeur du film de silicium (body + bridge) extraite des simulations des TCAD de la caractéristique statique de l'A2RAM $V_d = 50 \text{ mV}$ à différents V_g .

La figure 5.3.a sont reportées les caractéristiques statiques I_d - V_g (courant de drain I_d en fonction de tension de la grille face avant V_g) sur l'axe de gauche, et leurs transconductances (g_m) respectives sur l'axe de droite dans les cas de l' A2RAM et du transistor sur substrat SOI.

- Pour les tensions de grille face avant négatives ($V_g < 0 \text{ V}$), le courant de drain de l' A2RAM est essentiellement dû à la conduction dans le bridge et diminue

progressivement lorsqu'on se déplace vers les tensions de grille très négatives car le bridge est de plus en plus déserté.

- Pour des polarisations de grille face avant positives ($V_g > 0$ V) la différence entre la caractéristique statique de l'A2RAM et du transistor sur substrat SOI est réduite (mais ne s'annule pas) et reste constante puisque le bridge est aussi en conduction.

La comparaison de ces deux caractéristiques statiques révèle effectivement la présence de deux canaux dans l'A2RAM. C'est pourquoi sur la caractéristique de la transconductance de l'A2RAM nous observons un plateau constant à partir des tensions de grille face avant $V_g < 0$ V (contrairement au transistor SOI). La figure 5.3.b montre les extractions des profils verticaux de la densité d'électrons dans la profondeur du film de silicium (body + bridge) à différentes tensions de la grille face avant dans l'A2RAM.

Pour $V_g \geq V_{thf}$, en plus de la charge d'inversion habituelle à l'interface face avant, nous avons également une charge dans le bridge qui reste constante bien que la tension de la grille face avant V_g augmente : le body est en régime d'inversion et le bridge est en régime d'accumulation. Cependant, lorsque $V_g < V_{thf}$, la charge dans le canal face avant est quasiment inexistante et passe donc en régime de désertion, mais le bridge est toujours en conduction bien que sa densité d'électrons ait diminuée. En effet, pour des tensions de grille face avant négatives, le bridge est partiellement déserté, d'où la réduction progressive de la densité d'électrons et ceci jusqu'à $V_g \approx V_{thb}$. En dessous de V_{thb} , la densité d'électrons diminue de façon exponentielle et le bridge entre en régime de désertion.

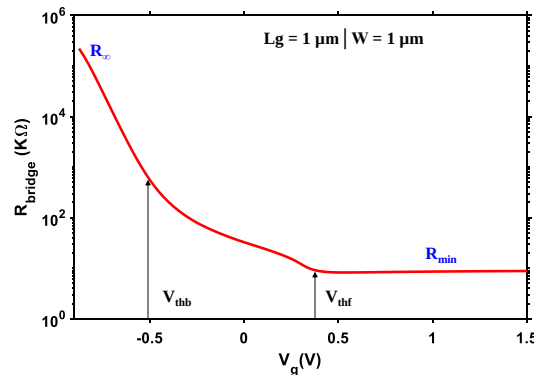


Figure 5.4 : Variation de la résistance du bridge de l'A2RAM en fonction de la polarisation de grille extraite de TCAD.

En résumé, la comparaison faite sur la figure 5.3 montre que la différence principale entre les caractéristiques statiques de l'A2RAM et du transistor SOI est due à la conduction dans le bridge. Ainsi, à partir des caractéristiques statiques (I_d - V_g) de la figure 5.3.a, nous avons extrait la variation de la 'résistance du bridge' R_{bridge} (avec $R_{bridge} = V_d / ((I_{drain-A2RAM} - I_{drain-FDSOI}))$) en fonction de la tension de la grille face avant, illustrée sur la figure 5.4.

- Lorsque $V_g \geq V_{thf}$, la résistance du bridge reste constante (R_{min}) ; cela est dû au fait que la densité d'électrons dans le bridge reste constante, comme indiqué sur la figure 5.3.b.

- Pour des tensions de grilles face avant tels que : $V_{thb} < V_g < V_{thf}$; la résistance du bridge augmente lentement. Comme nous l'avions noté sur la figure 5.3.b, la densité d'électron entre V_{thb} et V_{thf} diminue parce que le bridge commence à être déserté par le champ électrique vertical induit par la grille face avant.
- Pour $V_g \leq V_{thb}$, la résistance du bridge augmente de manière exponentielle jusqu'à R_{∞} , lorsque le bridge est complètement déserté. En fait, comme indiqué sur la figure 5.3.b, pour $V_g \ll 0$ V, la concentration des électrons (n) varie exponentiellement avec V_g . En considérant le bridge comme un bloc résistif implique que $R_{bridge} \propto 1 / n$, ainsi R_{bridge} augmentera de manière exponentielle lorsque V_g diminuera.

La variation de la résistance du bridge du régime de désertion partielle au régime de désertion totale peut être modélisée par l'équation empirique suivante que nous avons établie :

$$R_{bridge} = R_{\infty} - (R_{\infty} - R_{min}) \left[1 - \left[1 + \exp\left(\frac{V_g - (V_{thb} + V_{ref})}{smooth}\right) \right]^{-1} \right], \quad 5.1$$

où V_{ref} est une tension de référence utilisée comme paramètre d'ajustement, '*smooth*' est un facteur utilisé pour lisser la transition entre R_{∞} (bridge entièrement déserté) et R_{min} (bridge en accumulation). Il est important de noter que l'équation 5.1 et ses dérivés sont des fonctions continues ; c'est une condition fondamentale pour la modélisation SPICE.

Pour valider notre modèle décrivant la variation de la résistance du bridge R_{bridge} en fonction de la tension de la grille face avant, nous avons extrait de la simulation TCAD R_{min} , R_{∞} et V_{thb} pour les introduire dans l'équation 5.1. Sur la figure 5.5, nous avons superposé notre modèle R_{bridge} avec l'extraction faite à partir de simulations TCAD, et nous pouvons noter que notre modèle réussit à reproduire les résultats de simulations à éléments finis. Cependant sur la figure 5.5, nous observons un changement de concavité pour des polarisations de grille au-tour de V_{thf} que notre modèle ne réussit pas à reproduire. En effet, ces polarisations marquent le passage du régime de désertion au régime d'inversion dans le body, chose qui n'est pas modélisée par l'équation 5.1.

Cependant, même si R_{min} , R_{∞} , et V_{thb} sont des paramètres facilement extraits des simulations TCAD (ou de données expérimentales), il convient de souligner que la tension de seuil du bridge V_{thb} en plus de dépendre des paramètres technologiques de l'A2RAM, dépend aussi de la polarisation de la grille face arrière de l'A2RAM. Dans les chapitres 2 et 4, nous avons vu que la polarisation de la grille face arrière (V_b) impacte énormément sur les performances mémoires. Il est donc nécessaire d'évaluer V_{thb} avec une expression analytique, pour que la modélisation de $R_{bridge}(V_g)$ tienne compte des variations des paramètres de la structure de l'A2RAM et de la polarisation appliquée sur sa grille face arrière.

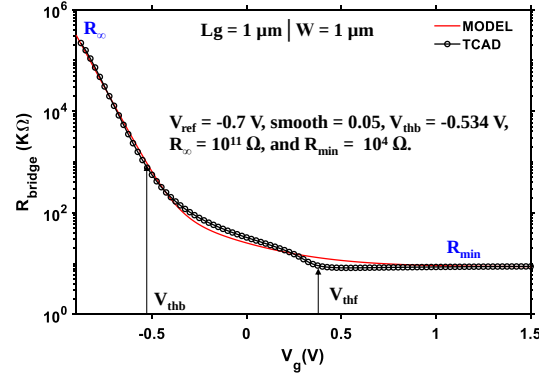


Figure 5.5 : Variation de la résistance du bridge de l'A2RAM en fonction de la polarisation de grille obtenue par TCAD (courbe en symbole) et l'équation 5.1 (courbe continue).

V.3. Modélisation de la tension de seuil du bridge

V.3.1. Expression potentiel électrostatique dans le film de silicium

Nous voulons modéliser la tension de seuil du bridge V_{thb} qui marque la transition entre le régime de désertion total et le régime neutre dans le bridge. Pour cette modélisation nous n'allons considérer le cas idéal d'un dispositif A2RAM à canal long et tenir juste compte des effets 1D. La figure 5.6 présente la structure 1D de l'A2RAM où sera résolue l'équation de Poisson pour en déduire le modèle de la tension seuil du bridge V_{thb} . Sur l'empilement vertical de la structure A2RAM sont indiqués les dopages avec un profil constant dans les régions du body (de type P, égal à N_A) et du bridge (de type N, égal à N_D). T_{sc} , T_{body} et T_{BOX} sont respectivement les épaisseurs totales semi-conducteur (body + bridge), du body et de l'oxyde enterré BOX. Ψ_{sf} et E_{sf} sont respectivement le potentiel électrostatique et le champ électrique de surface à l'interface avant, et Ψ_{sb} et E_{sb} , ceux de l'interface arrière. V_g et V_b sont les tensions de grille face avant et face arrière.

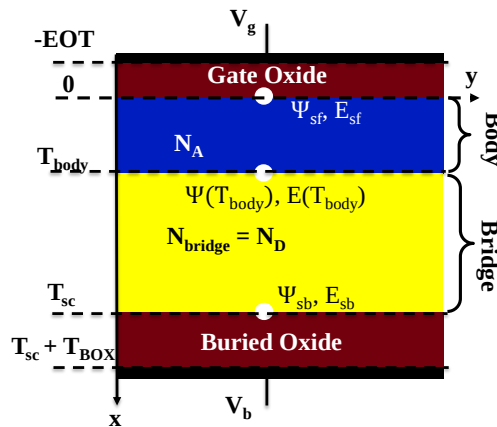


Figure 5.6 : Description de la structure A2RAM utilisée pour résoudre l'équation 1D de Poisson.

Dans l'hypothèse que la structure de la figure 5.6 soit en régime de désertion totale, et que tous les dopants soient ionisés ; l'équation 1D de Poisson dans chacune des parties du film de silicium de l'A2RAM s'exprime par :

$$\begin{cases} \frac{d^2\Psi(x)}{dx^2} = \frac{q}{\epsilon_{sc}} N_A, \text{ for } 0 \leq x \leq T_{body} \\ \frac{d^2\Psi(x)}{dx^2} = -\frac{q}{\epsilon_{sc}} N_D, \text{ for } T_{body} \leq x \leq T_{sc} \end{cases} \quad (5.2)$$

où $\Psi(x)$ est le potentiel électrostatique à un point x dans la profondeur du film du silicium, q la charge élémentaire, ϵ_{sc} la permittivité du matériau semi-conducteur. En intégrant deux fois de suite l'équation 5.2 toujours dans chacune des portions constituant le film de silicium, conduit à l'expression du potentiel électrostatique suivante :

$$\begin{cases} \Psi(x) = \frac{q}{2 \cdot \epsilon_{sc}} N_A x^2 - E_{sf} \cdot x + \Psi_{sf}, \text{ for } 0 \leq x \leq T_{body} \\ \Psi(x) = -\frac{q}{2 \cdot \epsilon_{sc}} N_D X^2 + E_{sb} \cdot X + \Psi_{sb}, \text{ for } T_{body} \leq x \leq T_{sc} \end{cases} \quad (5.3)$$

avec $X = (T_{sc} - x)$. L'équation 5.3 contient donc quatre variables inconnues Ψ_{sf} , Ψ_{sb} , E_{sf} et E_{sb} ; ainsi, nous avons besoin de quatre autres équations.

Les deux premières équations sont données par les conditions de continuité du vecteur champ électrique de déplacement aux interfaces oxyde de grille face avant/body et oxyde enterré/bridge exprimées par :

$$E_{sf} = \frac{C_{ox}}{\epsilon_{sc}} (V_{gf} - \Psi_{sf}) \quad (5.4)$$

$$E_{sb} = -\frac{C_{box}}{\epsilon_{sc}} (V_{gb} - \Psi_{sb}) \quad (5.5)$$

avec $C_{ox} = \frac{\epsilon_{ox}}{EOT}$, $C_{box} = \frac{\epsilon_{ox}}{T_{BOX}}$, $V_{gf} = (V_g - V_{FB-f})$, $V_{gb} = (V_b - V_{FB-b})$; ϵ_{ox} est la permittivité de l'oxyde (SiO_2) de la grille face avant, V_{FB-f} and V_{FB-b} sont les tensions de bandes plates des grilles avant et arrière, respectivement.

Les deux dernières équations seront exprimées en considérant également la continuité du vecteur déplacement du champ électrique, et la continuité du potentiel électrostatique à l'interface body/bridge. Ces conditions limites sont données par les équations 5.6 et 5.7 respectivement.

$$\frac{q}{\epsilon_{sc}} N_A T_{body} - E_{sf} = \frac{q}{2 \cdot \epsilon_{sc}} N_D (T_{sc} - T_{body}) - E_{sb} \quad (5.6)$$

$$\frac{q}{2 \cdot \epsilon_{sc}} N_A T_{body}^2 - E_{sf} \cdot T_{body} + \Psi_{sf} = -\frac{q}{2 \cdot \epsilon_{sc}} N_D (T_{sc} - T_{body})^2 + E_{sb} (T_{sc} - T_{body}) + \Psi_{sb} \quad (5.7)$$

ce qui permet d'obtenir les expressions analytiques de Ψ_{sf} (équation 5.8) et de Ψ_{sb} (équation 5.9).

$$\Psi_{sf} = \frac{V_{gf}(C_{ox} \cdot C_{sc} + C_{box} \cdot C_{ox}) + V_{gb} \cdot C_{box} \cdot C_{ox} + \gamma}{C_{ox} \cdot C_{sc} + C_{box} \cdot C_{sc} + C_{box} \cdot C_{ox}} \quad (5.8)$$

$$\Psi_{sb} = \frac{(V_{gb} + \delta) C_{box} + \Psi_{sf} \cdot C_{sc}}{C_{sc} + C_{box}} \quad (5.9)$$

avec $C_{sc} = \frac{\epsilon_{sc}}{T_{sc}}$; les paramètres γ et δ sont des constantes qui dépendent des paramètres technologiques de l'A2RAM dont les expressions sont données par les équations 5.10 et 5.11 respectivement :

$$\gamma = \epsilon_{sc} \cdot \alpha \cdot C_{sc} - [\beta - \alpha(T_{sc} - T_{body})] \quad (5.10)$$

$$\delta = (\beta + \alpha \cdot T_{body}) \frac{C_{sc}}{C_{box}} \quad (5.11)$$

α et β sont donnés par les équations 5.11 et 5.12.

$$\alpha = \frac{q}{\epsilon_{sc}} N_D (T_{sc} - T_{body}) - \frac{q}{\epsilon_{sc}} N_A \cdot T_{body} \quad (5.12)$$

$$\beta = \frac{q}{2 \cdot \epsilon_{sc}} N_D (T_{sc} - T_{body})^2 - \frac{q}{2 \cdot \epsilon_{sc}} N_A (T_{body})^2 \quad (5.13)$$

V.3.2. Validation par TCAD

Pour pouvoir valider l'expression du potentiel électrostatique dans la profondeur du film de silicium donnée par l'équation 5.3, nous allons d'abord valider nos modèles de potentiel électrostatiques de surface avant et arrière des équations 5.5 et 5.6.

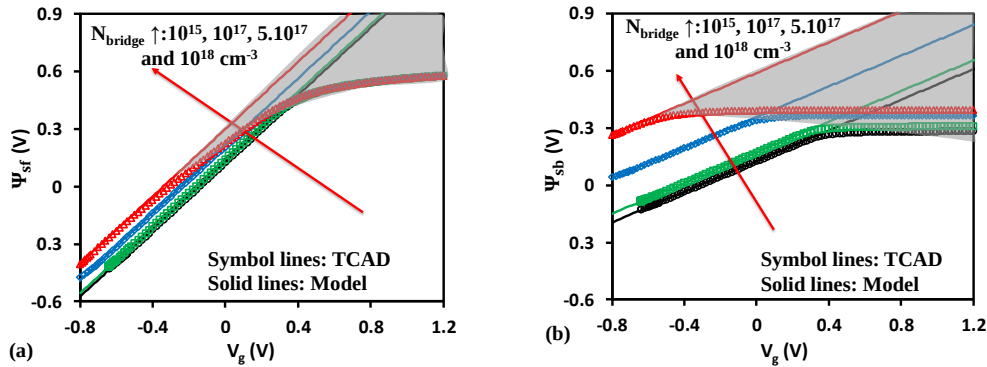


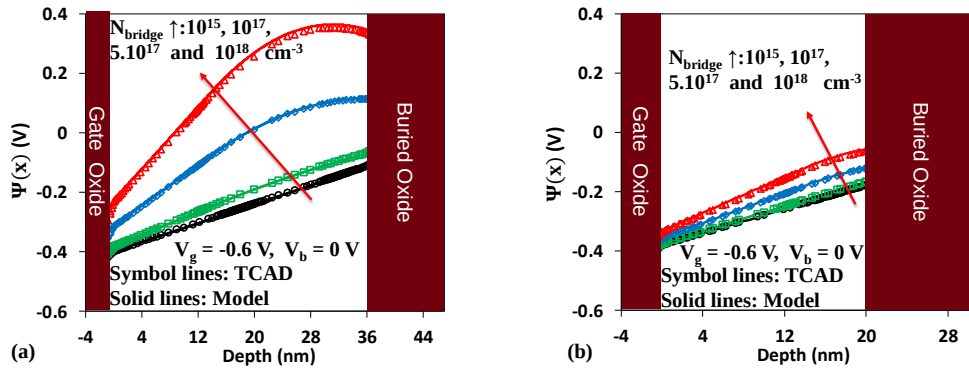
Figure 5.7 : (a) Potentiel à l'interface avant, et (b) potentiel à l'interface arrière en fonction de la tension de la grille face avant extraits de TCAD (symboles) et modèle (lignes) pour différents N_{bridge} .

En TCAD, durant les simulations petits signaux comme décrit dans le Chapitre 4, il est possible d'avoir accès au potentiel électrostatique en tous points et pour tous jeux de polarisations appliquées sur le dispositif simulé. Ainsi, nous avons extrait les valeurs du potentiel électrostatique face avant et arrière à différents dopages du bridge. Par ailleurs, pour pouvoir comparer les extractions TCAD avec les modèles donnés par les équations 5.8 et 5.9, il fallait tenir compte des conditions de bande plate données en TCAD : $V_{FB-f} = V_{FB-b} = 0.16 \text{ V}$ quel qu'en soit le profil de dopage le long du film de silicium.

Sur les figures 5.7.a 5.7.b, nous reportons les comparaisons entre les extractions TCAD et nos modèles des variations du potentiel électrostatique aux interfaces avant et arrière en fonction

de la polarisation de la grille face avant pour différents dopages du bridge. Les zones grisées sur les figures 5.7.a et 5.7.b correspondent aux polarisations de la grille face avant qui mettent en conduction le bridge, et donc hors des gammes de validité de notre modèle. Lorsque le bridge est en régime de désertion (parties non grisées des figures 5.7.a et 5.7.b), nous observons des variations linéaires de Ψ_{sf} et Ψ_{sb} en fonction de V_g , et nos modèles se superposent parfaitement avec les extractions TCAD.

La validation de nos modèles de potentiels électrostatiques aux interfaces avant et arrière (équations 5.8 et 5.9) nous permet à présent de valider l'expression du potentiel électrostatique dans la profondeur du film de silicium en régime de désertion donnée par l'équation 5.3. Les figures 5.8.a et 5.8.b montrent la comparaison entre les extractions TCAD et notre modèle du profil vertical du potentiel électrostatique dans le film de silicium $\Psi(x)$ pour différents dopages du bridge N_{bridge} et pour deux épaisseurs totales du film de silicium T_{sc} (36 ou 20 nm) à une polarisation de grille face avant donnée ($V_g = -0.6$ V) assurant un régime de désertion dans la structure entière. Encore une fois, notre modèle de l'équation 5.3 reproduit parfaitement les variations données par les extractions TCAD.



V.3.3. Modélisation V_{thb} d'une cellule A2RAM

A partir de la description 1D du potentiel électrostatique dans la profondeur du film de silicium de l'A2RAM en régime de désertion, l'expression analytique de la tension de seuil du bridge V_{thb} peut être déterminée. La première étape consiste à définir un critère de seuil marquant le passage du régime de désertion totale dans la profondeur film de silicium (body + bridge), au régime neutre dans le bridge. Ayant déjà un modèle du potentiel électrostatique de surface à l'interface arrière Ψ_{sb} (équation 5.9) qui dépend de V_g par le biais de Ψ_{sf} , il nous suffit de trouver un critère de seuil sur ce dernier pour pouvoir déterminer V_{thb} . Du Chapitre 4, nous avons vu que grâce à la dérivée première de la caractéristique C_{gg} - V_g il est possible d'évaluer (extraire) la tension de la grille face avant qui met en conduction le bridge ; ainsi, on pourrait donc aussi évaluer en même temps la valeur de Ψ_{sb} correspondante pour en déduire un critère de seuil.

Dans cette optique, nous avons fait des simulations petits signaux, en faisant varier le dopage dans le bridge. Sur la figure 5.9, nous reportons l'évolution du potentiel électrostatique de surface de l'interface arrière Ψ_{sb} (axe à gauche) et de la dérivée première de la capacité de la gille face avant C_{gg} de l'A2RAM en fonction de la tension de la grille face avant V_g (axe à droite) pour différentes valeurs de dopage du bridge ($N_{bridge} = 10^{18}, 9 \cdot 10^{17}, 8 \cdot 10^{17}$ et $6 \cdot 10^{17} \text{ cm}^{-3}$). Pour chaque structure, nous avons extrait la valeur de Ψ_{sb} lorsque le bridge passe en régime d'accumulation et nous l'avons comparée au potentiel de Fermi φ_F dans le bridge. Nous avons remarqué que le bridge passe en régime d'accumulation lorsque $\Psi_{sb} = \Psi_{sb_{th}} = 0.8\varphi_F$. Cette égalité nous a servi de critère pour la tension seuil, et lorsque l'on l'injecte dans les équations 5.8 et 5.9, nous obtenons l'expression analytique de la tension du bridge V_{thb} :

$$V_{thb} = (C_{ox} \cdot C_{sc} + C_{box} \cdot C_{ox})^{-1} \left\{ \frac{C_{ox} \cdot C_{sc} + C_{box} \cdot C_{sc} + C_{box} \cdot C_{ox}}{C_{sc}} [\Psi_{sb_{th}} (C_{sc} + C_{box}) - (V_{gb} + \delta) C_{box}] - \gamma - V_{gb} \cdot C_{box} \cdot C_{ox} \right\} + V_{FM-f}. \quad (5.14)$$

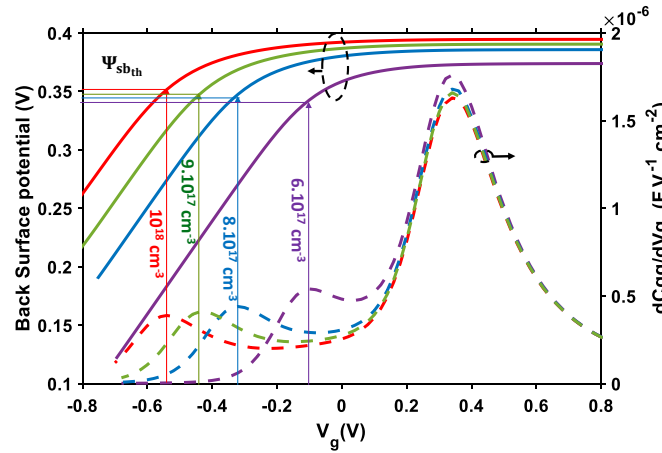


Figure 5.9 : Potentiel à l'interface arrière (axe à gauche) en fonction de V_g et dérivée première des caractéristiques C_{gg} - V_g (axe à droite) extraits des simulations TCAD à différents dopages du bridge.

V.3.4. Validation de la tension seuil du bridge V_{thb} par TCAD

Dans cette partie, il est question de valider le modèle de tension seuil du bridge V_{thb} donnée par l'équation 5.14, lorsqu'on varie les paramètres technologiques de l'A2RAM : dopage et épaisseur du bridge, épaisseur totale du film de silicium et polarisation de la grille face arrière V_b . Pour cela nous avons fait des simulations TCAD petits signaux comme dans le Chapitre 4. Des caractéristiques C_{gg} - V_g obtenues, nous avons évalué leurs dérivées premières pour extraire la tension seuil du bridge. Les valeurs extraites de TCAD sont comparées à celles évaluées avec notre modèle (équation 5.14).

V.3.4.1. Variation du dopage du bridge (N_{bridge})

Sur la figure 5.10, nous comparons les extractions TCAD et les valeurs obtenues de notre modèle des variations de la tension seuil du bridge V_{thb} en fonction de l'épaisseur du bridge T_{bridge} pour différentes valeurs de concentration du bridge N_{bridge} . D'après nos études du Chapitre 4, nous savons qu'une fois l'épaisseur de l'oxyde enterré T_{BOX} et de la polarisation de la grille face arrière fixées, la gamme de valeurs des paramètres (T_{body} , T_{bridge} et N_{bridge}) de

l'A2RAM est limitée. Ceci est dû au fonctionnement (structure) très particulier de l'A2RAM. En effet, réduire le dopage du bridge (N_{bridge}) ou l'épaisseur du bridge (T_{bridge}) favorise la déplétion du bridge, car le couplage capacitif entre la grille face avant et arrière augmente. Dans cette condition, la tension seuil du bridge augmente et tends à se confondre à la valeur de la tension seuil du canal face avant, car les deux canaux s'activent simultanément. C'est pour cette raison que sur la figure 5.10, lorsque l'on réduit le dopage du bridge ($N_{\text{bridge}} \leq 10^{18} \text{ cm}^{-3}$ pour $T_{\text{bridge}} < 10 \text{ nm}$), notre modèle de la tension seuil du bridge V_{thb} ne peut plus être validé.

En revanche, lorsque nous augmentons N_{bridge} ou T_{bridge} , le bridge reste en conduction, se comportant comme une interface arrière en régime d'accumulation, car le champ électrique vertical induit par la grille face avant nécessaire pour la désertion du bridge est élevé. Ainsi, avant d'atteindre le régime de désertion dans le bridge, des trous sont accumulés à l'interface face avant (en dessous de l'oxyde de grille) dans le body, faisant écran au couplage capacitif entre les grilles face avant et arrière. Il n'est plus possible de définir une tension de seuil du bridge dans ces conditions. C'est pourquoi, sur la figure 5.10, pour $N_{\text{bridge}} = 1.5 \times 10^{18} \text{ cm}^{-3}$ et $T_{\text{bridge}} > 18 \text{ nm}$, notre model n'est plus valable.

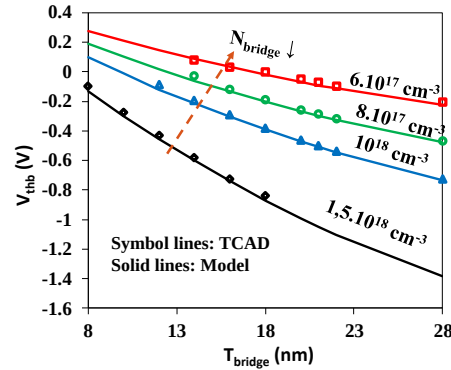


Figure 4.10 : Comparaison des valeurs de V_{thb} extraites avec TCAD ou calculées avec l'équation 4.14 en fonction de l'épaisseur du bridge pour une variations du dopage du bridge.

V.3.4.2. Variation de l'épaisseur totale du film de silicium (T_{sc})

Sur la figure 5.11, nous comparons les extractions TCAD avec les valeurs obtenues de notre modèle des variations de la tension seuil du bridge V_{thb} en fonction de l'épaisseur du bridge T_{bridge} pour différentes valeurs d'épaisseurs du film de silicium T_{sc} . Notons que pour la figure 11, nous considérons un dopage du bridge N_{bridge} fixe égal à 10^{18} cm^{-3} . Les observations de la figure 5.10 restent toujours valables pour la figure 5.11 ; la réduction de l'épaisseur du film de silicium favorise la désertion du bridge. C'est pourquoi sur la figure 5.11 lorsque l'épaisseur du film de silicium est inférieure à 30 nm, nous observons un léger décalage entre notre modèle et les extractions TCAD.

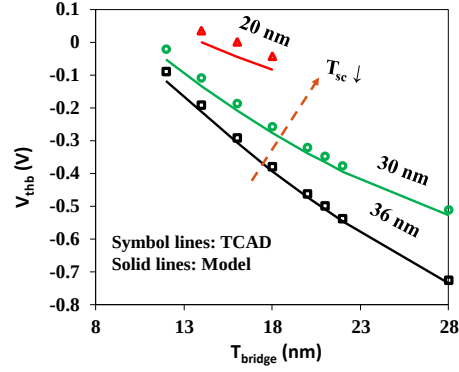


Figure 5.11 : Comparaison des valeurs de V_{thb} extraites avec TCAD ou calculées avec l'équation 5.14 en fonction de l'épaisseur du bridge pour une variation de l'épaisseur totale du film de silicium T_{sc} avec $N_{bridge} = 10^{18} \text{ cm}^{-3}$.

V.3.4.3. Variation de la polarisation de grille face arrière (V_b)

Sur la figure 5.12, nous comparons les extractions TCAD avec les valeurs obtenues de notre modèle des variations de la tension seuil du bridge V_{thb} en fonction de l'épaisseur du bridge T_{bridge} pour différentes valeurs de polarisations de la grille face arrière V_b . Ici, nous avons fixé le dopage du bridge N_{bridge} à $1.5 \times 10^{18} \text{ cm}^{-3}$, l'épaisseur de l'oxyde enterré est toujours égale 10 nm. Si la tension V_b appliquée sur la grille face arrière est trop élevée ($> 0 \text{ V}$), l'interface arrière reste en régime de forte accumulation et le fonctionnement de l'A2RAM est modifié, de sorte que V_{thb} n'est plus défini pour $V_b \geq 300 \text{ mV}$ et $T_{bridge} > 10 \text{ nm}$. De même, pour des polarisations V_b inférieures à 300 mV, lorsqu'on augmente l'épaisseur du bridge, il est de moins en moins facile de déserrer le bridge comme nous l'avons aussi souligné sur la figure 10. De ces observations, il s'ensuit que l'équation 5.14 est valable dans une gamme limitée mais réaliste de structures A2RAM.

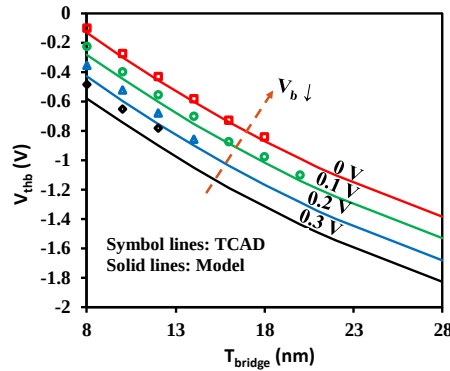


Figure 5.12 : Comparaison des valeurs de V_{thb} extraites avec TCAD ou calculées avec l'équation 5.14 en fonction de l'épaisseur du bridge en variant la tension de la grille face arrière V_b avec $N_{bridge} = 1.5 \times 10^{18} \text{ cm}^{-3}$.

V.4. Modèle compact DC de l'A2RAM

Nous allons à présent construire le modèle complet quasi statique de l'A2RAM par son macro modèle : un modèle compact d'un transistor et le modèle compact de la résistance du bridge branchés en parallèle (figure 5.2). Le courant de drain total de l'A2RAM sera la somme des courants du transistor et de la résistance du bridge.

V.4.1. Extraction carte modèle du transistor sur substrat SOI

En ce qui concerne le modèle compact du transistor, nous utilisons MASTAR VA [Lacord 14], et nous avons fait l'extraction de la carte modèle à partir de simulations classiques TCAD de transistor SOI. Pour cela, nous avons lancé des simulations TCAD comme sur la figure 5.3, à différentes tensions de drain. Puisque le modèle quasi statique de l'A2RAM sera utilisé pour l'opération de lecture, nous avons limité la tension de drain V_d à 200 mV. De plus, nous avons désactivé dans nos simulations les mécanismes de génération tels que l'ionisation par impact et la génération par effet tunnel bande à bande car le modèle MASTAR VA ne les prend pas en compte. Ainsi, la méthodologie de calibration consiste à fixer les bonnes valeurs des paramètres qui influencent la mobilité, la tension seuil, les effets de canaux courts (SEC) et le DIBL (Drain Induced Barrier Lowering). Sur la figure 5.13.a en échelle linéaire et en échelle logarithmique sur figure 5.13.b, nous rapportons la caractéristique transfert typique I_d - V_g du modèle MASTAR VA en comparaison avec les extractions TCAD. Globalement, nous avons une bonne correspondance entre le modèle et les extractions TCAD, bien qu'en régime d'inversion lorsque la tension de drain V_d est supérieure à 50 mV il y'a un léger décalage entre le modèle MASTAR VA et les simulations TCAD.

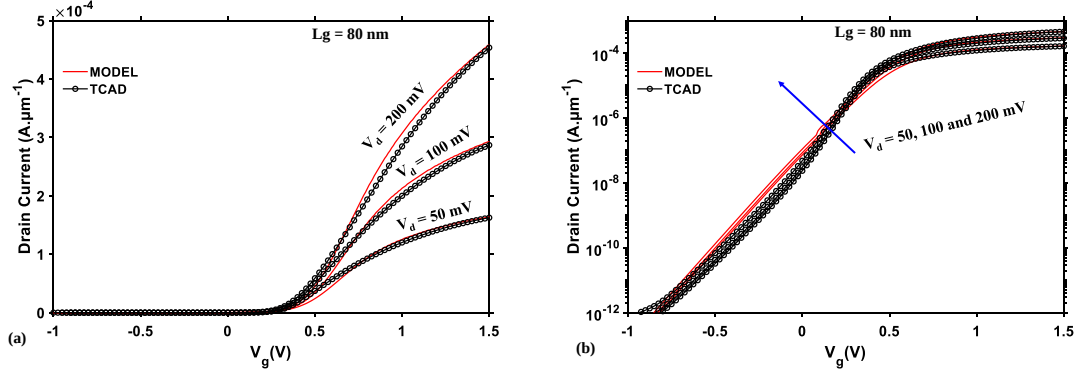


Figure 5.13 : Comparaison des courbes I_d - V_g obtenues par simulation TCAD et simulation ELDO avec le modèle MASTAR VA: (a) en échelle linéaire et (b) en échelle logarithmique pour une longueur de grille $L_g = 80$ nm.

V.4.2. Validation du modèle quasi statique de l'A2RAM

Nous avons validé le modèle MASTAR VA utilisé pour le modèle du transistor SOI ; et la modélisation de la résistance variable a déjà été validée sur la figure 5.5. Il reste encore à introduire le modèle de la tension seuil du bridge V_{thb} dans l'équation 5.1. Nous pouvons maintenant valider la modélisation de la caractéristique quasi statique de la A2RAM implémentée dans ELDO. Pour valider notre approche, nous avons extrait du simulateur SPICE ELDO, les caractéristiques statiques de notre modèle compact DC de l'A2RAM et

comparé avec les simulations TCAD. Sur la figure 5.14, la validation est faite pour une structure de longueur 80 nm et pour des tensions de drain limitées à 200 mV. Nous pouvons noter une bonne correspondance entre le modèle et les prédictions TCAD ; le décalage observé en inversion est dû en partie de notre calibration du modèle MASTAR VA (figure 5.13).

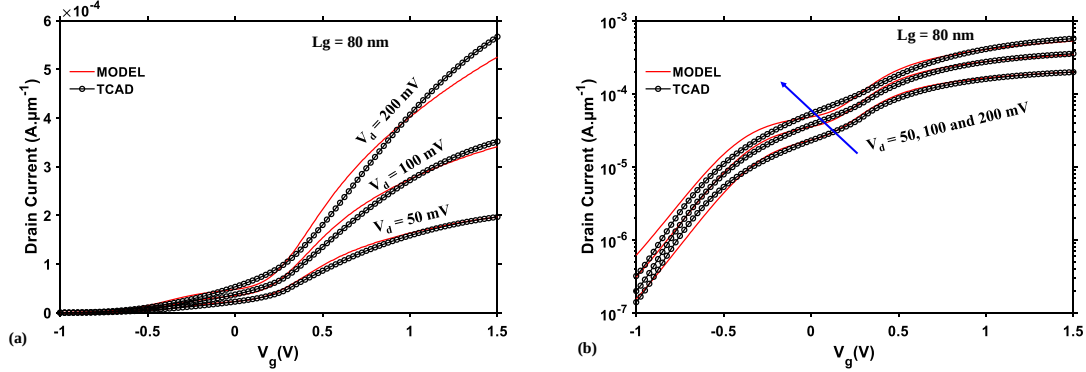


Figure 5.14 : Comparaison des courbes I_d - V_g obtenues par TCAD et par simulation ELDO avec le macro modèle DC de l'A2RAM : (a) en échelle linéaire et (b) en échelle logarithmique pour une structure A2RAM avec $L_g = 80$ nm.

V.5. Modélisation d'opérations mémoires de l'A2RAM

Dans cette partie, nous allons discuter de la modélisation des opérations mémoires de l'A2RAM (figure 2.1 avec $L_g = 80$ nm) comme décrit sur la figure 2.2 du Chapitre 2.

V.5.1. Circuit équivalent : stratégie de modélisation de l'effacement et de l'écriture

Nous allons tout d'abord décrire le comportement attendu du circuit équivalent qui modélise les états mémoires ('1' et '0'). Partant du motif de tension de la figure 5.15.a, la réponse attendue de notre circuit équivalent est présentée sur la figure 5.15.b : sa sortie STATE en fonction du temps. La variable STATE doit changer de valeur (d'état) seulement lorsqu'il y'a une opération d'effacement ou de lecture comme nous pouvons le noter sur la figure 5.15.b. L'implémentation du circuit équivalent est identique à l'approche utilisée dans [Martinie 17], et est présenté sur la figure 5.16. Il est composé d'un comparateur à deux résistances (R_{m1} et R_{m2}) dont les valeurs dépendent des tensions appliquées sur la grille face avant et sur le drain de la mémoire. Pour stocker la valeur de l'état mémoire correspondante : l'état '0' et l'état '1' ; les résistances R_{m1} et R_{m2} sont combinées à une capacité C_s afin de construire un circuit RC. Le circuit RC est alimenté par 0 V et 1 V pour que la sortie du circuit corresponde à la valeur de l'état mémoire (0 ou 1). Les valeurs des paramètres R_{m1} , R_{m2} et C_s de notre circuit équivalent sont choisies de façon à garantir l'écriture et l'effacement pour un temps de 300 ns. Pour notre étude, nous avons fixé la valeur de $C_s = 5.10^{-7}$ F, R_{m1} et R_{m2} sont égaux et oscillent entre 0 Ω et $10^3 \Omega$.

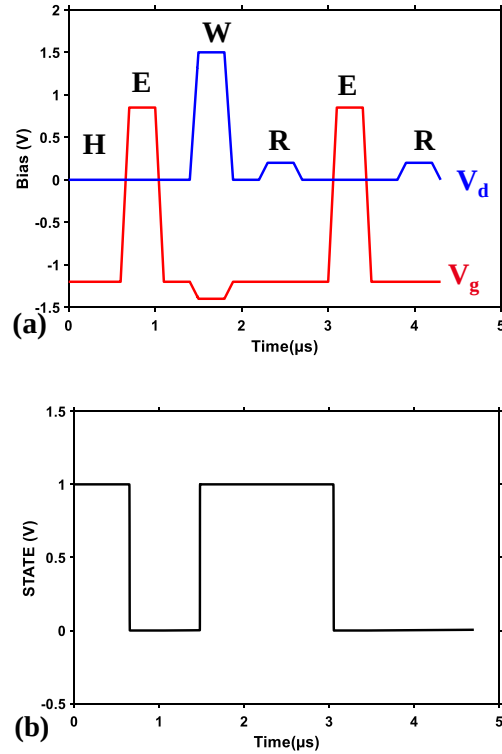


Figure 5.15 : a) Les tensions de drain V_d et de la grille face avant V_g en fonction du temps pour la séquence E-W-R avec une largeur d'impulsion de 300 ns et un front montant/descendant de 100 ns. (b) Variation de STATE correspondante obtenue par simulation Eldo du comparateur de la figure 5.16.

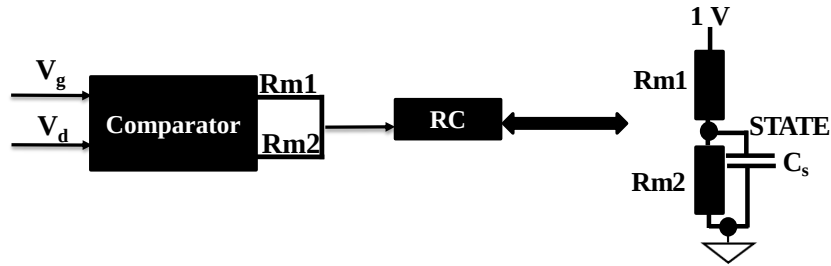


Figure 5.16 : Schéma du circuit équivalent décrivant l'écriture et l'effacement de la mémoire.

Cependant, si nous voulons que STATE varie seulement lorsque l'on écrit ou efface effectivement dans la cellule A2RAM, nous devons borner les polarisations qui garantissent les conditions d'écriture et d'effacement. En utilisant le même chronogramme de tensions de la figure 5.17.a, nous avons fait des simulations TCAD pour évaluer les couples de tensions (V_{gW1} , V_{dW1}) qui garantissent un courant de lecture I_1 de l'état '1' élevé avec des temps d'écriture et de lecture de 300 ns. Les tendances de la variation de I_1 avec V_{gW1} et V_{dW1} ont déjà été discutées dans le Chapitre 2. Comme nous pouvons constater sur la figure 5.17.a, même en imposant un critère sur la valeur minimale du courant lu de l'état '1' I_1 ($6\mu A.\mu m^{-1}$)

assurant une bonne fonctionnalité sous matrice, plusieurs couples de tensions (V_{gW1} , V_{dW1}) remplissent ce critère. Mais ici, pour simplifier l'implémentation de notre modèle, nous allons considérer que l'état '1' est écrit avec $V_{dW1} \geq 1$ V et $V_{gW1} \leq -0.5$ V.

Sur la figure 5.17.b, une méthodologie similaire est appliquée au courant de lecture de l'état '0' I_0 qui est tracé en fonction de V_{gErase} . Pour assurer une bonne marge entre I_1 et I_0 , c'est-à-dire une fenêtre de programmation $I_1/I_0 > 40$ pour ce travail, nous considérons que la cellule est effacée lorsque $V_{gErase} \geq -0.4$ V.

Après avoir défini les conditions limites de tension de la grille face avant et du drain qui garantissent la programmation de l'état mémoire, nous étudions maintenant le comportement attendu du circuit équivalent :

- R_{m1} doit être égal à 0Ω lorsque $(V_d - V_g) / (V_{dW1} - V_{gW1}) = 1$. Si ce n'est pas le cas, sa valeur est de $10^3 \Omega$.
- R_{m2} passe à 0Ω lorsque $V_g/V_{gErase} = 0$ et/ou $(V_d - V_g)/(V_{dW1} - V_{gW1}) = 1$; sinon, sa valeur est de $10^3 \Omega$.

Pendant la programmation de l'état '1', les résistances R_{m1} et R_{m2} sont égales à 0 ; la charge est stockée dans la capacité C_s , la variable STATE bascule donc à 1. Pour l'état '0', R_{m1} est à $10^3 \Omega$ et R_{m2} à 0Ω ; C_s est déchargé et STATE passe à 0.

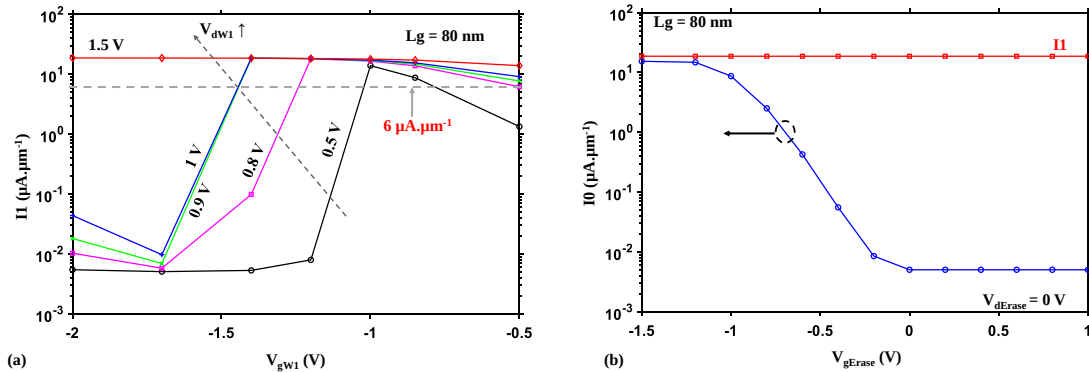


Figure 5.17 : (a) Courant de lecture de l'état '1' en fonction de la tension d'écriture de la grille face avant V_{gW1} pour différentes tensions d'écriture de drain V_{dW1} . La ligne en pointillé indique le courant de drain minimal assurant la fonctionnalité sous matrice. (b) Courant de l'état '0' en fonction de la tension d'effacement de la grille face avant V_{gErase} , le courant de lecture de l'état '1' étant toujours fixe. Les deux simulations des figures 5.17.a et 5.17.b sont extraites de la simulation TCAD sur la structure montrée à la figure 2.1 avec $L_g = 80$ nm.

V.5.2. Modélisation de l'opération de lecture de l'information : stratégie

Nous allons tout d'abord étudier la façon dont le modèle compact DC de l'A2RAM va interagir avec le circuit équivalent modélisant les états mémoires ('1' ou '0') pour pouvoir modéliser

la lecture de l'information. Pour cela, nous avons extrait des simulations TCAD les courants lus des états '1' et '0' en faisant varier la tension de grille face avant de lecture comme indiqué sur la figure 5.18.a de -2 V à 1.5 V.

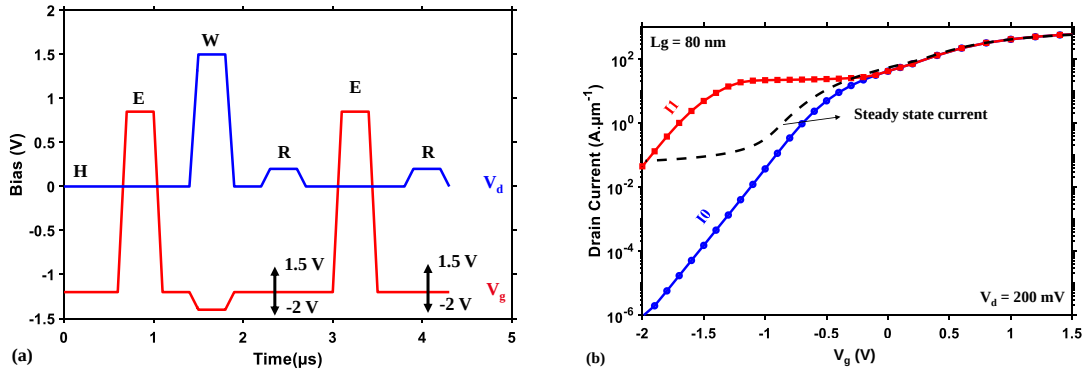


Figure 5.18 : (a) Motif de tensions appliqué sur la grille face avant V_g et sur le drain V_d pour la programmation et la lecture des états mémoires. (b) Courant des états lus I1 et I0 comparés au courant de la caractéristique statique de l'A2RAM en fonction de la tension de grille face avant durant la lecture de l'information, extraits des simulations TCAD.

Sur la figure 5.18.b, nous comparons les courants lus des états '1' et '0' avec le courant de drain de la caractéristique statique de l'A2RAM tous extraits de simulations TCAD (au même $V_d = 200$ mV). La caractéristique du courant de lecture de l'état '0' est similaire à celle de la caractéristique statique, mais nous avons un changement de pente pour des polarisations de grille face avant négatives ($V_g < 0$ V). Ceci est causé par le fait que la cellule est hors équilibre après la programmation de l'état '0'.

La caractéristique du courant de lecture de l'état '1' est semblable à la caractéristique de lecture de l'état '0', mais à une tension de seuil du bridge V_{thb} inférieure. Pour la mise en œuvre du modèle, nous considérons que la valeur du courant de lecture de l'état '0' peut être déduite de la caractéristique statique en ajustant le facteur de lissage '*smooth*' et la valeur de la résistance R_∞ lorsque le film de silicium est totalement déserté. Ensuite, la valeur du courant de lecture de l'état '1' est déduite de celle de la caractéristique du courant lu de l'état '0' avec V_{thb} donné par l'équation 5.15.

$$V_{thb} = V_{thb-DC} - (STATE * SWMEMV) \quad (5.15)$$

V_{thb-DC} est la valeur de la tension de seuil du bridge en régime statique donnée dans l'équation 5.14, $SWMEMV$ est la fenêtre mémoire extraite des simulations TCAD ou des données expérimentales, et la valeur $STATE$, donnée par le circuit équivalent, représente l'état actuel de la mémoire, qui peut être égal à '1' ou à '0'.

La figure 5.19.b montre le circuit complet qui modélise les opérations mémoires dans la cellule A2RAM. La sortie du circuit équivalent est connectée au modèle compact DC de l'A2RAM via un 'signal' interne : $STATE$. La valeur de $STATE$ changera la valeur de V_{thb} comme modélisée par l'équation 5.15 qui modulera la valeur du courant en fonction de l'état mémoire ('1' ou '0').

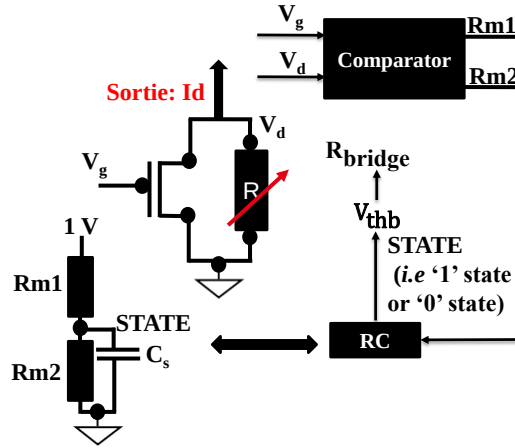


Figure 5.19 : Schéma montrant la stratégie de la macro-modélisation de l'A2RAM qui sera par la suite implémenté sous ELDO.

V.5.3. Extraction du paramètre SWMEMV

SWMEMV est le paramètre important qui garantit le bon fonctionnement de notre modèle. Ici, pour l'extraire nous partons toujours de nos simulations de la figure 5.18. Le but étant d'évaluer la fenêtre mémoire du dispositif que l'on veut modéliser c'est-à-dire le plus grand décalage de la tension de la grille face avant induit par l'effet mémoire. Ce décalage correspond à la plus grande différence des valeurs de courants lus des états '1' et '0'. Pour l'A2RAM, la lecture se fait pour des polarisations négatives de la grille face avant ($V_g < 0$ V). C'est pourquoi, sur la figure 5.20 nous avons repris les résultats de la figure 5.18.b en faisant un 'zoom' sur les polarisations négatives de la grille face avant où l'extraction de SWMEMV y est illustrée. Nous obtenons une valeur $SWMEMV = 1$ V. Cette valeur sera injectée dans l'équation 5.15, pour obtenir la tension seuil du bridge lorsqu'on écrira l'état '1'.

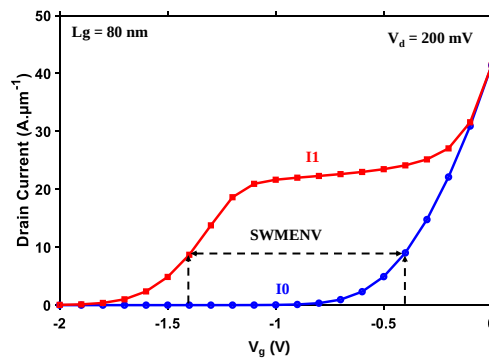


Figure 5.20 : Zoom sur $-2 \text{ V} \leq V_g \leq 0 \text{ V}$ de la figure 5.18.b, montrant comment extraire la fenêtre mémoire utile pour la modélisation des opérations mémoires.

V.5.4. Validation de la modélisation des opérations mémoires

Nous allons dans cette partie valider notre stratégie de modélisation des opérations mémoire de l'A2RAM illustrée sur la figure 5.19.

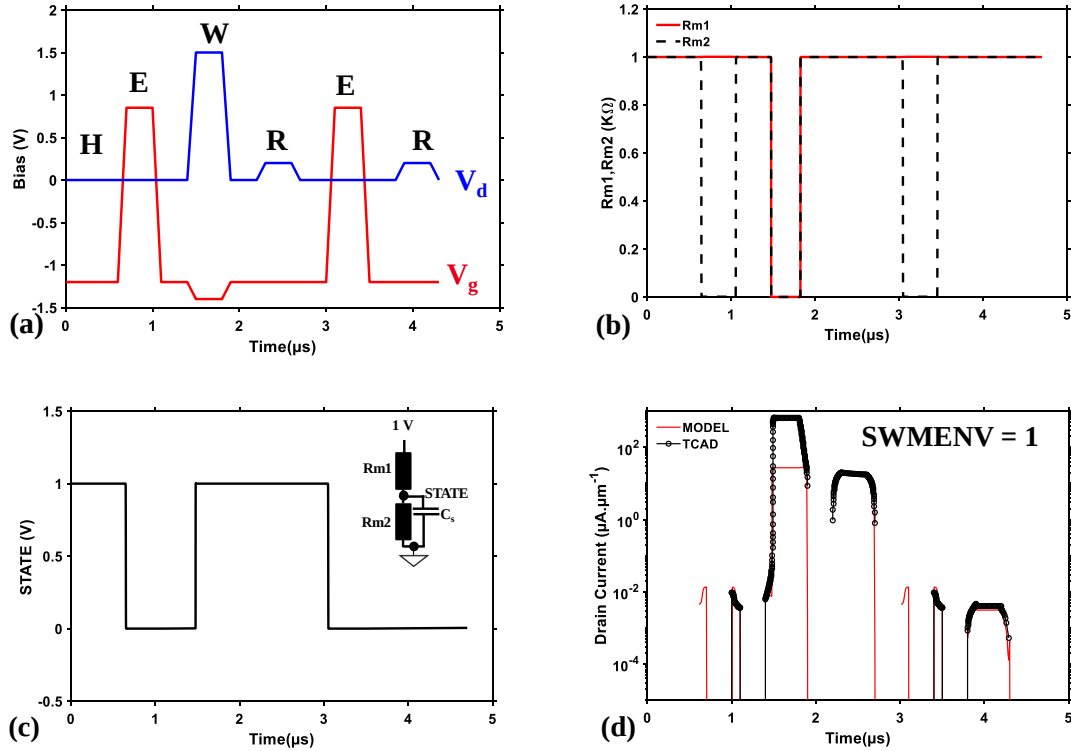


Figure 5.21 : a) Les tensions de drain V_d et de la grille face avant V_g en fonction du temps pour la séquence E-W-R avec une largeur d'impulsion de 300 ns et un front montant/descendant de 100 ns. (b) $Rm1$, $Rm2$ sont les sorties du comparateur sur la figure 5.19, et (d) la comparaison du courant de drain de notre macro-modèle avec l'extraction TCAD lorsque la séquence de tension de la figure 5.21.a est appliquée sur la cellule A2RAM de la figure 2.1.b avec $L_g = 80$ nm.

Sur la figure 5.21.a, nous avons une fois de plus notre chronogramme type des tensions de drain et de grille face avant utilisé pour les opérations mémoires durant ce travail de thèse (la source et la grille face arrière sont mises à la masse), présenté et expliqué dans le Chapitre 2, section 1. Nous allons appliquer ce motif de tension à notre modèle (figure 2.19) et les résultats obtenus seront comparés aux prédictions de TCAD.

Les figures 5.21.b et 5.21.c montrent les sorties des sous-circuits internes de notre modèle. Les sorties du comparateur (figure 5.21.b) et du circuit RC (figure 5.21.c) sont cohérentes avec les descriptions faites ci-dessus: la variable STATE passe à 0 après l'initialisation de l'état mémoire (phase E de la figure 5.21.a), et à 1 après la programmation de l'état '1' et n'est pas affectée par les opérations de maintien (H) et lecture (R) qui suivent ; il en est de même pour l'état '0'.

Sur la figure 5.21.d, nous comparons les courants de drain obtenus avec notre modèle et celui issu des extractions TCAD lorsqu'on applique le chronogramme de la figure 5.21.a sur la cellule A2RAM. Nous pouvons remarquer que le courant de drain pendant la programmation à l'état '1' n'est pas décrit avec grande précision par le modèle. Cela était prévu car la programmation est définie par le comparateur RC et n'inclut pas les mécanismes physiques qui entrent en jeu durant la programmation de l'information. Pour les autres opérations, nous observons une bonne correspondance entre TCAD et le modèle, en particulier pour les courants de lecture I1 et I0 des états lus '1' et '0' qui sont les données importantes pour évaluer les performances mémoires.

L'étude de la figure 5.21, nous a permis de valider notre approche macro-modèle compacte de la cellule A2RAM. Nous allons à présent l'utiliser dans un simulateur de circuit de type SPICE : ELDO.

V.5.5. Etude des opérations mémoires d'A2RAM dans une matrice 2x2: simulation SPICE

Le but ici est de montrer un exemple d'utilisation de notre modèle compact de l'A2RAM. Nous considérons une matrice 2x2 (schéma illustré sur la figure 5.22) composée de quatre points mémoire A2RAM A1, A2, A3 et A4 (la structure de la cellule est décrite sur la figure 2.1, avec $L_g = 80$ nm). L'accès de chaque ligne de mot (Word Line, WL_i avec $i = 0$ ou 1) est contrôlé par un NMOSFET de la technologie 28FDSOI [Planes 12]: si la ligne de mot est sélectionnée (une impulsion positive est appliquée sur WL_i). Elle connecte l'électrode source (S_i avec $i = 0$ ou 1) de chaque cellule A2RAM à la masse. Sur une même ligne, les électrodes des grilles face avant des points mémoires A2RAM sont connectées à la ligne de grille (Gate Line, GL_i avec $i = 0$ ou 1). Sur une même colonne, les électrodes des drains des points mémoires A2RAM sont connectées à la ligne de bit (Bit Line, BL_i avec $i = 0$ ou 1).

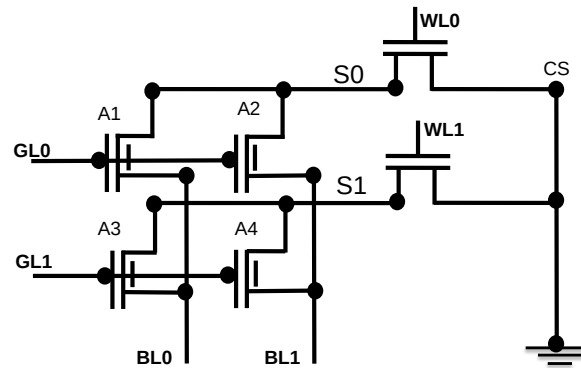


Figure 5.22 : Schéma d'une matrice 2x2 de points A2RAM (A1, A2, A3 et A4) utilisant deux transistors MOS de sélection. WL = ligne de mot BL : ligne de bit, GL: ligne de grille et S: source.

Les figures 5.23.a et 5.23.b décrivent les tensions appliquées sur chaque nœud pendant les opérations de la matrice. Le motif de la figure 5.23.a est utilisé pour contrôler les transistors d'accès ; le motif en ligne continue rouge est appliquée sur la WL0 et la ligne pointillée noire sur la WL1. Celui de la figure 5.23.b est appliqué sur les cellules A2RAM. Sur l'axe de gauche,

les lignes continues représentent les tensions appliquées sur la GL0 (rouge) et la GL1 (noir). Sur l'axe de droite, les lignes pointillées représentent les tensions appliquées sur la BL0 (rouge) et la BL1 (noir).

Dans cette étude, l'effacement est effectué sur une ligne complète tandis que la programmation et la lecture de l'information sont effectuées cellule par cellule pour une ligne donnée. Puisque nous voulons écrire le mot 1001 dans la matrice de la figure 5.22 correspondant aux données des cellules A1, A2, A3 et A4 respectivement ; après l'initialisation complète de la matrice, nous programmons l'état '1' sur A1 (W: A1) et par la suite nous programmons l'état '1' dans A4 (W: A4). Afin de vérifier les opérations de programmation précédentes, nous lisons la matrice complète (ligne par ligne) : lecture de l'information dans A1 (R: A1), lecture de l'information dans A3 (R:A3), lecture de l'information dans A2 (R:A2), et lecture de l'information dans A4 (R:A4). Pour chaque opération, la largeur d'impulsion de chaque phase est de 300 ns et les temps de montée et de descente sont de 100 ns. Pour éviter l'écriture parasite dans les cellules partageant la même colonne, par exemple, lors de l'écriture de l'état '1' dans A1, GL1 est mis à -0.4 V, il en est de même durant l'écriture du '1' dans A4.

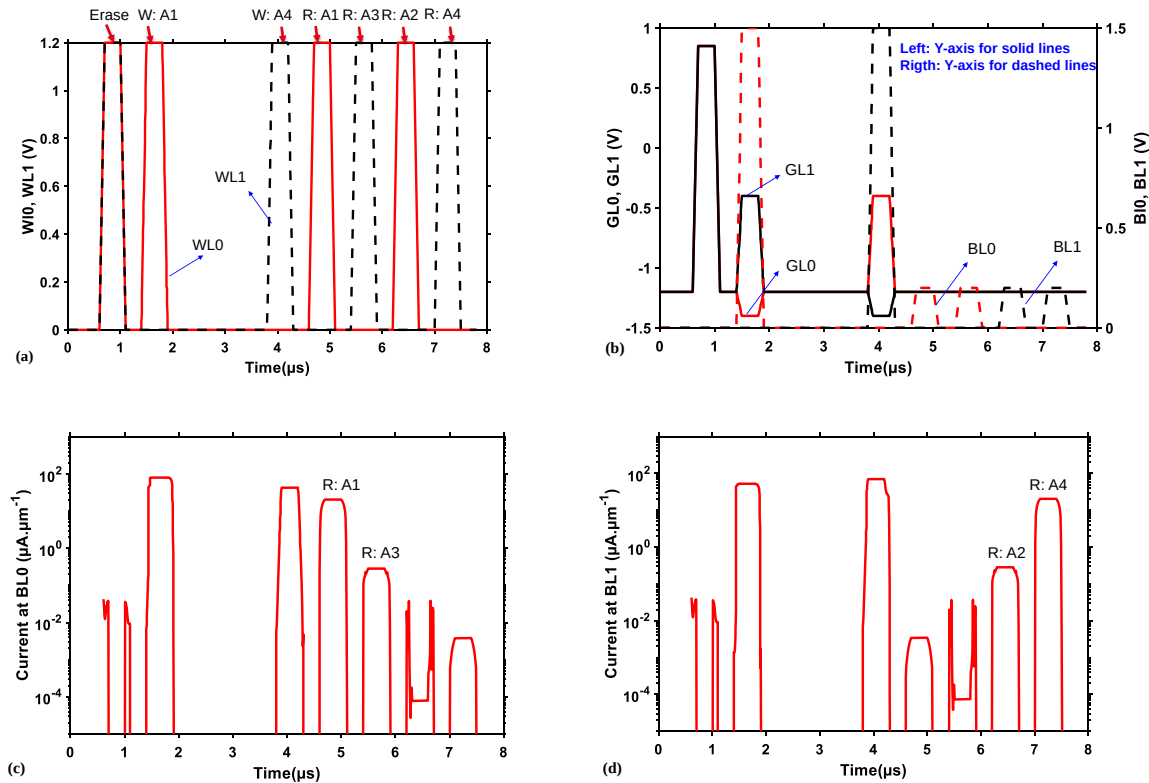


Figure 5.23 : Simulation SPICE de la matrice 2x2 de points A2RAM: motif de tensions appliquées sur (a) WL, (b) GL à gauche et BL à droite, pendant la séquence suivante: effacement de matrice complète - programmation de l'état '1' dans A1 - programmation de l'état '1' dans A4 - lecture de la matrice complète. Courant de drain en fonction du temps de chaque cellule A2RAM au cours de la séquence mémoire de 5.23.b à la (c) BL0 et (d) BL1.

Pour observer l'évolution des opérations de la matrice, nous avons extrait les courants sur chaque ligne de donnée (figure 5.23.c et figure 5.23.d) pendant toute la séquence de mémoire (figure 5.23.b). Comme prévu, dans la première colonne (BL0), nous lisons le courant d'état '1' I_1 (R: A1) dans A1 et l'état '0' $I_0 < 6\mu A.\mu m^{-1}$ (R: A3) en A3. De manière analogue, sur la deuxième colonne (BL1), $I_0 < 6\mu A.\mu m^{-1}$ est lu en A2 (R: A2) et I_1 en A4 (R: A4) ; au final nous avons bien le mot 1001 comme prévu. Ces résultats préliminaires démontrent d'une part la robustesse de notre modèle lorsqu'il est utilisé pour un cas réel. D'autre part, ils suggèrent aussi que notre modèle compact de la cellule A2RAM peut être facilement utilisé dans un simulateur de circuit SPICE, et donc implémenté par des concepteurs de circuits.

V.6. Conclusion

Dans ce travail, nous avons proposé la modélisation compacte de la cellule A2RAM. Le modèle repose sur la combinaison de la modélisation de la caractéristique statique de la cellule A2RAM et d'un circuit équivalent. La modélisation statique de l'A2RAM est obtenue par la combinaison d'un transistor sur substrat SOI et d'un modèle de résistance du bridge. Le circuit équivalent permet de modéliser la dynamique du stockage de l'information ('1' ou '0'), et sa sortie sert à moduler la valeur de la tension seuil du bridge, donc de la résistance du bridge. Grâce à la description analytique de la tension de seuil du bridge V_{thb} , nous avons pu utiliser la variation V_{thb} pour décrire l'opération de lecture. Ce modèle compact est calibré par TCAD, implémenté en Verilog-A et utilisable dans un simulateur SPICE, comme nous l'avons montré par l'étude des opérations matricielles A2RAM 2X2. Cependant, l'opération de maintien nécessaire à l'évaluation du temps de rétention de l'information reste à modéliser.

CHAPITRE VI :
FABRICATION D'UNE
CELLULE A2RAM
NANOFIL TRIPLE-GRILLES

Les études présentées dans ce chapitre font partie de la contribution du CEA-Leti au projet européen H2020 REMINDER [REM2020]. Le CEA-Leti avait 2 objectifs :

- démontrer un gain en performance apporté par une architecture 3D de la cellule A2RAM.
- démontrer le gain sur les performances de l'A2RAM apporté par l'intégration d'un matériau III-V tel que le silicium-germanium dans le body.

La fabrication de ce nouveau lot de mémoires a été suivie par les experts du laboratoire d'intégration du CEA-Leti. La contribution de ce travail de thèse porte sur le dimensionnement des paramètres technologiques, et sur la définition des étapes de procédé formation du bridge puis sur la caractérisation électrique des dispositifs fabriqués.

VI.1. Intégration de la cellule A2RAM de type 3D

Dans cette partie, nous présentons les étapes de fabrication d'une cellule 3D A2RAM. De nos études des chapitres 2, 3 et 4, nous savons que le profil de dopage du bridge est primordial. Le challenge est alors de définir les bonnes conditions d'implantation du bridge associées aux recuits adéquats.

VI.1.1. Etapes de fabrication de la cellule A2RAM 3D

Le procédé de fabrication de dispositifs A2RAM 3D repose sur une adaptation du procédé pour les nanofils du CEA-Leti [Barraud 13]. Sur la figure 6.1.a, nous rappelons les principales étapes de fabrication de transistors nanofils, et les figures 6.1.b et 6.1.c montrent les images TEM des structures nanofils standards.

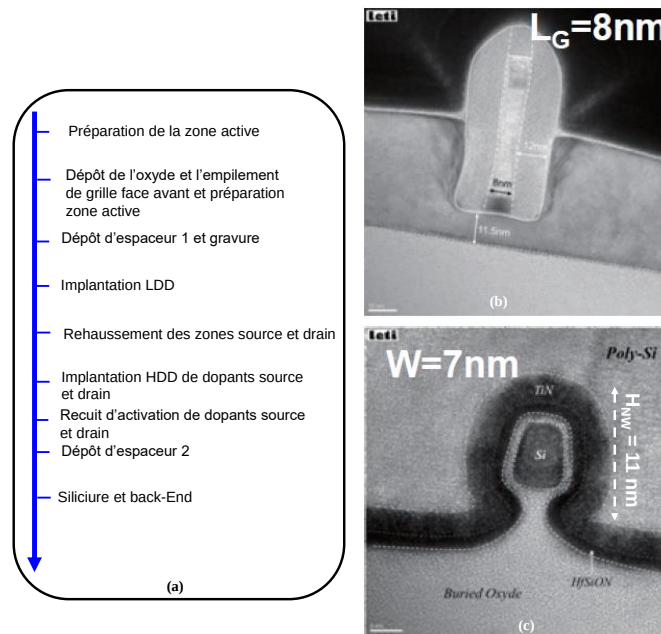


Figure 6.1 : (a) Procédé de fabrication de transistor nanofils standard au CEA-Leti, image TEM d'un dispositif nanofil sur substrat SOI issu du procédé de fabrication de la figure 6.1.a

(b) dans le sens de la longueur de la grille, et (c) dans le sens de la largeur de la grille qui met évidence la forme omega de la grille.

Ce procédé d'intégration a été modifié pour tenir compte des spécificités de la structure A2RAM, c'est-à-dire pour former le bridge entre source et drain, et contrôler son profil de dopage. Ainsi quatre nouvelles étapes seront ajoutées pour obtenir celui de la figure 6.2.

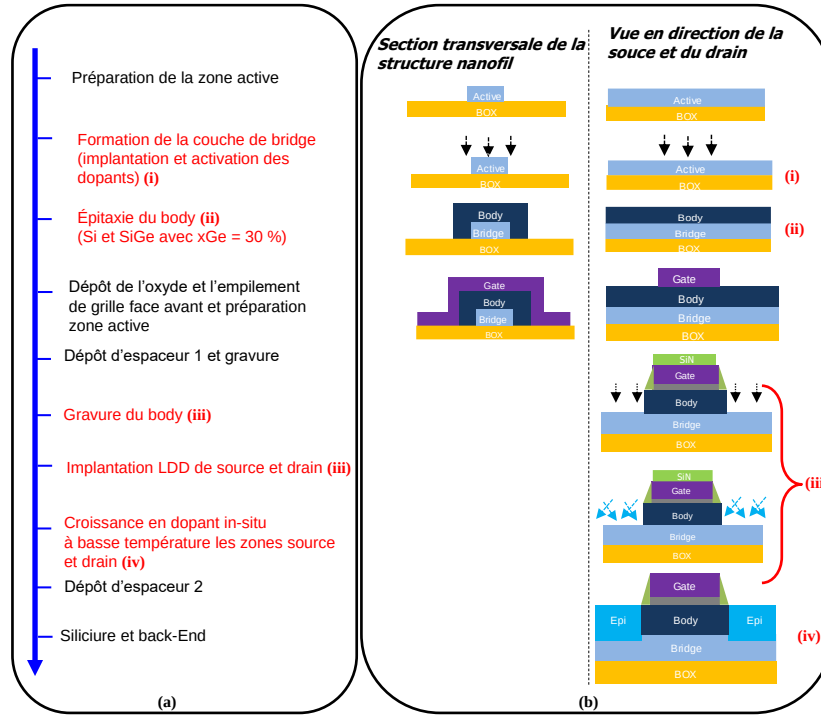


Figure 6.2 : Procédé de fabrication de structures nanofil sur substrat SOI sub-14 nm du CEA-Leti. (a) Procédé de fabrication modifié (en rouge) pour fabriquer les cellules A2RAM 3D, et (b) schémas illustratifs des étapes critiques de la fabrication de la cellule A2RAM 3D.

Entre la préparation de la couche active et le dépôt de l'empilement de la grille deux étapes sont rajoutées :

- Etape (i) : la formation de la couche du bridge par une implantation d'arsenic à faible dose suivie d'un recuit d'activation. Ces étapes ont été dimensionnées par simulation TCAD dans la section 2.4.
- Etape (ii) : après la formation du bridge, une épitaxie est effectuée pour former le body. Cette étape, qui a un budget thermique non négligeable pour la diffusion des dopants a été simulée en TCAD pour évaluer son impact sur le profil du bridge (section 2.4).

Ensuite, entre les formations des deux espaceurs il y a deux nouvelles étapes :

- Etape (iii) : la gravure du body suivie d'une implantation des zones LDD (*Lightly Doped Drain*) source et drain. La gravure du body a pour but de permettre un bon contact (créer un court-circuit) entre le bridge faiblement dopé de type n (10^{18} cm^{-3}), et les zones source et drain aussi dopées de type n à $2 \cdot 10^{20} \text{ cm}^{-3}$. La formation de LDD

permet d'assurer un dopage important à l'entrée de la grille pour maximiser le GIDL [shenck 09] donc la programmation. Son dimensionnement est présenté dans la section 2.2.

- Etape (iv) : le surélevèrent des zones source et drain par une épitaxie sélective dopée in-situ de type n à basse température. L'objectif est de remplacer l'implantation HDD (*Highly Doped Drain*) des zones source et drain du procédé de fabrication standard qui consiste à implanter une forte dose de dopants puis de les activer par un recuit haute température (1050 °C). Sachant que nous souhaitons contrôler le profil du bridge, c'est-à-dire limiter le plus possible la diffusion des dopants , il faut :
 - Limiter l'augmentation de la température d'activation des dopants source et drain à environ 735 °C, le but étant d'éviter autant que possible la diffusion des dopants du bridge dans le body.
 - Réussir à obtenir un fort dopage de type n de $2.10^{20} \text{ cm}^{-3}$ dans les zones source et drain avec une faible température de recuit.

Pour pouvoir obtenir ces résultats, nous avons utilisé l'expertise du CEA-Leti sur les procédés de fabrication à basse température utilisée pour l'intégration 3D séquentielle (CoolCube™) [Lu 16].

A la fin du procédé de fabrication, l'objectif est d'obtenir l'empilement vertical des différentes couches représenté sur la figure 6.3. Comme on peut le voir, l'empilement de la grille est constitué d'une fine couche intermédiaire de SiO_2 d'une épaisseur de 1 nm, de 1.9 nm de HfSiON , de 5 nm de TiN (métal de la grille définissant le travail de sortie), et du poly-silicium dopé à 10^{20} cm^{-3} d'une épaisseur de 50 nm comme définie dans [Barraud 13]. Le dimensionnement des épaisseurs du body (T_{body}) et du bridge (T_{bridge}) sera décrit dans la section 2.1.

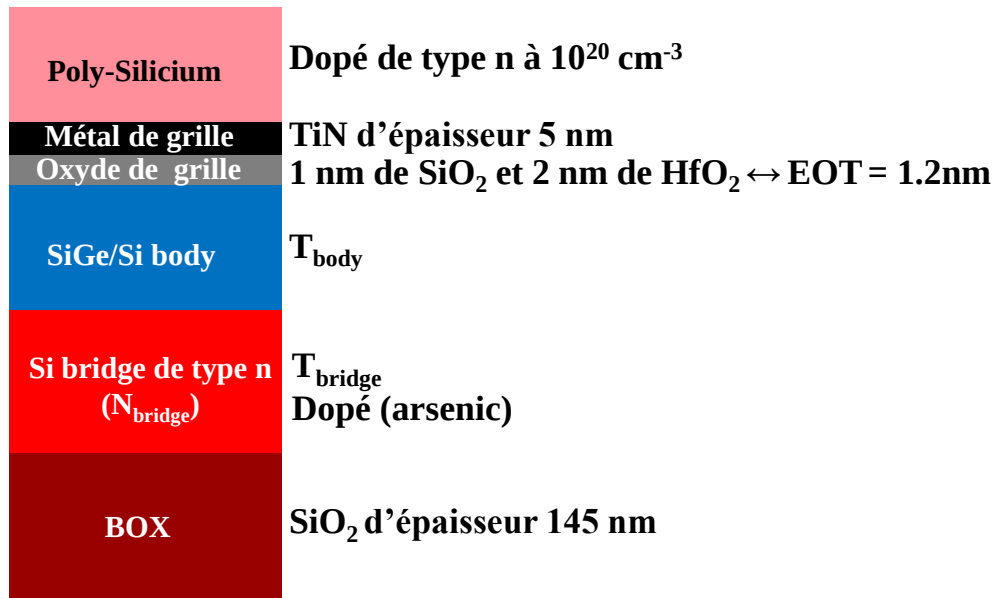


Figure 6.3 : Empilement vertical de la structure A2RAM 3D.

VI.2. Simulations prédictives TCAD

Dans cette partie, nous présentons les résultats de différentes simulations TCAD qui ont permis d'établir le cahier des charges pour la fabrication des dispositifs A2RAM en tenant compte :

- des contraintes technologiques imposées par le procédé de fabrication :
 - Epaisseur totale de silicium (body + bridge) ne peut pas être supérieure à 25nm
 - Réutilisation d'un jeu de masques, la longueur de grille minimale peut être ajustée par le procédé.
- des contraintes liées à la caractérisation électrique de performance mémoire : le courant doit être suffisamment fort pour pouvoir être lu. Les dispositifs devront respectés les critères suivants :
 - $I1 > 6\mu A.\mu m^{-1}$
 - $I1/I0 > 20$.

VI.2.1. Dimensionnement de la longueur de la grille et des épaisseurs du body et du bridge

Dans ce paragraphe, nous allons présenter les résultats de performances mémoires extraites des simulations TCAD attendues pour le choix des épaisseurs du film de silicium ($T_{body} + T_{bridge}$), et aussi des longueurs de grille (L_g). D'une part, ces résultats ont été obtenus en tenant compte des exigences du procédé de fabrication de la route nanofil du Leti qui nous impose une épaisseur maximale de la zone active de 25 nm. D'autre part, dans la section 2.5 du Chapitre 3, nous avons vu qu'avec un dopage du bridge de $3.10^{18} \text{ cm}^{-3}$, il n'était pas possible de viser les longueurs de grille L_g inférieures à 40 nm même avec une épaisseur de film de silicium de 20 nm, car $I1 = I0$. Ainsi, étant donné que nous visons une forte densité d'intégration, pour avoir des structures fonctionnelles avec $L_g < 40 \text{ nm}$ il faut réduire le dopage, c'est pourquoi pour nos simulations nous avons choisi comme dopage du bridge 10^{18} cm^{-3} .

Les simulations ont été effectuées avec la structure de la figure 6.4.a, avec un EOT de 1.12 nm, un BOX de 145 nm et un dopage du bridge $N_{bridge} = 10^{18} \text{ cm}^{-3}$. La longueur de grille L_g variera de 20 nm à 80 nm et l'épaisseur du bridge de 5 nm à 15 nm tout en conservant une épaisseur totale du film de silicium inférieure à 25 nm (tableau 6.1). Pour évaluer les performances mémoires ($I1$ et $I1/I0$), le chronogramme de tensions de la figure 6.4.b sera utilisé, il est similaire à ceux déjà utilisés dans les chapitres précédents.

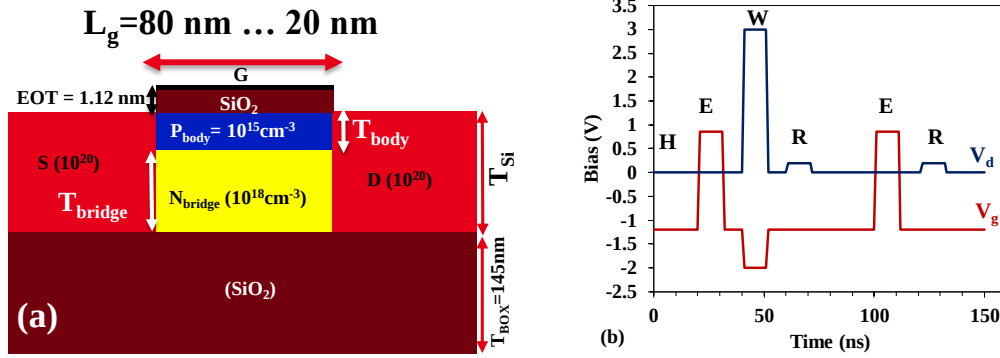


Figure 6.4 : (a) Structure A2RAM 2D qui nous a servie de base pour le dimensionnement des paramètres de l'A2RAM 3D. (b) Chronogramme de tensions du drain V_d et de la grille V_g utilisé pour la caractérisation mémoire.

$T_{\text{Si}} \text{ (nm)}$	$T_{\text{body}} \text{ (nm)}$	$T_{\text{bridge}} \text{ (nm)}$
25	5	20
	10	15
	15	10
	20	5
20	5	15
	10	10
	15	5
15	5	10
	10	5

Tableau 6.1: Variations des épaisseurs du film de silicium de la structure de la figure 6.4.a.

La figure 6.5 montre les évolutions du courant I_1 , et du ratio I_1/I_0 en fonction de L_g pour différentes épaisseurs du bridge lorsque l'épaisseur du film de silicium vaut 15 nm (figure 6.5.(a, b)), 20 nm (figure 6.5.(c, d)) et 25 nm (figure 6.5.(e, f)). Sur chaque figure est indiquée la limite de fonctionnalité : $I_1 > 6 \mu\text{A} \cdot \mu\text{m}^{-1}$ et $I_1/I_0 > 20$. On remarque d'abord qu'une épaisseur totale de silicium de 15nm n'est pas suffisante, car le critère de fonctionnalité est respecté juste pour $L_g < 30 \text{ nm}$. Ensuite, pour $T_{\text{Si}} \geq 20 \text{ nm}$, même si elles permettent des dispositifs fonctionnels en simulation, les épaisseurs de body et de bridge égales à 5 nm sont exclues car trop difficile à fabriquer de manière fiable.

Par conséquent les structures qui seront fabriquées auront les combinaisons d'empilement bridge/body suivantes :

- $T_{\text{Si}} = 20 \text{ nm}$ avec $T_{\text{body}} = 10 \text{ nm}$ et $T_{\text{bridge}} = 10 \text{ nm}$.
- $T_{\text{Si}} = 25 \text{ nm}$ avec $T_{\text{body}} = 15 \text{ nm}$ et $T_{\text{bridge}} = 10 \text{ nm}$.
- $T_{\text{Si}} = 25 \text{ nm}$ avec $T_{\text{body}} = 10 \text{ nm}$ et $T_{\text{bridge}} = 15 \text{ nm}$.

Ces empilements seront associés à une longueur de grille minimale de 30nm.

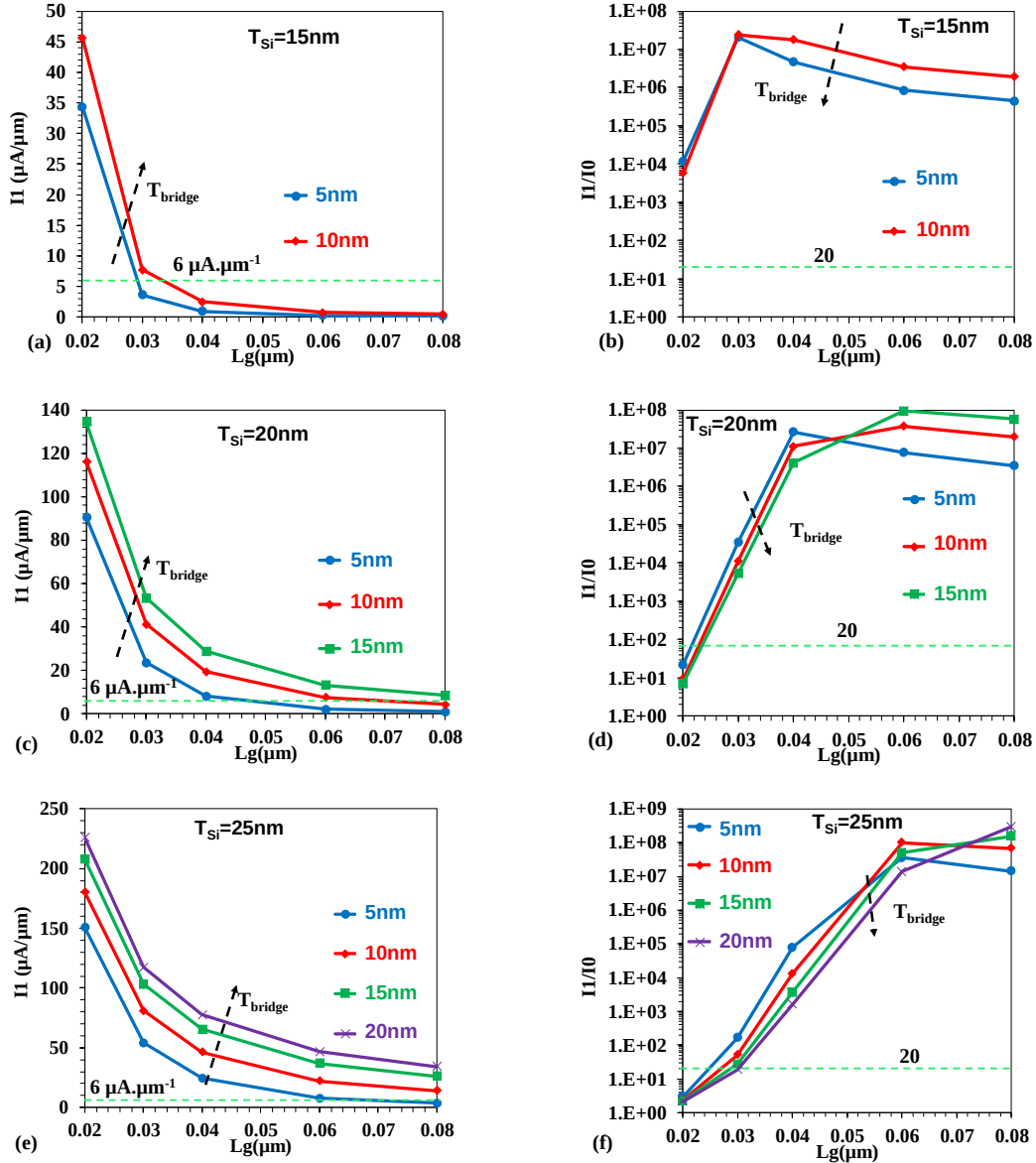


Figure 6.5 : Variations du courant lu de l'état '1' et de la fenêtre de programmation I_1/I_0 en fonction de la longueur de la grille pour différentes épaisseurs du bridge T_{bridge} lorsque : (a, b) $T_{\text{Si}} = 15\text{ nm}$, (c, d) $T_{\text{Si}} = 20\text{ nm}$, (e, f) $T_{\text{Si}} = 25\text{ nm}$.

VI.2.2. Dimensionnement des zones de recouvrement source et drain

Dans cette partie, nous évaluons l'impact des zones de recouvrement source et drain sur les performances mémoires (I_1 , I_0 et temps de rétention). La figure 6.6.a présente la structure simplifiée utilisée dans les simulations. Nous appelons 'Overlap' la longueur de la zone de recouvrement du dopage de la source (drain) en dessous de la grille et de l'espaceur. Nous avons fait varier l'Overlap de 0 nm à 12 nm, sachant que la longueur des espaceurs est la même et vaut 10 nm ; cette valeur a été choisie en accord avec l'équipe d'intégration du Leti.

Deux longueurs de grille sont considérées (65 nm et 80 nm), avec une épaisseur totale de film de silicium de 25 nm (épaisseur du bridge 10 nm et du body de 15 nm). Les autres paramètres technologiques sont les mêmes que ceux de la figure 6.4.a.

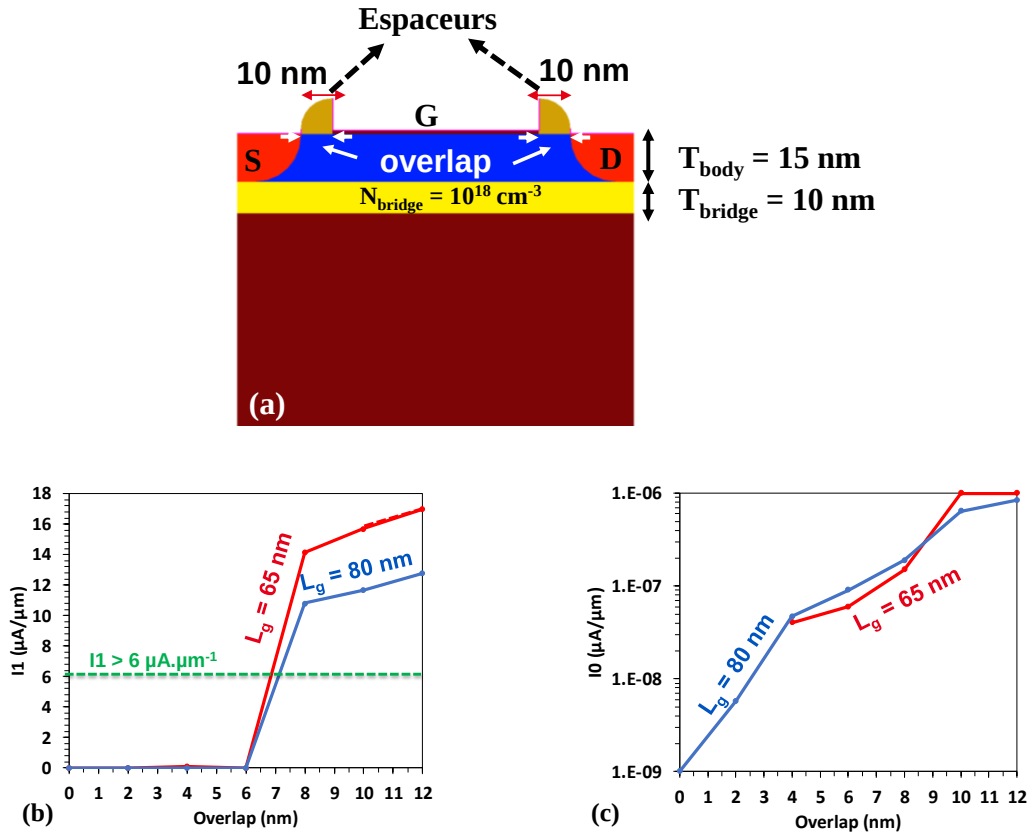


Figure 6.6 : (a) Structure A2RAM simulée pour évaluer l'impact des zones de recouvrement sur les performances mémoires. Variations du (b) courant lu de l'état '1' et (c) courant lu de l'état '0' en fonction de la longueur de la zone de recouvrement 'Overlap'.

Dans un premier temps, nous étudions l'impact de l'overlap sur I_1 et I_0 indiqués sur les figures 6.6.b et 6.6.c respectivement. Ces résultats sont obtenus avec le même motif de tension de la figure 6.4.b, dont le temps de programmation de l'information vaut 10 ns.

Sur la figure 6.6.b, nous pouvons noter qu'indépendamment de la longueur de la grille, l'amplitude de I_1 augmente avec la longueur de l'overlap. Ceci est dû au fait que le courant de programmation par GIDL augmente lorsque les zones de recouvrement source et drain augmentent. Nous avons confirmation de cette observation sur les cartographies de la densité de génération par effet tunnel bande à bande (B2B) durant la programmation de l'état '1' sur la figure 6.7 pour différents Overlap. Le taux de génération diminue avec la longueur d'overlap, ce qui justifie la réduction de I_1 pour Overlap courts.

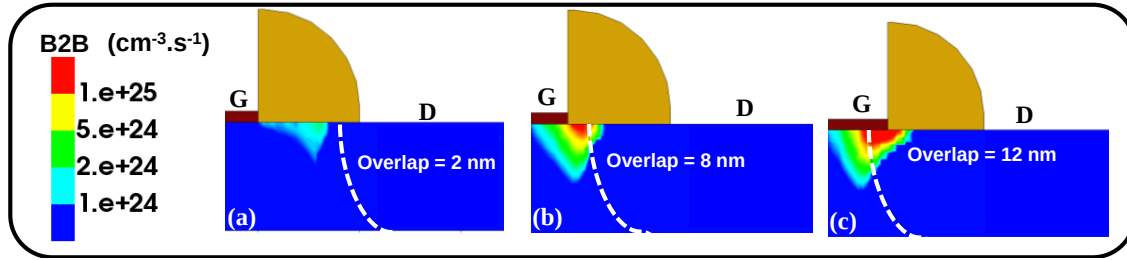


Figure 6.7 : Cartographie du taux de génération de la charge par effet tunnel bande à bande (B2B) durant l'écriture de l'état '1' pour trois valeurs d'Overlap : (a) 2 nm, (b) 8 nm, et (c) 12 nm.

Sur la figure 6.6.c, le courant I_0 augmente lorsque les zones de recouvrement augmentent indépendamment de la longueur de la grille. En effet, lorsque l'on augmente l'Overlap, la tension seuil du canal face avant (V_{thf}) est réduite à cause de la réduction de la longueur électrique du dispositif. Ceci a pour conséquence d'augmenter le courant I_{off} qui est essentiellement dû aux courants de fuite des diodes source/body et drain/body.

Donc, augmenter la zone de recouvrement améliore I_1 mais dégrade I_0 . Il convient alors d'évaluer le temps de rétention afin de déterminer le meilleur compromis. Nous avons étudié le temps de rétention pour une longueur de grille de 65 nm, et un overlap variant de 8 à 12 nm. Le protocole utilisé reste le même que celui discuté dans les chapitres 2 et 3. Sur la figure 6.8.a, nous présentons les variations du courant I_1 et I_0 durant une lecture continue pour différentes valeurs d'Overlap : 8 nm, 10 nm et 12 nm. Les extractions du temps de rétention sont montrées sur la figure 6.8.b ; le critère d'extraction tient compte de nos critères de fonctionnalité c'est-à-dire : $I_1 > 6 \mu A \cdot \mu m^{-1}$ et $I_1/I_0 > 20$. Globalement, le temps de rétention reste inférieur à la milliseconde quelle que soit la valeur de l'Overlap. Pour un Overlap < 10 nm, la faible amplitude de I_1 à l'équilibre thermodynamique réduit la valeur du temps de rétention. Lorsque l'Overlap ≥ 10 nm, l'augmentation de I_0 n'est pas responsable de la faible valeur du temps de rétention, mais c'est à cause de notre critère sur la valeur minimale de I_1 qui doit être toujours supérieure à $6 \mu A \cdot \mu m^{-1}$ au moment de l'extraction. Mais, si nous considérons comme unique critère de fonctionnalité la fenêtre de programmation ($I_1/I_0 > 20$), les variations du temps de rétention sur la figure 6.8.c montrent qu'augmenter la longueur de la zone de recouvrement dégrade le temps de rétention.

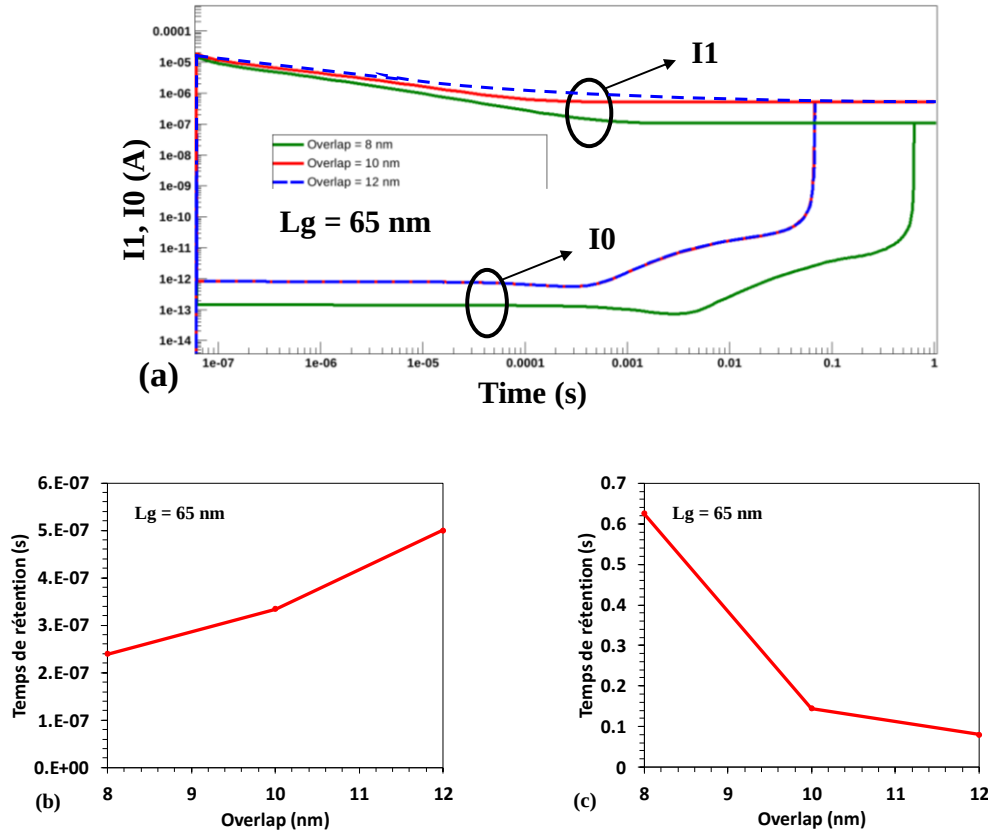


Figure 6.8 : (a) Variations des courant I_1 et I_0 durant une lecture continue pour évaluer le temps de rétention pour différentes longueurs d'Overlap. Evolution du temps de rétention extrait de la figure 6.8.a avec les critères : (b) $I_1 = 6$ A.μm $^{-1}$ et $I_1/I_0 = 20$, (c) $I_1/I_0 = 20$ seulement.

Des observations faites sur les figures 6.6 et 6.8, il est important d'avoir des zones de recouvrements dans nos structures pour maximiser le stockage de la charge, mais pas trop grandes pour ne pas dégrader le temps de rétention. Le meilleur compromis pour un espaceur de 10 nm est un Overlap de 8 nm. Pour mieux évaluer ces effets, nous avons aussi décidé d'inclure dans le lot des dispositifs des structures sans zone de recouvrement, c'est-à-dire sans implantation LDD.

VI.2.3. Evaluation de la concentration optimale du germanium dans SiGe

Il est question dans cette partie de confirmer les résultats que nous avons obtenus dans la section 3.2 du Chapitre 3. Il ressort de cette étude que la concentration du germanium idéale à intégrer dans le body est de 30%. L'un des avantages d'avoir du silicium-germanium dans le body est qu'il augmente la quantité de charges stockées et donc du courant lu de l'état '1'. Mais, nous avons une dégradation en terme du courant de fuite source-drain de l'A2RAM, ainsi le courant lu I_0 de l'état '0' augmente. Nous allons ici vérifier si l'intégration du SiGe dans le body de nos cellules 3D A2RAM pour les épaisseurs choisies garantit toujours une fonctionnalité mémoire, c'est-à-dire $I_1 > 6$ μA.μm $^{-1}$ et $I_1/I_0 > 0$.

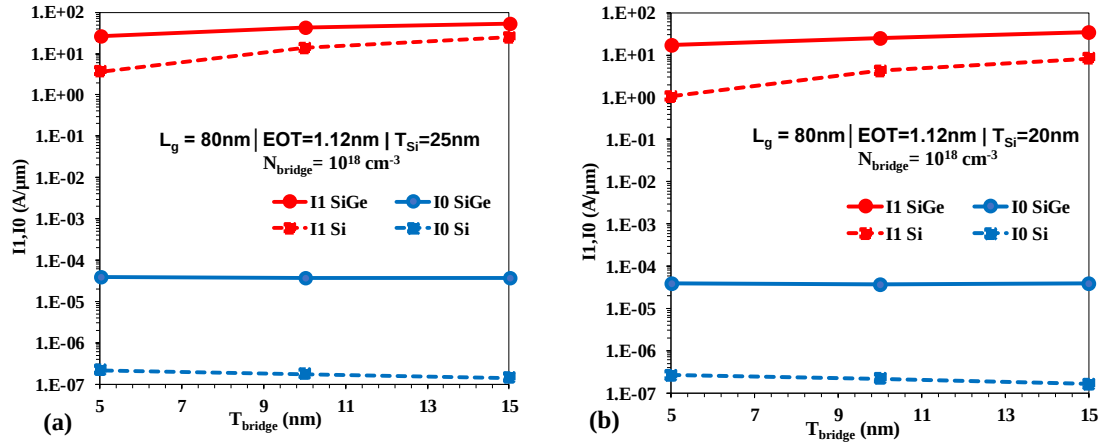


Figure 6.9 : Comparaisons des variations des courants lus des états '1' (I1) et '0' (I0) en fonction de l'épaisseur du bridge T_{bridge} pour une structure avec le body en SiGe et une autre avec le body en Si lorsque (a) $T_{\text{Si}} = 25\text{ nm}$, et (b) $T_{\text{Si}} = 20\text{ nm}$.

Les figures 6.9.a et 6.9.b montrent les variations des courants I1 et I0 en fonction de l'épaisseur du bridge lorsque les épaisseurs du film de silicium valent 25 nm (figure 6.9.a) et 20 nm (figure 6.9.b) pour deux structures dont l'une a un body en silicium (Si) et l'autre a un body en silicium-germanium (SiGe). Nous avons confirmation que l'intégration du SiGe dans le body augmente la valeur de I1 comparée à la structure en silicium. Cette augmentation de I1 est essentiellement dû à l'augmentation de la quantité de la charge stockée dans le body qui est supérieure à la structure totalement en silicium. Le courant I0 augmente aussi, mais les structures restent fonctionnelles car la fenêtre de programmation I1/I0 est toujours supérieure à 20.

VI.2.4. Dimensionnement de recette d'implantation du bridge

Comme nous l'avons mentionné précédemment, un des éléments importants pour la fabrication de mémoires A2RAM 3D est de contrôler le profil de dopage entre le body et le bridge. Le body à la fin du procédé de fabrication doit être 'intrinsèque' et le bridge avec un dopage d'environ 10^{18} cm^{-3} . Pour identifier les conditions idéales d'implantation du bridge et les températures de recuit pour l'activation des dopants de chaque phase, nous avons fait des simulations de procédé de fabrication via l'outil TCAD : Synopsys (Sentaurus) Process en utilisant les modèles d'implantation et de diffusion par défaut.

La figure 6.10 montre l'enchaînement d'étapes de formation du bridge :

- a) L'empilement qui sera implanté est constitué d'un BOX de 145nm, une couche de silicium de 15 ou 10 nm, une couche d'oxyde (SiO₂) de 1.5 nm.
- b) Implantation arsenic de dose 5.10^{12} cm^{-2} avec des variations d'énergie 12 keV à 20 keV.
- c) Une fois les dopants implantés et le retrait de la couche d'oxyde sacrificiel effectué, nous faisons un recuit de 10 minutes à 1050°C afin d'activer les dopants et uniformiser le profil dans la couche de silicium.

- d) Le body est formé par épitaxie de 10 ou 15nm. Afin d'évaluer le profil de dopage du bridge final, le recuit d'activation finale est simulé après l'épithaxie body (735°C – 15min) car les autres étapes n'ont pas d'effet sur la diffusion des dopants.

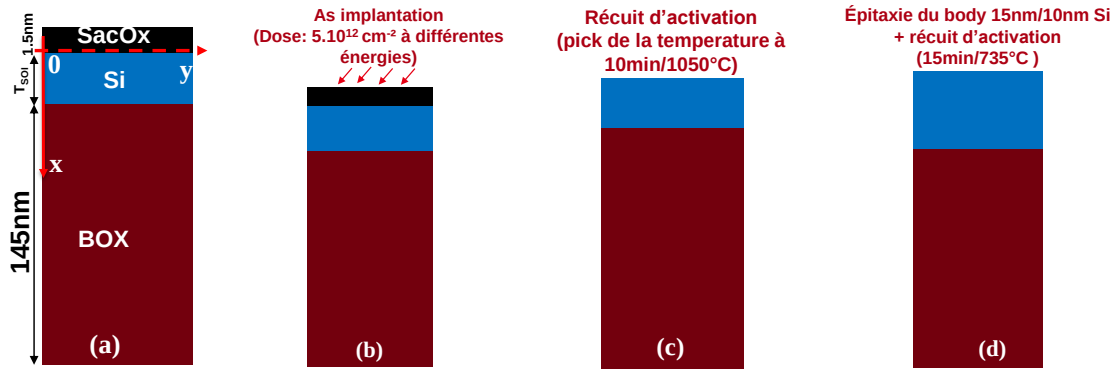


Figure 6.10 : Différentes étapes de simulation TCAD du procédé d'implantation et épithaxie du body ; (a) Structure simple utilisée pour la simulation avec un oxyde sacrificiel (SacOx) au-dessus de la zone active, (b) implantation d'arsenic pour dopage du bridge, (c) recuit d'activation des dopants du bridge, (d) épithaxie et recuit d'activation des dopants source et drain.

La figure 6.11.a montre le profil de dopants d'arsenic juste après la phase d'implantation de la figure 6.10.b, avec un $T_{\text{SOI}} = 10 \text{ nm}$, pour trois différentes valeurs d'énergie d'implantation. Lorsque l'énergie est supérieure à 15 keV, une bonne partie des dopants traversent la couche de SOI et finalement, il y a moins de dopant dans le bridge. Si l'énergie d'implantation est inférieure ou égale 15 keV, les dopants sont très localisés à la face arrière du SOI ; le bon compromis est pour une énergie d'implantation de 15 keV. Sur la figure 6.11.b, nous représentons le profil de dopants après la phase d'activation de la figure 6.10.c. Nous constatons qu'effectivement, l'énergie d'implantation garantissant un dopage du bridge d'environ 10^{18} cm^{-3} est donnée par l'énergie de 15 keV. Enfin, en considérant tous les bilans thermiques, nous obtenons le profil de dopage dans la profondeur du film de silicium sur la figure 6.11.c. Le profil (valeur) du dopage dans le bridge ne change pas quel qu'en soit l'énergie d'implantation. Elle reste égale à celle de la phase précédente. L'unique changement que nous pouvons constater est la diffusion des dopants du bridge dans le body. Le dopage dans le body atteint une valeur de 10^{15} cm^{-3} à 7 nm du bridge pour toutes les énergies d'implantation.

Sur la figure 6.12, nous présentons le profil de dopage d'arsenic après implantation, recuit d'activation, épithaxie du body et recuit final comme pour la figure 6.11.c mais ici avec un T_{SOI} de 15 nm. Nous pouvons noter que pour obtenir un dopage de 10^{18} cm^{-3} dans le bridge, il faut utiliser une énergie d'implantation de 20 keV. Comme pour le cas avec $T_{\text{SOI}} = 10 \text{ nm}$, le profil de dopage final le long du film de silicium met évidence une diffusion de dopants d'arsenic dans le body d'une longueur de 7 nm.

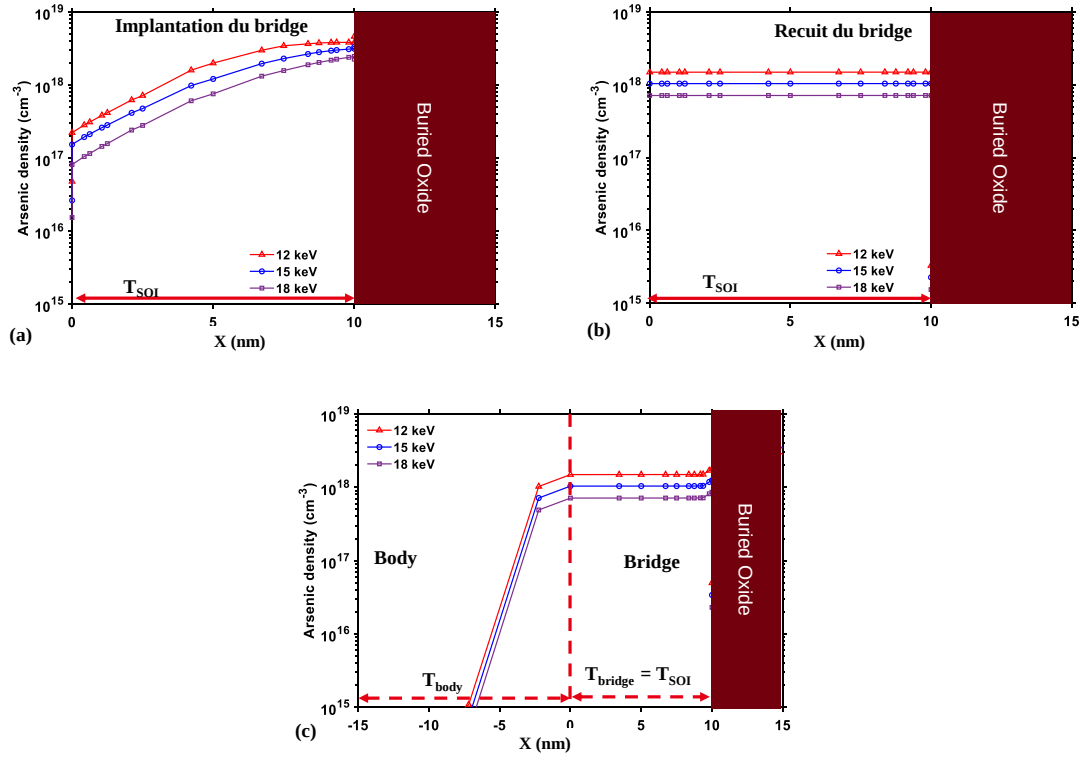


Figure 6.11 : Profil vertical de la concentration d'arsenic durant les différentes étapes pour différentes énergies d'implantation bridge : (a) Après implantation du bridge de la figure 6.10.b avec $T_{SOI} = 10$ nm, (b) après le recuit d'activation de la figure 6.10.c, et (c) après l'épitéxie du body (15 nm) et le dernier recuit du dopage in-situ des source et drain de la figure 6.10.d.

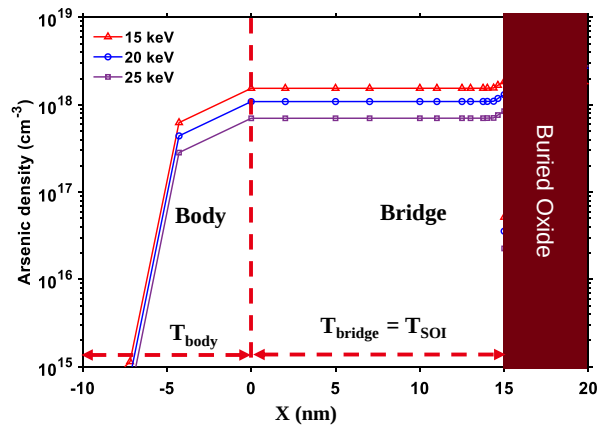


Figure 6.12 : Profil vertical de la concentration d'arsenic obtenu pour différentes énergies d'implantation bridge après l'épitéxie du body (10 nm) et le dernier recuit du dopage in-situ des source et drain partant d'un $T_{SOI} = 15$ nm.

Des descriptions faites sur les figures 6.11 et 6.12, il en ressort que les températures de recuits sont des éléments importants si nous voulons limiter autant que possible la diffusion des dopants du bridge dans le body. Mais, les températures utilisées pour ces simulations sont les minimales imposées par le procédé de fabrication. Ainsi, d'après les résultats obtenus sur les figures 6.11 et 6.12, pour réduire le risque d'avoir des structures avec un dopage du bridge largement supérieur ou inférieur à 10^{18} cm^{-3} , sur chaque plaque, nous avons inclus différentes énergies d'implantation en fonction de l'épaisseur T_{SOI} de départ (figure 6.13).

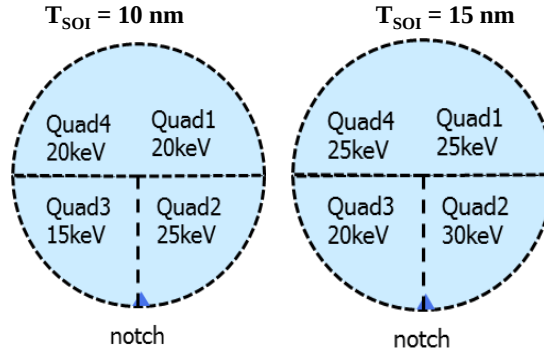


Figure 6.13 : Energie d'implantation du bridge en fonction de l'épaisseur de la couche active T_{SOI} et variation inter plaques.

VI.3. Résumé des choix des procédés pour la fabrication des cellules A2RAM 3D

Grâce aux simulations TCAD, nous avons défini toutes les étapes critiques du procédé nécessaires pour la fabrication de cellules 3D A2RAM. Ainsi, sur la base de nos études de simulations TCAD, les principaux paramètres critiques sont présentés dans le tableau 6.1. Notons que l'implantation des LDD est effectuée à faible énergie (2 keV) dans le but de respecter autant que possible une longueur des zones de recouvrement de 8 nm.

Étape du procédé	Options technologiques :	
Zone active / épaisseur du bridge	Silicium d'une épaisseur de 10 nm	Silicium d'une épaisseur de 15 nm
Implantation du bridge pour avoir un dopage de 10^{18} cm^{-3}	20 KeV +/- 25% et $5.10^{12} \text{ cm}^{-2}$	25 KeV +/- 25% et $5.10^{12} \text{ cm}^{-2}$
Épitaxie du body	Épaisseurs de 10 et 15 nm	Épaisseur de 10 nm
Matériaux du body	Silicium ou Silicium-Germanium (xGe = 30 %)	Silicium ou Silicium-Germanium (xGe = 30 %)
Implantation LDD à faible dose (2 KeV / 10^{13} cm^{-2})	Oui et non	Oui et non

Tableau 6.1: Variations des paramètres technologiques des cellules A2RAM 3D.

VI.4. Caractérisation électrique de la cellule A2RAM de type 3D

Pour la caractérisation électrique, 8 plaques sont disponibles, uniquement avec des body SiGe avec $x_{\text{Ge}} = 30\%$ car les plaques avec body silicium n'ont pas pu terminer le procédé de fabrication. L'apport de l'hétérostructure sur les performances de l'A2RAM ne pourra donc pas être mis en évidence. Les variantes technologiques présentes sur ces 8 plaques sont décrites dans le tableau 6.2. 12 puces ont été mesurées, 3 par quadrant qui ont chacun vu différentes énergies d'implantation du bridge. La figure 6.14 montre la disposition des puces mesurées, Q1, Q2, Q3 et Q4 se réfèrent à quad1, quad2, quad3 et quad4 comme indiqués sur la figure 6.13.

Plaques	Épaisseur du Bridge Si (nm)	Épaisseur du Body SiGe (nm)	Implantation LDD
18	10	15	Yes
19	10	15	No
20	10	15	No
21	10	15	Yes
22	15	10	No
23	15	10	Yes
24	10	10	No
25	10	10	Yes

Tableau 6.2 : Variations des paramètres technologiques en fonction des plaques.

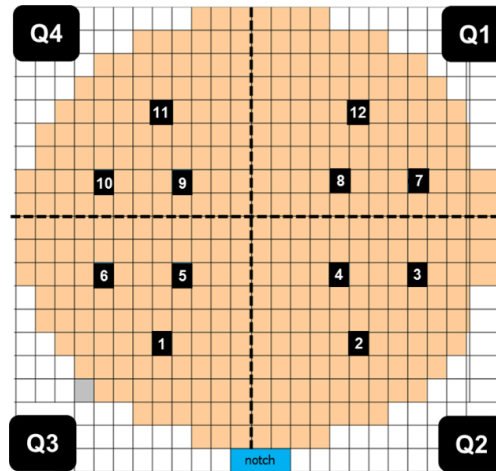


Figure 6.14 : Cartographie des différents dispositifs mesurés en fonction des énergies d'implantation du bridge comme montré sur la figure 6.13.

La figure 6.15 montre le dessin de masque simplifié représentatif des structures mesurées. Sur la figure 6.15.a, nous avons choisi les longueurs de la grille (L_g) qui varient entre 30 nm à 1 μm , et les largeurs (W) entre de 80 nm à 10 μm . De plus, dans le jeu des structures que nous avons à notre disposition, nous avons aussi sélectionné des structures multidoigts (x 50,

figure 6.15.b) avec des largeurs beaucoup plus étroites allant de 30 nm à 100 nm. Celles-ci nous permettront d'étudier les avantages apportés par la structure 3D en terme de contrôle électrostatique de la grille face avant.

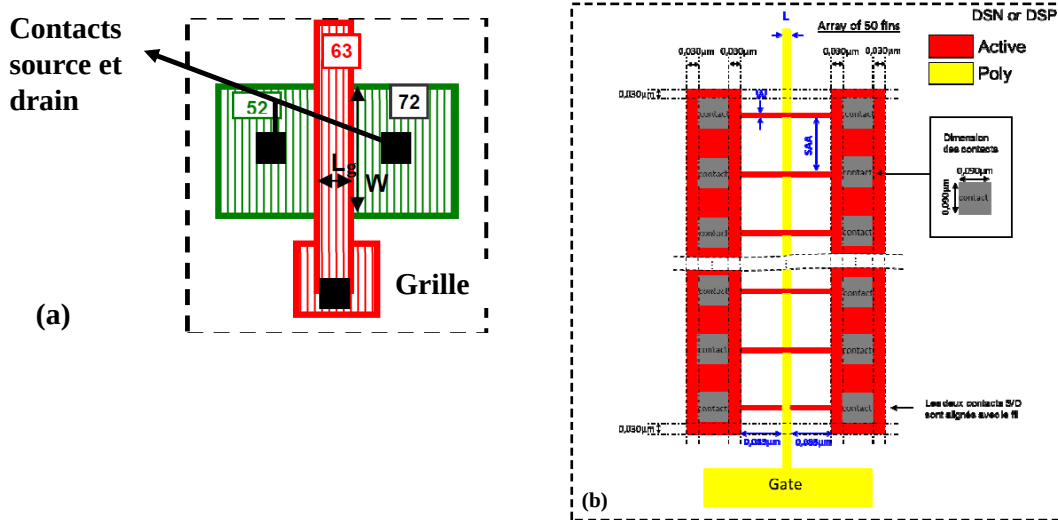


Figure 6.15 : Dessin de masque représentatif des dispositifs mesurés: a) grande largeur d'active pour fabriquer des dispositifs planaires, b) structures multi-doigts avec zone active étroite pour obtenir une structure 3D nanofil.

VI.4.1. Caractéristique statique de la cellule A2RAM 3D

Dans un premier temps, nous avons extrait la caractéristique statique (I_d - V_g) courant de drain (I_d) en fonction de la tension de la grille face avant (V_g) pour une tension de drain $V_d = 50$ mV ou forte tension $V_d = 900$ mV. Ici, nous présentons les caractéristiques I_d - V_g obtenues sur un dispositif avec une longueur de grille L_g de 80 nm et une largeur W de 10 μm .

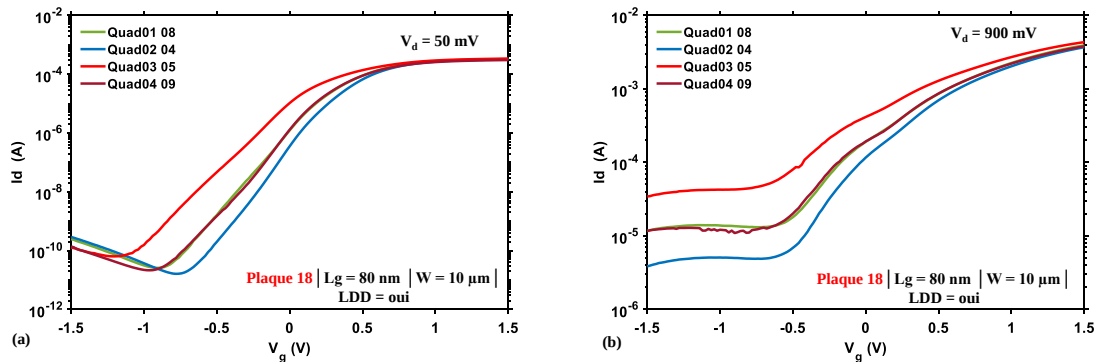


Figure 6.16 : I_d - V_g à (a) $V_d = 50$ mV et (b) $V_d = 900$ mV pour des dispositifs issus de la plaque 18 de longueur $L_g = 80$ nm et de largeur $W = 10$ μm .

Sur la figure 6.16, nous présentons les mesures effectuées sur 4 dispositifs (1 par quadrant) de la plaque 18 ($T_{\text{body}} = 15 \text{ nm}$ et $T_{\text{bridge}} = 10 \text{ nm}$). A $V_d = 50 \text{ mV}$, la conduction dans le bridge n'est pas clairement visible comme on peut le constater sur la figure 6.16.a ; ceci pourrait être due à la forte résistivité du bridge pour une épaisseur $T_{\text{bridge}} \approx 10 \text{ nm}$ et un dopage $N_{\text{bridge}} \approx 10^{18} \text{ cm}^{-3}$. Sur la figure 6.16.b, lorsque $V_d = 900 \text{ mV}$, la conduction dans le bridge est visible pour $-0.5 \text{ V} < V_g < 0.5 \text{ V}$, cette observation ne démontre pas un effet mémoire, mais est un indice de la présence d'une couche dopée de type n en bas du film de silicium.

Sur les figures 6.16.(a, et b), nous pouvons noter que les amplitudes de courant sont fonction de l'énergie d'implantation du bridge. À la section 2.4, nous avons observé qu'une forte énergie d'implantation implique un faible dopage dans le bridge ; vice versa, une faible énergie implique un fort dopage dans le bridge (Figure 6.11.c et 6.12). C'est pourquoi les amplitudes de courant élevées sont obtenues ici pour le quad03 qui correspond à la plus faible énergie d'implantation (15 KeV, figure 6.13). Les courbes issues des quadrants 01 et 04 se superposent parce qu'elles ont une même énergie d'implantation du bridge (figure 6.13).

Les observations faites sur la plaque 18 ont été retrouvées sur les autres plaques à notre disposition. Même si les caractéristiques statiques ne nous garantissent pas l'effet mémoire, il est important de noter que tous les dispositifs du tableau 6.2 semblent présenter une couche dopée de type n au fond du film de silicium. Pour confirmer la présence du bridge des mesures petit signal et le tracé de courbes capacité-tension est nécessaire.

VI.4.2. Caractéristique petit signal de la cellule A2RAM 3D

Pour confirmer la présence du bridge dopé de type n dans nos dispositifs, et pour essayer d'extraire le dopage du bridge comme nous l'avons fait dans le Chapitre 4, nous avons fait des mesures électriques petit signal et tracer la variation de la capacité de grille face avant C_{gg} en fonction de la tension de la grille face avant V_g ($C_{\text{gg}}-V_g$). Le protocole de la mesure est identique à celui utilisé dans le Chapitre 4 ; le dispositif mesuré est long et large ($L_g = 1 \mu\text{m}$ et $W = 10 \mu\text{m}$).

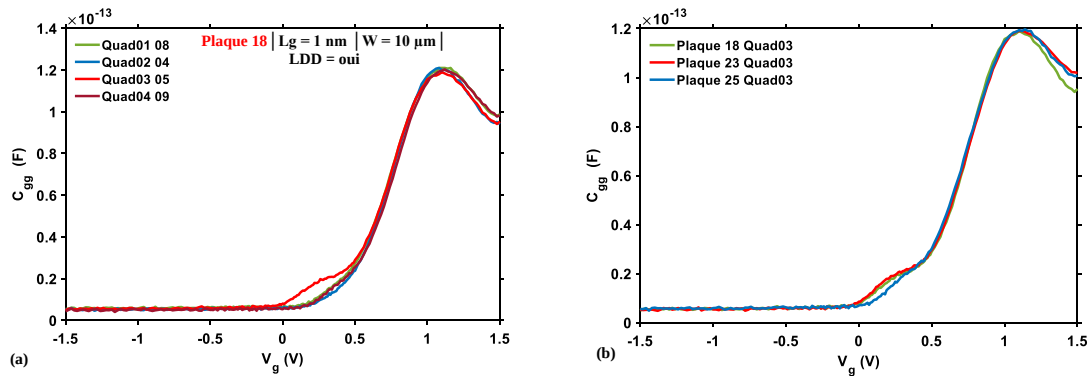


Figure 6.17 : Caractéristiques de la capacité de la grille face avant C_{gg} en fonction de la tension de la grille face avant V_g pour (a) 4 dispositifs de la plaque 18, (b) trois différentes plaques. Les dispositifs mesurés ont une longueur $L_g = 1 \mu\text{m}$ et une largeur $W = 10 \mu\text{m}$

La figure 6.17.a présente la caractéristique C_{gg} - V_g pour 4 dispositifs (un par quadrant, donc pour des implantations de bridge différentes) sur la plaque 18 ($T_{body} = 10$ nm et $T_{bridge} = 15$ nm). Lorsque $0 \text{ V} \leq V_g < 0.5 \text{ V}$, nous observons une bosse sur les caractéristiques C_{gg} - V_g qui indique la présence du bridge comme expliqué dans le Chapitre 4. Cette bosse est plus élevée pour le quadrant 03 correspondant à l'énergie d'implantation la plus basse. Ce résultat est en accord avec les observations faites sur les caractéristiques statiques de la figure 6.16. Cependant, lorsque la tension de grille augmente, en régime de forte inversion dans le canal face avant ($V_g \geq 1 \text{ V}$) où la capacité doit saturer à une valeur constante, nous observons une chute de la capacité qui sera investiguée dans la suite de ce paragraphe.

Pour vérifier la présence du bridge sur toutes les combinaisons d'épaisseur body/bridge, les caractéristiques C_{gg} - V_g des plaques 18 ($T_{body} = 15$ nm et $T_{bridge} = 10$ nm), 23 ($T_{body} = 10$ nm et $T_{bridge} = 15$ nm) et 25 ($T_{body} = 10$ nm et $T_{bridge} = 10$ nm) pour un dispositif du quadrant 03 (car il permet de mieux observer l'effet du bridge) sont reportées sur la figure 6.17.b :

- pour $0 \text{ V} \leq V_g < 0.5 \text{ V}$, le dispositif ayant le bridge le plus épais est au-dessus des autres courbes, ceci est cohérent avec nos études du Chapitre 4. Car, un bridge épais implique plus de charge dans le bridge, par conséquent le champ électrique vertical induit par la grille face avant le déplete très peu pour une même valeur de polarisation comparé aux autres cas. Cependant, en régime de forte inversion comme pour la figure 6.17.a, nous avons une chute de la capacité. La capacité d'oxyde de la grille face avant est donnée par

$$C_{ox} = \frac{\epsilon_{ox}}{EOT} WL_g \quad (6.1)$$

Pour un dispositif avec une longueur de grille $L_g = 1 \text{ } \mu\text{m}$, une largeur $W = 10 \text{ } \mu\text{m}$, et une épaisseur équivalente d'oxyde de grille face avant de 1.2 nm, nous devrions avoir $C_{gg} = C_{ox} \approx 2.88 \times 10^{-13} \text{ F}$ pour $V_g > 0.5 \text{ V}$. La valeur de la capacité d'oxyde de grille face avant obtenue avec l'expression 6.1 reste deux fois supérieure aux valeurs C_{gg} pour $V_g \geq 1 \text{ V}$. Cette chute de la capacité ne peut être due à la fuite de la grille face avant. En effet, la figure 6.18.a montre le courant (valeurs logarithmiques) de la grille face avant extrait durant la caractérisation petit signal. L'amplitude du courant de grille reste à un niveau faible, et standard pour cet empilement de grille et est la même quelle que soit la puce mesurée. De plus, sa valeur est typique pour un empilement de grille dont l'épaisseur équivalente de l'oxyde vaut 1.2 nm [Kang 00].

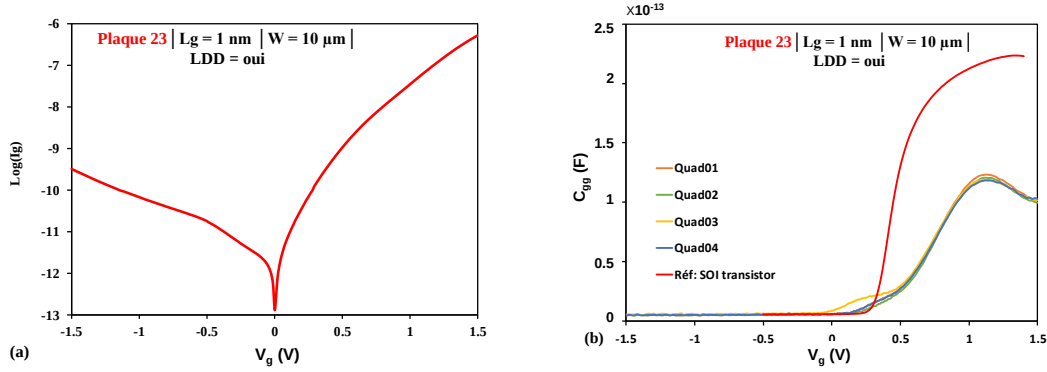


Figure 6.18 : (a) Variation du courant de la grille face avant extrait des échantillons de la plaque 23 quadrant 03 lors de la caractéristique petit signal de la figure 6.17. (b) Comparaisons des caractéristiques C_{gg} - V_g des dispositifs de la plaque 23 avec une cellule nanofil (Réf : SOI transistor) ayant le même empilement de la grille face avant tel que décrit dans la section 1.

Dans l'optique pourquoi en régime d'inversion nous avons une chute de la capacité de la grille face avant. Nous avons comparé les caractéristiques obtenues sur ces plaques, avec celle mesurée sur un dispositif de référence ayant subi le même procédé de fabrication à l'exception des étapes spécifiques de l'A2RAM. Sur la figure 6.18.b, nous présentons la comparaison entre les dispositifs de la plaque 23 et le dispositif de référence. Nous pouvons noter que :

- Pour des polarisations de grille face avant telles que $0 \text{ V} < V_g < 0.5 \text{ V}$, la capacité de la grille face avant C_{gg} de l'A2RAM est supérieure à celle de la cellule de référence. Ceci est dû à la présence du bridge dans une structure A2RAM.
- Cependant, lorsque la tension de la grille face avant augmente, la capacité de la grille face avant de l'A2RAM n'atteint pas sa valeur maximale comme la structure de référence qui ont pourtant le même empilement de grille.

Ce comportement peut être symptomatique d'un polysilicium non dopé sur le métal de grille. Pour le confirmer, sur la figure 6.19, nous comparons les caractéristiques C_{gg} - V_g obtenues par des simulations TCAD de différentes structures avec le même empilement de la grille face avant, mais en faisant varier le dopage du poly silicium. Le résultat obtenu confirme effectivement qu'un faible dopage du poly silicium ($< 10^{15} \text{ cm}^{-3}$) implique une chute de la capacité de la grille face avant.

Ce résultat implique que le contrôle électrostatique de la grille face avant sera dégradé, car la tension appliquée sur la grille sera en partie répartie sur le poly silicium non dopé, sur l'oxyde de la grille face avant, et le reste sur le film de silicium. Par conséquent les performances mémoires de la cellule 3D A2RAM risquent aussi d'être dégradées.

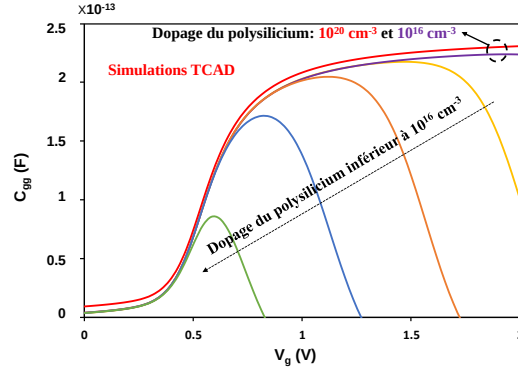


Figure 6.19 : Caractéristiques C_{gg} - V_g extraites de simulations TCAD de dispositifs A2RAM avec des variations du dopage du polysilicium dans l'empilement oxide de grille-Metal-polysilicium.

VI.4.3. Essai d'extraction du dopage du bridge

Bien que l'étude de la caractéristique de la capacité de la grille face avant en fonction des variations de la tension de grille face avant montre que le polysilicium n'est pas dopé, nous avons essayé d'extraire le dopage du bridge. L'idée étant d'utiliser la méthodologie développée dans le Chapitre 4 ; dans un premier temps, nous avons évalué la dérivée première qui nous permet de quantifier la tension seuil du bridge.

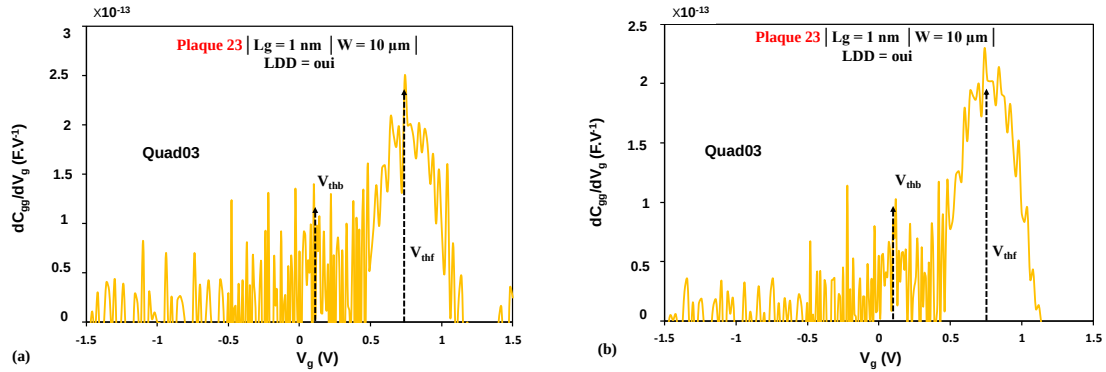


Figure 6.20 : Dérivée première de la caractéristique C_{gg} - V_g de la figure 6.17.b pour (a) la plaque 23, (b) moyennée sur trois dispositifs de la plaque 23.

Sur la figure 6.20.a, nous avons la dérivée première extraite d'un dispositif de la plaque 23 du quadrant 3 ($T_{body} = 10$ nm et $T_{bridge} = 15$ nm avec un énergie d'implantation du bridge de 20 keV). Nous constatons que la caractéristique est très bruitée, comme la caractéristique C_{gg} - V_g de la figure 5.17.b. Notons que dans le Chapitre 4, nous avons déjà utilisé le même protocole de mesure pour extraire la caractéristique petit signal sur les échantillons fabriqués au Leti en 2011. Il est difficile de donner une explication claire sur la provenance de ce bruit sur nos mesures, mais il est probable que le polysilicium de la grille non dopée en soit la cause. Pour réduire l'effet du bruit, nous avons refait la mesure sur trois dispositifs du même quadrant,

et la moyenne des résultats obtenus est présenté sur la figure 6.20.b. Dans ce cas, la caractéristique est 'moins' bruitée, et nous pouvons distinguer deux pics, le premier correspondant à la tension seuil du bridge V_{thb} se trouve à 100 mV. Le second pic se trouve à 700 mV et correspond à la tension seuil du canal face avant.

Nous avons utilisé la même technique pour extraire la tension seuil du bridge sur l'ensemble des quadrants de la plaque 23. Sur les figures 6.21, nous avons la dérivée première moyennée sur des dispositifs des quadrants 01 (énergie d'implantation du bridge de 25 KeV), 02 (énergie d'implantation du bridge de 30 KeV) et 04 (énergie d'implantation du bridge de 25 KeV). Le pic correspondant à la tension de seuil du bridge V_{thb} se rapproche de plus en plus de celui de la tension du canal face avant, car ces dispositifs ont un dopage de bridge faible comparé aux dispositifs du quadrant 3. Même en moyennant les mesures des caractéristiques petits signaux de nos échantillons, la dérivée première reste encore très bruitée. Ainsi, lorsque nous utilisons l'équation 4.3 pour extraire le dopage du bridge, le résultat obtenu n'est pas exploitable.

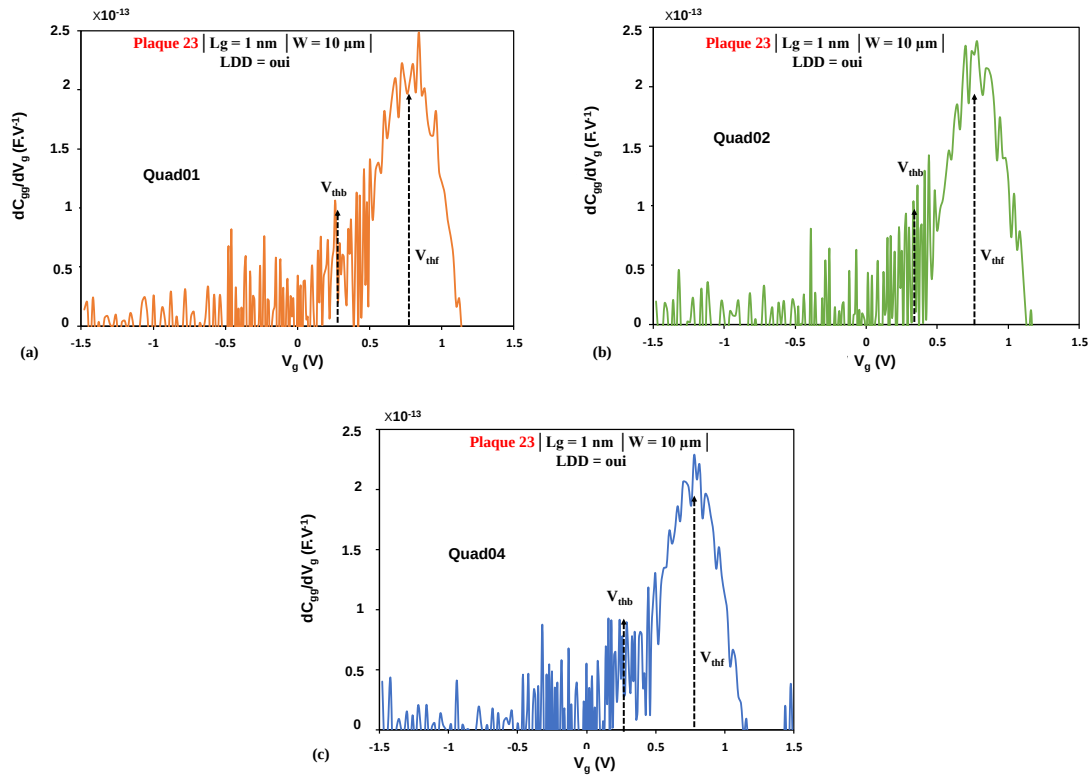


Figure 6.21 : Dérivées premières des caractéristiques C_{gg} - V_g de 3 dispositifs de la plaque 23 ($T_{body} = 10$ nm, $T_{bridge} = 15$ nm) avec une énergie d'implantation du bridge qui vaut (a) 25 keV, (b) 30 keV et (c) 25 keV.

L'étude de la caractéristique petits signaux nous a permis de mettre en évidence la présence d'un polysilicium non dopé. Mais, malgré le fort bruit présent sur ces caractéristiques, il est tout de même possible de confirmer la présence d'un bridge. Le quadrant qui a un bridge

moins résistif (fort dopage de bridge) reste celui avec une faible énergie d'implantation du bridge (20 keV).

VI.4.4. Analyse des performances mémoires

Dans cette partie, nous allons mettre en évidence la présence d'un effet mémoire sur les dispositifs A2RAM 3D dont les caractéristiques statiques et petits signaux ont été étudiées précédemment. Pour cela, nous avons utilisé le motif de tension de la figure 6.22, le choix des valeurs de tensions a été fait de façon à garantir l'observation d'un effet mémoire.

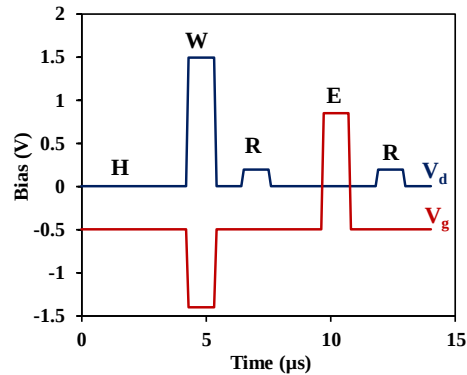


Figure 6.22 : Chronogramme de tensions appliquées sur le drain V_d (en bleu) et la grille face avant V_g (en rouge) pour la séquence programmation : lecture-effacement-lecture.

Nous avons appliqué ce motif de tension sur un dispositif ayant $T_{\text{body}} = 10 \text{ nm}$, $T_{\text{bridge}} = 15 \text{ nm}$ et une énergie d'implantation du bridge de 20 KeV. Les valeurs du courant de drain des différentes opérations mémoires sont présentées sur la figure 6.23. Nous pouvons noter que le courant en écriture se comporte comme nous l'avons souligné dans le Chapitre 2 lorsqu'il y a un stockage de charge dans le body. Ainsi, le courant I_1 lu est dû effectivement à un courant circulant dans le bridge. Mais, ce courant est inférieur au courant lu de l'état '0' I_0 , comme montré sur la figure 6.23. La phase d'effacement avec un polysilicium non dopé doit alors être investiguée.

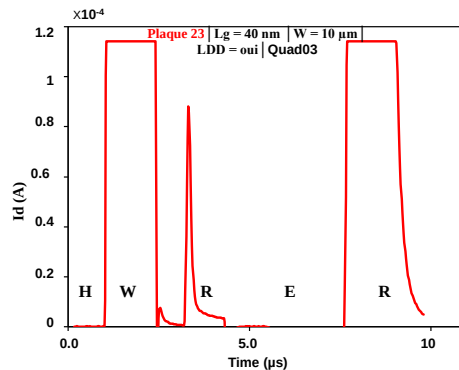


Figure 6.23 : Variations expérimentales du courant de drain durant les opérations mémoires (figure 6.22) d'un dispositif 3D A2RAM.

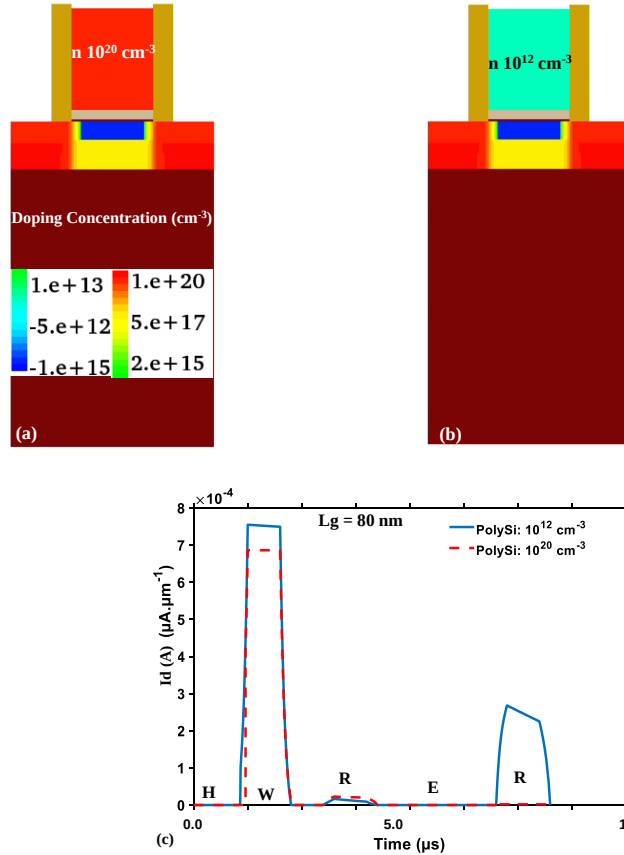


Figure 6.24 : Simulation TCAD d'opérations mémoire de la figure 6.22 sur une structure A2RAM avec (a) le polysilicium dopé à 10^{20} cm⁻³ et (b) le polysilicium dopé à 10^{12} cm⁻³. (c) Comparaison des chronogrammes de courant de drain des structure (a) et (b).

Le chronogramme de la figure 6.22 est simulé par TCAD sur une structure A2RAM représentative des dispositifs fabriqués ($T_{\text{body}} = 10$ nm, $T_{\text{bridge}} = 15$ nm, $N_{\text{bridge}} = 10^{18}$ cm⁻³ et $L_g = 80$ nm), mais avec le polysilicium dopé à 10^{20} cm⁻³ (figure 6.24.a), ou à 10^{12} cm⁻³ (figure 6.24.b). La figure 6.24.c montre le chronogramme de courants de drain issus des deux cellules. Les courants en écriture et lecture de l'état '1' suivent quasiment les mêmes variations dans les deux cas. Cependant, lorsque nous lisons le courant de l'état '0', le courant I_0 reste très faible pour la structure avec un polysilicium dopé. Mais, le courant lu de l'état '0' de la cellule avec polysilicium faiblement dopé est largement supérieur au courant I_1 , comme ce qui est vu expérimentalement.

En effet, durant l'effacement de l'état '0', l'impulsion positive appliquée sur la grille face avant met le canal face avant en régime d'inversion. Les électrons, une fois dans le canal vont se recombiner en partie avec les trous stockés, et lorsqu'on repasse à la tension de maintien sur la grille face avant ($V_g \ll 0$ V), le body reste en régime de désertion. Ceci n'est pas le cas lorsque le polysilicium est non dopé. Dans ce cas, les électrons restent dans le body bien que la tension de la grille face avant repasse à sa valeur de maintien $V_{g\text{Hold}} = -1.2$ V négative. L'évacuation des électrons du canal n'étant pas instantanée, lorsque nous appliquons un

signal positif sur le drain nous lisons un courant de canal en plus de la conduction dans le bridge.

De l'étude faite sur les figures 6.23 et 6.24, il en ressort qu'à cause du polysilicium non dopé il n'est pas possible d'effacer l'information. Notons que nous avons aussi essayé d'effacer l'information en utilisant la technique de polarisation des diodes en direct, mais l'amplitude du courant I_0 reste toujours égale ou supérieur à I_1 . Ainsi, pour la suite de ce travail nous allons nous focaliser sur la programmation et la lecture de l'état '1'.

VI.4.4.1. Caractérisation de l'écriture par GIDL sur nos échantillons

Pour cette partie, nous allons utiliser le chronogramme de tension de la figure 6.25.a où nous avons simplement une phase de programmation de l'état '1'. En effet, pour pouvoir observer le courant d'écriture, nous devons augmenter la compliance du courant à 10 mA ce qui correspond à un bruit de courant de l'instrument de mesure de 1.9 mA rendant impossible la mesure du courant durant la phase de lecture (l'amplitude maximale est d'environ 100 μ A) avec une bonne résolution. Par ailleurs, le chronogramme de la figure 6.25.a a une durée de 5 μ s pour maximiser le nombre de points d'acquisition et donc augmenter la résolution du courant de programmation en fonction du temps.

Sur la figure 6.25.b, nous avons le chronogramme du courant de drain lorsqu'on applique le motif de tensions de la figure 6.25.a sur différentes cellules de la plaque 23 ayant toutes $L_g = 40$ nm et $W = 10$ μ m. Notons tout d'abord que le courant en écriture se comporte comme nous l'avons déjà décrit dans notre Chapitre 2 : il augmente progressivement jusqu'à saturation à une valeur constante, démontrant ainsi que le stockage de la charge a atteint sa limite maximale.

Plus en détail, la figure 6.25.b, montre les variations du courant en écriture en fonction de l'énergie d'implantation du bridge. Nous constatons une fois de plus que l'amplitude du courant la plus élevée est donnée pour le quadrant dont l'énergie d'implantation est la plus faible (Quad03 qui correspond à une énergie de 20 KeV). Car, de nos analyses TCAD (section 2.4) ce quadrant a un dopage de bridge le plus élevé, donc son bridge est moins résistif. Lorsque le courant sature en écriture dans une A2RAM, sa valeur dépend majoritairement de la conduction à travers le bridge. Ainsi, le courant est élevé s'il y a plus de charge dans le bridge donc un fort dopage qui est donnée par une faible énergie d'implantation. Ce résultat est cohérent avec les observations déjà faites sur les caractéristiques statiques et petits signaux précédentes. Le même résultat est observé sur les autres plaques.

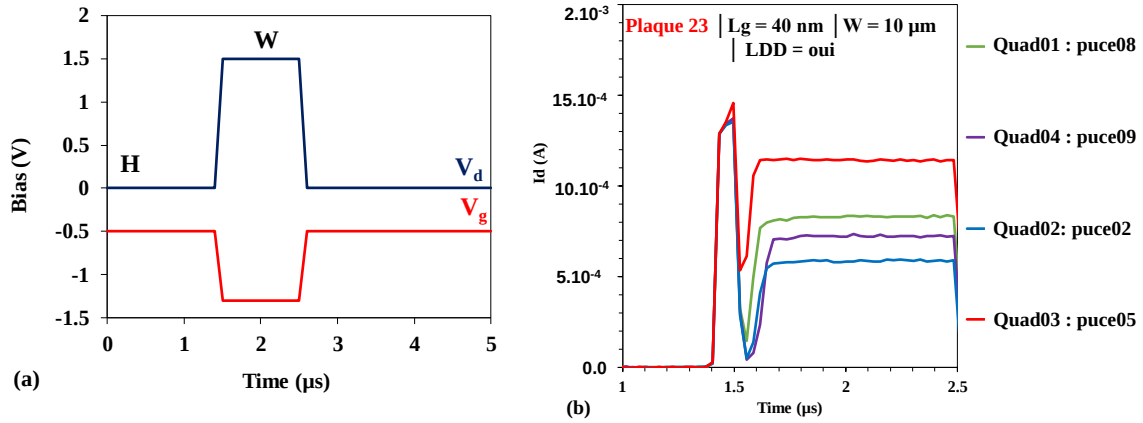


Figure 6.25 : (a) Chronogramme de tensions appliqué sur le drain V_d (en bleu) et la grille face avant V_g (en rouge) pour la programmation de l'état '1'. (b) Variation du courant de drain pour des dispositifs de la plaque 23.

Sur la figure 6.26, nous comparons les variations du courant d'écriture en fonction de la longueur de la grille en considérant le quadrant avec la plus faible énergie d'implantation du bridge (Quad03). Nous pouvons remarquer que :

- le plateau de saturation est fonction de la longueur de la grille. Ceci est dû au fait que le courant de drain dans la structure est essentiellement dû au courant dans le bridge. Si on augmente la longueur du bridge, sa résistance augmente d'où le comportement de la figure 6.26.
- le plateau de saturation est atteint beaucoup plus lentement pour des structures de plus en plus longues. La capacité du réservoir de stockage de trous (le body) augmente avec la longueur de la grille, donc nécessite plus de temps pour stocker de la charge.

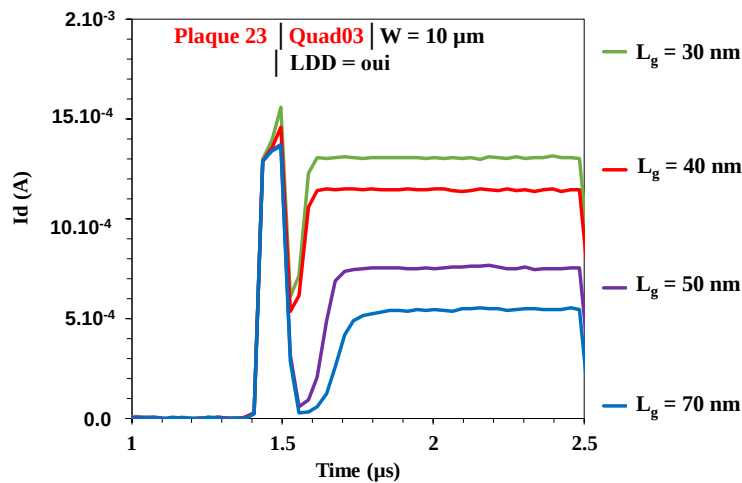


Figure 6.26 : Variation du courant drain pendant l'opération d'écriture de l'état '1' (figure 6.25) pour des dispositifs de la plaque 23 du quadrant numéro 3.

Dans la suite, nous allons nous concentrer sur l'amplitude du courant lu de l'état '1'. Nous allons tour à tour étudier l'impact de l'énergie d'implantation sur la valeur de I_1 , ensuite les variations de I_1 lorsqu'on varie les paramètres technologiques de la mémoire, à savoir :

- les épaisseurs du body et du bridge (plaques 18, 23 et 25),
- les implantations LDD (plaques 22 et 23).

Les variations de I_1 seront également évaluées lorsqu'on applique une polarisation en face arrière V_b .

Puisque l'effacement de l'information n'est pas possible avec un polysilicium non dopé, nous avons changé le chronogramme de tension utilisé pour caractériser la cellule A2RAM. La figure 5.27 montre le nouveau motif de tension que nous allons utiliser. Il se compose d'une longue phase de lecture (R) afin d'assurer que le film de silicium (body + bridge) soit totalement déserté. L'amplitude du courant durant cette phase est faible et se rapproche de la valeur du courant lu de l'état '0', c'est pourquoi nous la considérons comme une référence pour la valeur de I_0 . Ensuite nous programmons l'état '1' (W), et nous finissons par deux lectures consécutives (R). Pour la mesure du courant, nous avons considéré deux compliances, la première est de 10 μA et le bruit de l'instrument de mesure associé est d'environ 2.5 nA : cette compliance sera utilisée pour évaluer la première lecture du courant de drain avant la programmation. La seconde compliance est de 100 μA et le bruit de l'instrument de mesure associé est de 47 nA : elle est utilisée pour évaluer le courant de lecture de l'état '1', après programmation (figure 6.27).

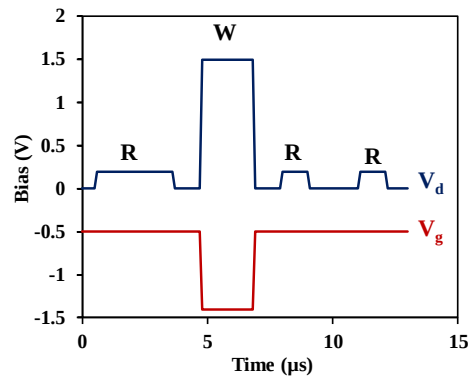


Figure 6.27 : Chronogramme de tensions appliqué sur le drain V_d (en bleu) et la grille face avant V_g (en rouge) durant différents types d'opérations mémoires.

VI.4.4.2. Lecture de l'état '1' : impact de l'énergie d'implantation du bridge

La cellule que nous avons choisie pour cette caractérisation est issue de la plaque 23, avec une longueur de grille $L_g = 40$ nm et une largeur $W = 10$ μm . Sur la figure 6.28.a, nous reportons les variations du courant de drain, pour 4 cellules issues des 4 quadrants, durant la première lecture du motif de la figure 6.27. Sur la figure 6.28.b, nous avons les variations

du courant de drain de ces mêmes cellules durant les deux cycles consécutifs de lecture. Nous notons que :

- Le courant lu avant la programmation de l'état '1' reste constant à une valeur inférieure ou égale à $1.2 \mu\text{A}$. De plus, la valeur du courant est fonction de l'énergie d'implantation comme nous l'avons souligné sur la figure 6.25, et le maximum est donné par la plus petite énergie d'implantation. Notons que le courant issu du quadrant 2 est bruité parce qu'il se rapproche de la gaussienne du bruit de l'instrument de mesure dont le pic est à 2.5 nA .
- Les valeurs des courants lus de l'état '1' I1 (première lecture) sont comprises entre $0.2 \mu\text{A}$ et $5 \mu\text{A}$. Quel que soit le quadrant, les courants I1 lus sont 5 à 10 fois supérieurs à la première lecture avant la phase de programmation ; nous pouvons donc conclure que nous avons effectivement programmé l'état '1'.
- La seconde lecture de l'état '1' montre que le courant décroît avec le temps et converge vers la valeur d'équilibre thermodynamique de la première lecture. Par conséquent la rétention est très médiocre, ceci pourrait être encore un effet du polysilicium non dopé.

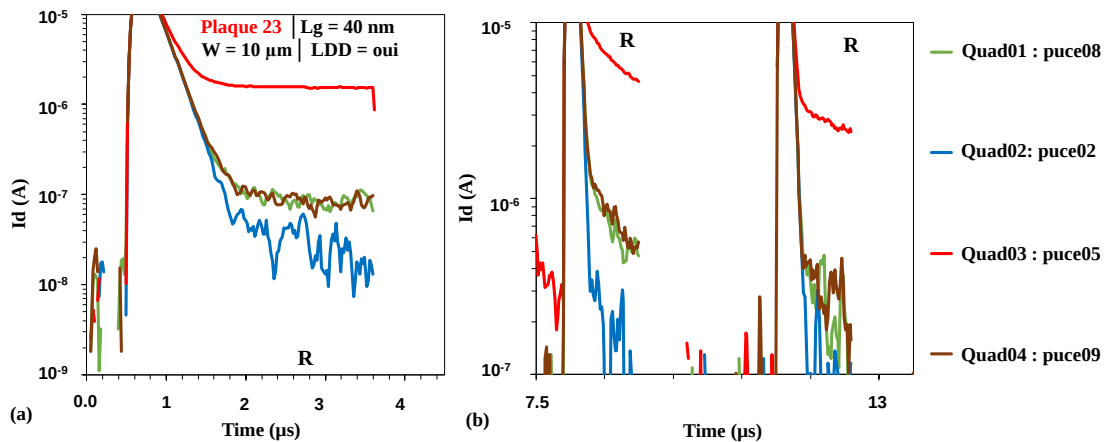


Figure 6.28 : Variations du courant de drain en fonction de l'énergie d'implantation du bridge durant les opérations mémoires de la figure 6.27, pour (a) la première phase de lecture avant programmation, (b) les deux autres phases de lectures après programmation.

Ainsi, grâce à l'étude faite sur la figure 6.28, nous pouvons dire que même s'il n'est pas possible de mettre évidence l'état '0' sur nos échantillons, nous pouvons programmer l'état '1'. Le meilleur compromis entre la valeur du courant à l'équilibre thermodynamique avant écriture et la valeur de I1 est garanti pour un dispositif issu du quadrant 3. Cependant, l'amplitude du courant de l'état '1' lu et le temps de rétention de l'information sont bien plus faibles qu'attendus.

VI.4.4.3. Lecture de l'état '1' : impact de la longueur de la grille

Nous avons aussi étudié l'impact de la longueur de la grille sur les amplitudes des courants lus avant et après la phase de programmation (figure 6.27). Sur la figure 6.29 nous reportons

les variations de courant de drain pour les dispositifs issus du quadrant 3 avec différentes longueurs de grille. Sur la figure 6.29.a, nous notons que le courant à l'équilibre thermodynamique augmente lorsque la longueur de la grille diminue. Ceci est dû fait que les effets canaux courts augmentent lorsqu'on réduit la longueur de la grille et de plus, la résistance du canal diminue avec la longueur de la grille. Sur la figure 6.29.b, nous avons les variations des courants lus de l'état '1'. Les tendances sont les mêmes que sur la figure 6.28.b. Lorsque $L_g < 40$ nm à cause de la très forte présence des effets de canaux courts, l'effet mémoire n'est pas très visible car le courant I_1 est quasiment égal à la valeur du courant avant programmation. Pour $L_g > 40$ nm, l'effet mémoire est visible mais les amplitudes de courant I_1 sont très faibles. Par ailleurs, même pour ces longueurs le temps de rétention n'est pas amélioré malgré leur forte immunité aux effets de canaux courts. Ces observations confirment une fois de plus que l'élément responsable de la dégradation de la rétention de l'état '1' est la grille non dopée.

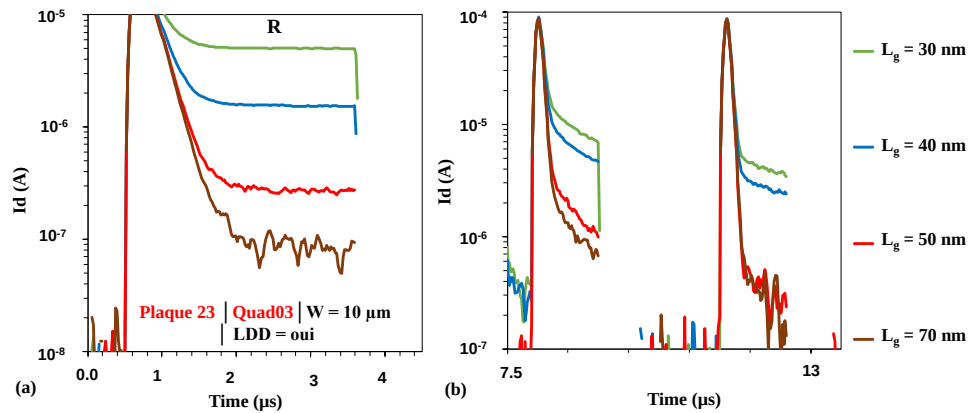


Figure 6.29 : Variations du courant de drain en fonction de la longueur de la grille durant les opérations mémoires de la figure 6.27, pour (a) la première phase de lecture avant programmation, (b) les deux autres phases de lectures après programmation.

VI.4.4.4. Lecture de l'état '1' : impact de l'épaisseur du bridge et/ou du body

Ici, nous allons étudier les performances mémoires des plaques 18, 23 et 25 ; le dispositif choisi sur ces plaques a une longueur de grille $L_g = 40$ nm ($W = 10$ μm). Comme indiqué dans le tableau 6.2, ces plaques présentent des différences au niveau de l'épaisseur du bridge (18/25 et 20) et de l'épaisseur totale du film de silicium (18/23 et 25). Notons que ces trois plaques ont une implantation LDD ; une fois la comparaison faite, nous évaluerons aussi l'impact d'une structure sans implantation LDD sur les performances obtenues. Pour ces études, le chronogramme de tensions de la figure 6.27 est toujours utilisé.

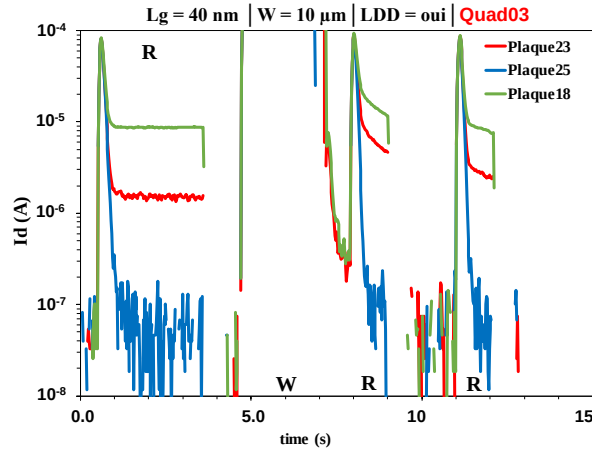


Figure 6.30 : Les variations du courant de drain durant les opérations mémoires de la figure 6.27 sur des dispositifs issus de plaque 18 ($T_{\text{body}} = 15 \text{ nm}$ et $T_{\text{bridge}} = 10 \text{ nm}$), plaque 23 ($T_{\text{body}} = 10 \text{ nm}$ et $T_{\text{bridge}} = 15 \text{ nm}$) et plaque 25 ($T_{\text{body}} = 10 \text{ nm}$ et $T_{\text{bridge}} = 10 \text{ nm}$) tous ayant une longueur $L_g = 40 \text{ nm}$ et une largeur $W = 10 \mu\text{m}$.

Sur la figure 6.30 sont montrées les variations du courant de drain des plaques 23, 25 et 18. Nous remarquons que le courant de drain reste faible durant toutes les opérations de la figure 6.28 pour le dispositif issu de la plaque 25 ($T_{\text{Si}} = 20 \text{ nm}$) : le film de silicium est très mince et le champ électrique vertical issu de la grille face avant maintient le bridge en 'désertion'. Pour des épaisseurs de film de silicium plus épaisses (plaques 18 et 25), le meilleur compromis entre le courant lu à l'équilibre thermodynamique et le courant lu de l'état '1' est garanti par la plaque 23.

VI.4.4.5. Lecture de l'état '1' : impact de l'implantation LDD

Nous allons à présent étudier l'impact des implantations LDD sur les performances mémoires. Considérant la plaque 23, nous comparons ses performances mémoires avec celles obtenues sur la plaque 22 lorsqu'on applique le chronogramme de tension de la figure 6.27. La figure 6.31 montre les variations de courant de drain en fonction du temps de deux dispositifs, l'un issu de la plaque 23 et l'autre de la plaque 22.

En présence d'une implantation LDD, le courant à l'équilibre thermodynamique est élevé. Ceci s'explique par le fait que l'implantation LDD réduit la longueur électrique du dispositif entraînant l'augmentation des courants de fuite par les diodes parasites source/body et drain/body comme nous l'avons vu de nos études TCAD de la section 2.2.

D'un autre côté, il est bénéfique d'avoir une implantation LDD car elle favorise le courant GIDL. Ainsi, la programmation par effet tunnel bande à bande assure un stockage efficace, d'où le fort courant lu de l'état '1' I1. De plus, vu que la longueur électrique du dispositif est réduite en présence d'implantation LDD, le courant lu I1 via la résistance du bridge augmente aussi.

Cependant pour nos cas de la figure 6.31, le courant I1 décroît très rapidement dans les deux cas, ceci veut dire que le contrôle électrostatique de la grille face avant est dégradé à cause de la présence du polysilicium non dopé.

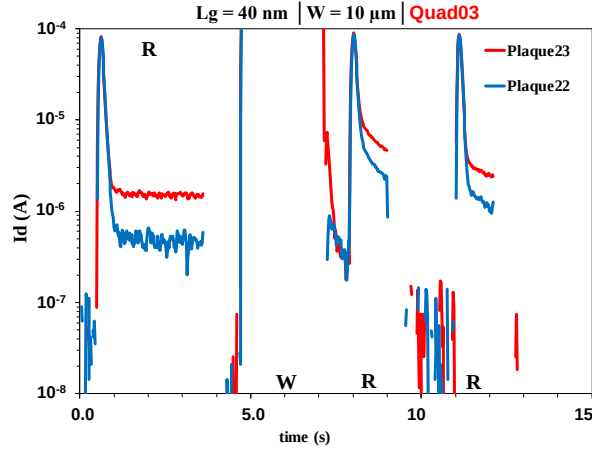


Figure 6.31 : Les variations du courant de drain durant les opérations mémoires de la figure 6.27 sur des dispositifs issus des plaques 22 (sans implantation LDD) et 23 (avec implantation LDD) ayant $T_{\text{body}} = 10 \text{ nm}$, $T_{\text{bridge}} = 15 \text{ nm}$, une longueur $L_g = 40 \text{ nm}$ et une largeur $W = 10 \mu\text{m}$.

De l'étude faite sur la figure 6.31, comme nous l'avons déjà mentionné de nos études TCAD, l'implantation LDD a un impact positif sur les performances mémoire, bien que le dispositif sans implantation LDD soit plus immunisé des courants de fuites par GIDL.

VI.4.4.6. Lecture de l'état '1' : impact de la polarisation face arrière

Pour finir, nous avons étudié l'impact de la polarisation de la grille face arrière sur les performances mémoire, comme déjà mis en évidence dans la section 2.3 du Chapitre 04. De même que pour les échantillons précédemment fabriqués, nous n'avons pas un contact direct sur la grille face arrière, par conséquent une tension continue est appliquée directement sur le chuck.

Nous reportons sur la figure 6.32.a le courant de drain durant la première phase de lecture du chronogramme de la figure 6.27 pour différentes polarisations sur la grille face arrière. La figure 6.32.b illustre les courants de drain en fonction du temps durant les deux phases de lecture de l'état '1' pour ces mêmes polarisations de la grille face arrière. Le dispositif qui fait l'objet de cette étude reste toujours celui de la plaque 23 du quadrant 03 avec une longueur de grille $L_g = 40 \text{ nm}$ et une largeur $W = 10 \mu\text{m}$.

Nous remarquons pour les deux figures que le courant de drain augmente lorsque la tension positive appliquée sur la grille face arrière augmente. De plus, les valeurs du courant lu de l'état '1' deviennent de plus en plus stables. Cependant, les amplitudes du courant de drain des figures 6.32.a et 6.32.b ne garantissent toujours pas une fenêtre de programmation ('1/10' au moins supérieur à 20) suffisante pour assurer une bonne fonctionnalité mémoire.

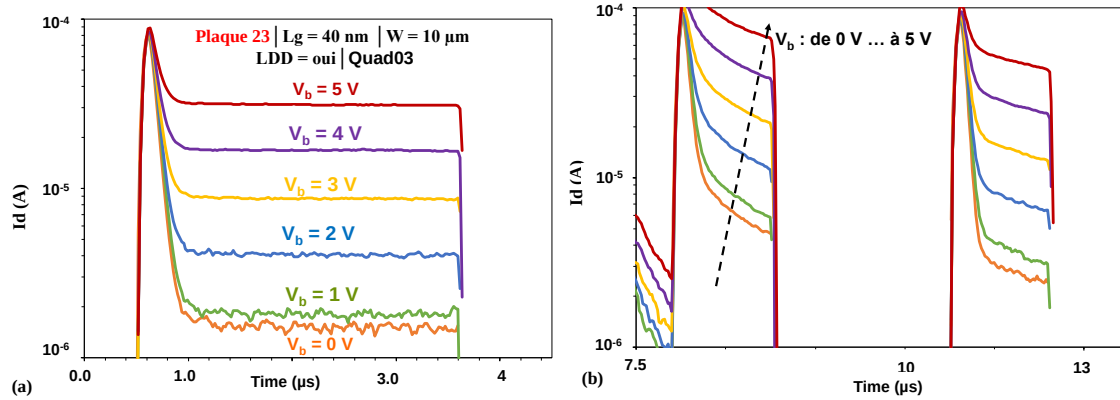


Figure 6.32 : Variations du courant de drain en fonction de la tension appliquée sur la grille face arrière durant les opérations mémoires de la figure 6.27, pour (a) la première phase de lecture avant programmation, (b) les deux autres phases de lectures après programmation.

Nous avons commencé cette étude de caractérisation électrique en mesurant les structures 3D A2RAM à canal large pour maximiser la valeur du courant, et vérifier la présence d'un effet mémoire. Par la suite, nous avons mesuré les structures multidoigts pour démontrer le gain apporté par une structures A2RAM 3D. dire également que tu as fait des CVs mais qu'on ne voyait rien Mais, lorsque nous avons utilisé nos motifs de tensions (figures 6.25.a et 6.27), nous n'avons pas observé un effet mémoire. En effet, les structures multidoigts avaient une largeur W totale entre 2 μm et 5 μm, donc très étroites comparées aux structures planaire. De plus, ayant un polysilicium non dopé nous ne pouvons pas dire que les structures mesurées étaient celle attendu du procédé de fabrication.

VI.5. Conclusion

Dans ce chapitre nous avons présenté nos études sur les cellules 3D A2RAM : de la définition du procédé de fabrication jusqu'à la caractérisation électrique. Nous avons commencé l'étude par la présentation du procédé de fabrication de cellules 3D A2RAM, basé sur celui de transistors nanofils du CEA-Leti [Barraud 13]. Quatre nouvelles étapes ont été ajoutées à ce procédé pour fabriquer des dispositifs A2RAM. Les trois premières étapes (dopage de la zone active, épitaxie et gravure du body) ont été optimisées grâce à des simulations TCAD en tenant compte des contraintes technologiques imposées par le procédé de fabrication, et celles données par la caractérisation électrique des opérations mémoires.

Trois types de caractérisations électriques ont été effectuées : statique (I_d - V_g), petit signal (C_{gg} - V_g), et dynamique (opération mémoire). Grâce aux caractérisations statiques et petits signaux, nous avons pu mettre en évidence la présence du bridge dans nos structures. Ce résultat nous a permis de valider nos recettes d'implantation du bridge définies par simulation TCAD. Nous avons aussi pu montrer grâce aux caractérisations électriques petits signaux, la présence d'un poly silicium non dopé dans l'empilement de la grille. Ceci a rendu

impossible l'utilisation de la méthodologie d'extraction de profil de dopage de bridge développée dans le Chapitre 4, mais aussi l'opération d'effacement.

La programmation et la lecture de l'état '1' ont pu être mises en évidence, bien que l'amplitude et le temps de rétention du courant I_1 étaient dégradés principalement à cause du polysilicium non dopé. Cependant, l'étude approfondit des opérations de programmation et lecture de l'état '1' a permis de valider les choix de split technologiques (L_g , T_{body} , T_{bridge} , implantation LDD).

CONCLUSION GENERALE

Conclusion générale

La réduction des dimensions de la cellule 1T/1C-DRAM nécessite des procédés de plus en plus complexes et donc coûteux pour intégrer la capacité (*stack* et *trench*). Pour résoudre ce problème, d'autres architectures mémoire DRAM de type 1T-DRAM ont vu le jour, parmi lesquelles : l'A2RAM. La cellule A2RAM a été proposée par [Rodriguez 11], son architecture est similaire à un transistor MOSFET sur substrat SOI, avec le film de silicium divisé en deux parties. La partie supérieure appelée *body* est dopée de type opposé aux source et drain, et la partie inférieure appelée *bridge* est dopée de même type que la source et le drain.

Dans le premier chapitre, nous avons présenté l'architecture d'un transistor MOSFET en expliquant son fonctionnement : caractéristiques statiques (I_d-V_g et I_d-V_d) et les effets parasites (courants de fuite des diodes, courant de fuite par GIDL, et ionisation par impact) tous nécessaires pour mieux appréhender le fonctionnement d'une 1T-DRAM. Ensuite, nous avons retracé partiellement l'historique de la miniaturisation du transistor MOSFET qui est régie par la loi de Moore. Nous avons noté que partant d'une architecture planaire sur substrat massif, lorsqu'on réduit la longueur de la grille le contrôle électrostatique de la grille sur le canal est dégradé. Ce qui entraîne la réduction de la tension seuil et l'augmentation des effets parasites. Une solution pour résoudre ces problèmes est l'introduire de nouvelles architectures de transistor MOSFET :

- le transistor PDSOI : en plus d'améliorer le contrôle électrostatique, ce transistor a l'avantage d'avoir une couche d'oxyde en-dessous des zones source et drain ce qui élimine les courants de fuite dus aux diodes source (drain)/canal. Cependant, durant son fonctionnement, le film de silicium s'autopolarise à cause de l'effet Kink [Colinge 85], [Colinge 88]. Ceci peut être résolu en ajoutant un contact sur le film de silicium, mais il se traduit par une augmentation de la surface occupée sur le silicium et des coûts de fabrication.
- le transistor FDSOI : il permet d'avoir un meilleur contrôle électrostatique comparé à un transistor PDSOI. De plus, lorsque l'oxyde enterré est mince, le couplage des grilles avant et arrière nous permet d'augmenter le contrôle électrostatique sur le canal et d'ajuster sa tension seuil.
- le transistor nanofil : pour les nœuds technologiques inférieurs à 7 nm ($L_g < 10$ nm) [Lacord 12], les effets de canaux courts sont importants, et même le transistor FDSOI ne peut plus garantir une bonne fonctionnalité. C'est pourquoi les transistors nanofils ou transistors à grille enrobante sont proposés comme garant d'un bon contrôle électrostatique pour les futurs nœuds technologique [ITRS 15].

Par la suite nous avons retracé l'historique qui a menée à l'introduction de l'A2RAM en passant par l'explication du fonctionnement d'une 1T-DRAM. Il en ressort que les différents mécanismes mis en exergue durant la programmation de l'état '1' tirent avantage des effets parasites de la partie flottante du film de silicium flottant du transistor MOSFET [Bawedin 11]. Pour la programmation de l'état '0', la charge piégée dans le *body* est évacuée en polarisant les diodes (drain) source/film de silicium en direct ou en utilisant le couplage dynamique entre les grilles face avant et arrière.

La première démonstration expérimentale d'une architecture 1T-DRAM fut la ZRAM, mais la dégradation de ses performances lorsqu'on réduit ses dimensions a conduit à la proposition de plusieurs architectures 1T-DRAM. Parmi elles, nous avons la MSDRAM [Bawedin 05], mais cette dernière souffre de super-coupling [Eminente 07], [Navrro 15] lorsqu'on réduit ses

dimensions. C'est ainsi que dans un premier temps, l'ARAM [Rodriguez 10] a été proposée mais la fabrication de cette architecture mémoire demanderait l'introduction de nouveaux procédés et de jeux de masque qui augmentent son coût, d'où finalement le développement de la cellule A2RAM.

Par la suite dans notre Chapitre 2, nous avons fait une étude approfondie d'un point mémoire 1T-DRAM via l'outil de simulation TCAD. Nous avons débuté en présentant le protocole de simulation : structure simulée, motif de tensions pour les opérations mémoires et modèles utilisés pour les simulations. Notons que pour rester en accord avec le Chapitre 1, des différents mécanismes existants dans la littérature pour stocker l'état '1', pour l'A2RAM nous avons choisi la génération de la charge par effet tunnel bande à bande. Notre motivation pour ce mécanisme est principalement une bonne fiabilité vis-à-vis de l'oxyde de grille comparé au mécanisme de génération de la charge par ionisation par impact [Puget 09]. Une fois notre stratégie de simulation validée, nous avons fait une étude de sensibilité des performances mémoires lorsqu'on varie les différentes polarisations des opérations mémoires : écriture de l'état '1', maintien de l'information, lecture de l'information, effacement de l'information, et polarisation de la grille face arrière. Cette étude du point mémoire A2RAM nous a permis de mieux appréhender son fonctionnement, et il en ressort que :

- pour garantir un stockage efficace en 10 ns par la génération par effet tunnel bande à bande il faut utiliser de fortes polarisations sur le drain (> 3 V) et la grille (< -1.7 V).
- le profil de dopage du bridge est le paramètre qui impacte le plus les performances mémoires de l'A2RAM.

Au cours du Chapitre 3, toujours à travers l'outil TCAD, nous avons présenté un nouveau mécanisme de programmation de l'état '1' qui dépend cette fois de la génération par ionisation par impact ayant lieu à l'interface bridge/drain. Grâce à ce nouveau mécanisme, nous pouvons garantir un stockage en moins de 10 ns à basses polarisations sur le drain (≥ 1 V) et sur la grille (≤ -0.5 V). Par la suite, pour pouvoir optimiser les dimensions de la cellule A2RAM, nous avons fait une étude de sensibilité des variations des paramètres technologiques de l'A2RAM sur ces performances mémoires.

En effet, la réduction de la longueur de la grille (L_g) de l'A2RAM comme pour un transistor MOSFET de la logique dégrade le contrôle électrostatique de la grille sur le film de silicium, à cause de l'apparition des effets canaux courts. La conséquence directe est l'augmentation du courant lu de l'état '0' I_0 et la réduction du temps de rétention qui passe de 0.44 ms pour $L_g = 80$ nm à 0.3 μ s pour $L_g = 30$ nm. Pour améliorer le contrôle électrostatique, nous avons tour à tour évalué l'impact de la réduction de l'EOT, T_{body} , T_{bridge} , N_{bridge} . Ceci nous a permis par la suite de proposer des structures A2RAM optimisées, ayant toutes une longueur de la grille de 30 nm pour garantir une forte densité d'intégration. En fonction de la valeur des épaisseurs du body et du bridge, on peut obtenir un temps de rétention allant jusqu'à 3 ms qui reste tout de même inférieur à la valeur minimale d'environ 65 ms d'un temps de rétention dans une cellule 1T-1C DRAM [Kinam 09]. A la fin de ce chapitre, nous avons examiné la possibilité d'améliorer les performances de la cellule A2RAM, en particulier la valeur du courant lu de l'état '1' I_1 . Une des possibilités consiste à utiliser une structure A2RAM de type p avec une mince couche (1 nm à 2 nm) de bore dopé à dégénérescence (10^{20} cm⁻³ à $5 \cdot 10^{20}$ cm⁻³) [Nanver 10] au fond du film de silicium. L'autre possibilité consiste à introduire un matériau à faible gap d'énergie dans le body, dans ce travail de thèse nous avons

Conclusion générale

travaillé avec du silicium-germanium (SiGe) unique matériau III-V disponible en salle blanche du Leti si on envisage fabriquer un démonstrateur. Une première étude nous a permis de constater que la concentration idéale du germanium (xGe) était de 30%. Ensuite, nous avons noté que le gain apporté par le SiGe intégré dans le body était essentiellement l'augmentation de la charge stockée, par conséquent l'augmentation de la valeur du courant I_1 .

De notre étude du Chapitre 2 et du Chapitre 3, nous avons eu confirmation que le profil de dopage en profondeur du film de silicium impacte les performances mémoires de l'A2RAM. Dans l'optique d'analyser les résultats de nos caractérisations électriques, nous avons développé une technique d'extraction du profil de dopage dans une structure de type SOI. Cette étude a fait l'objet du Chapitre 4, où nous avons premièrement étudié la caractéristique petit signal de l'A2RAM via l'outil TCAD. Ceci nous a permis de déterminer les conditions électriques nécessaires afin d'adapter la formule de Marsajien [Maserjian 74] pour extraire le dopage du bridge, et ensuite extraire les épaisseurs du body et du bridge. Nous avons aussi analysé les limites de validité de notre technique d'extraction : elle garantit des résultats fiables pour un dopage du bridge supérieur à 10^{18} cm^{-3} et une épaisseur du bridge supérieure à 20 nm. Cette technique a été ensuite utilisée sur nos échantillons de cellules A2RAM précédemment fabriqués au Leti ; les résultats obtenus étaient cohérents avec nos prédictions TCAD. Par ailleurs, grâce à cette technique nous avons pu mettre en évidence la possibilité d'augmenter la valeur du courant I_1 , tout en maintenant la valeur de I_0 toujours basse.

La connaissance de la physique qui entre en jeu durant le fonctionnement de l'A2RAM, étudiée dans les premiers chapitres de ce manuscrit, nous ont permis de développer un modèle compact de l'A2RAM dans le Chapitre 5. La stratégie suivie est un macro-modèle qui repose sur la combinaison d'un modèle compact du comportement statique de l'A2RAM avec un comparateur pour modéliser la dynamique du changement d'état mémoire [Martinie 17]. Le modèle compact de la caractéristique statique (I_d - V_g) de l'A2RAM est obtenu en associant le modèle compact d'un transistor MOSFET sur substrat SOI en l'occurrence le modèle MASTAR VA [Lacord 14], en parallèle avec une résistance variable pour tenir compte de l'impact du bridge. Ce modèle repose essentiellement sur la modélisation précise de la tension seuil du bridge qui a été obtenue grâce à la résolution de l'équation de Poisson 1D dans toute la profondeur de l'empilement vertical de la structure de l'A2RAM. Le modèle complet d'opérations mémoires est implémenté en Verilog-A, ensuite il est calibré et validé via des simulations TCAD. Pour finir, nous avons utilisé notre modèle dans un simulateur de circuit commercial type SPICE : ELDO [ELDO]. Nous avons fait une étude de cellules A2RAM intégrées dans une matrice 2x2, les résultats obtenus sont accord avec le comportement attendu de notre modèle.

Enfin, dans le Chapitre 6, nous avons utilisé notre expertise acquise durant ce travail de thèse pour servir de support à l'équipe d'intégration du CEA-Leti et définir le cahier de charge pour la fabrication d'un nouveau lot de mémoire A2RAM. Mais aussi, pour caractériser les cellules mémoire une fois fabriquées. La définition du cahier de charge a été effectuée en tenant compte des contraintes du procédé de fabrication (limite sur la lithographie et le budget thermique) et de celles imposées par la caractérisation électrique mémoire ($I_1 > 6 \mu\text{A} \cdot \mu\text{m}$ et $I_1/I_0 > 20$). Ainsi, nous avons fait des simulations TCAD pour dimensionner les épaisseurs

Conclusion générale

du body et bridge, mais également la longueur des zones de recouvrements drain et source. Le bridge étant crucial dans le fonctionnement de l'A2RAM, nous avons fait des simulations de procédé de fabrication pour optimiser la recette d'implantation du bridge et les recuits d'activation associés. La caractérisation électrique des échantillons nous a permis de valider dans un premier temps le cahier de charge. En mêlant caractérisation électrique et simulations TCAD, nous avons mis en évidence la présence d'un polysilicium de grille non dopée. Ceci explique en partie pourquoi les performances mémoires obtenues étaient très inférieures de celles prédites par simulations TCAD.

De ce travail de thèse, il reste encore des questions ouvertes, en particulier sur la caractérisation électrique. Les limites imposées par l'instrument de mesure au niveau de la vitesse des variations des signaux à appliquer sur la cellule mémoire n'ont pas permis d'évaluer les performances de cellule A2RAM dans les conditions d'une cellule DRAM standard. Cette limitation est en partie responsable de l'impossibilité de pouvoir mettre en évidence expérimentalement notre nouveau mécanisme de programmation. Par ailleurs, le modèle compact proposé dans ce travail de thèse permet de modéliser seulement les opérations d'écriture des états '1' et '0', et la lecture de l'information. Les opérations d'effacement et de programmation et leur dynamique respective, mais aussi la modélisation de l'opération de maintien de l'information (pour des études d'endurance) sont des voies d'optimisations du modèle compact de l'A2RAM.

Pour la suite de ce travail de thèse, l'aboutissement de notre modèle compact de l'A2RAM pourrait être utilisé dans les simulateurs commerciaux SPICE pour évaluer et optimiser ses performances dans un environnement matriciel. Ce travail permettrait aussi de faire une étude d'optimisation des autres circuits auxiliaires nécessaires à la sélection, la lecture, et l'écriture de l'information du point mémoire sous matrice. Pour finir, il serait intéressant de reprendre nos travaux du Chapitre 6 pour fabriquer un nouveau lot de cellules A2RAM. Pour la caractérisation électrique mémoire de ce nouveau lot, il faudrait utiliser les mêmes conditions que celles d'une cellule DRAM 1T/1C afin de vérifier expérimentalement si les performances de la cellule A2RAM sont à mêmes de pouvoir rivaliser avec celles des DRAM déjà présentes sur le marché.

BIBLIOGRAPHIE

- [Arnaud 18] F. Arnaud et al., 'Truly Innovative 28nm FDSOI Technology for Automotive Micro-Controller Applications embedding 16MB Phase Change Memory', 2018 IEEE International Electron Devices Meeting.
- [Ban 08] I. Ban et al., 'Scaled Floating Body Cell (FBC) Memory with High-k+Metal Gate on Thin-Silicon and Thin-BOX for 16-nm Technology Node and Beyond', VLSI Symp. Circuit Dig., Jun. 2008, pp. 92-93.
- [Barraud 13] S. Barraud et al., 'Scaling of Ω -Gate SOI Nanowire N- and P-FET down to 10nm Gate Length: Size- and Orientation-Dependent Strain Effects', 2013 Symposium on VLSI Technology Digest of Technical Papers.
- [Bawedin 05] M. Bawedin et al., 'A new memory effect (MSD) in fully depleted SOI MOSFETs', Solid-State Electronics 49 (2005) 1547-1555.
- [Bawedin 07] M. Bawedin et al., 'Innovating SOI memory devices based on floating-body effects', Solid-State Electronics 51 (2007) 1252-1262.
- [Bawedin 08] M. Bawedin et al., 'A Capacitorless 1T-DRAM on SOI Based on Dynamic Coupling and Double-Gate Operation', IEEE Electron Device Letters, Vol. 29, no. 7, July 2008
- [Bawedin 09] M. Bawedin et al., 'Floating-Body SOI Memory: Concepts, Physics and Challenges', ECS Transactions, 19 (4) 243-256 (2009).
- [Bawedin 11] M. Bawedin et al., 'Floating-Body SOI Memory: The Scaling Tournament', Semiconductor-On-Insulator Materials for Nanoelectronics Applications, Springer-Verlag Berlin Heidelberg 2011.
- [Bouhdad 97] A Bouhdad et al., 'Modelling of Gate-Induced Drain Leakage in relation to technological parameters and temperature', Microelectron. Reliab., Vol. 37, No. 4, pp. 649-652, 1997.

BIBLIOGRAPHIE

- [Chan 87] T.Y. Chan et al., 'The impact of gate-induced drain leakage current on MOSFET scaling', 1987 International Electron Devices Meeting.
- [Chang 87] C. Chang et al., 'Corner-field induced drain leakage in thin oxide MOSFET's', 1987 International Electron Devices Meeting.
- [Chattopadhyay 03] S. Chattopadhyay et al., 'C-V characterization of strained Si/SiGe multipleheterojunction capacitors as a tool for heterojunction MOSFET channel design', *Semicond Sci Technol* 2003 ; 18:738-44.
- [Chen 01] J.-H. Chen et al., 'An Analytic Three-Terminal Band-to-Band Tunneling Model on GIDL in MOSFET', *IEEE Transactions on electron devices*, Vol. 48, no. 7, July 2001.
- [Colinge 85] J.-P. Colinge, 'Reduction of floating substrate in thin-film SOI MOSFETs', *ELECTRONICS LETTERS* 13th February 1986 Vol. 22 No. 4.
- [Colinge 88] J.-P. Colinge, 'Reduction of Kink Effect in Thin-Film SOI MOSFET's', *IEEE Electron Device Letters*, Vol. 9, no. 2 February 1988.
- [Colinge 91] J.-P. Colinge, 'SILICON-ON-INSULATOR TECHNOLOGY: Materials To VLSI' Copyright © 1991 by Springer Science+Business Media New York.
- [Cooper 16] D. Cooper, 'Off-axis electron holography for the measurement of active dopants in silicon semiconductor devices', *J. Phys. D: Appl. Phys.* 49 474001.
- [Cristoloveanu 86] S. Cristoloveanu et al., 'Profiling of inhomogeneous carrier transport properties with the influence of temperature in silicon-on-insulator films formed by oxygen implantation', *J Appl Phys* 1986 ; 60:3199.
- [Dennard 68] R. H. Dennard, « Filed-Effect Transistor Memory », patent, 1968.

BIBLIOGRAPHIE

- [Dutartre 12] D. Dutartre et al., 'Faceting and nanostructure effects in Si and SiGe epitaxy', *Thin Solid Films* 520 (2012) 3163–3169.
- [ELDO] ELDO, www.mentor.com
- [Eminente 07] S. Eminente et al., 'Ultra-thin fully-depleted SOI MOSFETs: Special charge properties and coupling effects', *Solid-State Electronics* 51 (2007) 239–244
- [Endoh 90] T. Endoh et al., 'An Accurate Model of Subbreakdown Due to Bandto-Band Tunneling and Some Applications', *IEEE Transactions on electron devices*, Vol. 37, no. 1, January 1990.
- [Ertosun 08] M. G. Ertosun et al., 'A Highly Scalable Capacitorless Double Gate Quantum Well Single Transistor DRAM: 1T-QW DRAM', 0741-3106/\$25.00 © 2008 IEEE.
- [Fenouillet 12] C. Fenouillet-Beranger et al., 'Enhancement of Devices Performance of hybrid FDSOI/Bulk Technology by using UTBOX sSOI substrates', in *VLSI Tech. Dig.*, 115-116, 2012.
- [Fossum 87] J. G. Fossum, 'Anomalous subthreshold current-voltage characteristics of n-channel SOI MOSFET's', *IEEE Electron Device Letters*, Vol. EDL-8. no. 11, November 1987.
- [Fredeman 16] G. Fredeman et al., 'A 14 nm 1.1 Mb Embedded DRAM Macro With 1 ns Access', *IEEE Journal of solid-state circuits*, Vol. 51, no. 1, January 2016.
- [Gelder 71] W. V. Gelder et al., 'Silicon impurity distribution as revealed by pulsed MOS C-V measurements', *Solid State Sci* 1971.
- [Ghetti 18] A. Ghetti et al., 'Emerging Memory Modeling Challenges', in *SISPAD* 2018.

BIBLIOGRAPHIE

- [Golonzka 18] O. Golonzka et al., 'MRAM as Embedded Non-Volatile Memory Solution for 22FFL FinFET Technology', 2018 IEEE International Electron Devices Meeting.
- [GUY 15] J. Guy, « Evaluation des performances des mémoires CBRAM (Conductive Bridge Memory) afin d'optimiser les empilements technologiques et les solutions d'intégration », thèse, 2015.
- [IBM SRC eDRAM 09] J. Barth et al., 'eDRAM to the Rescue', SRC eDRAM 2009.
- [Intel, EETimes 14] R. Brain et al., 'Integration of a 3-D Capacitor into a Logic Interconnect Stack for High performance Embedded DRAM SoC Technology', IEEE International Interconnect Technology Conference, 20-23 May 2014.
- [ITRS 07] <http://www.itrs2.net/itrs-reports.html>
- [ITRS 15] <http://www.itrs2.net/itrs-reports.html>
- [Jang 05] M. W. Jang et al., 'Enhancement of Data Retention Time in DRAM using Step gated Asymmetric(STAR) Cell Transistors', Proceedings of ESSDERC, Grenoble, France, 2005.
- [Jang 09] T.-S. Jang et al., 'Highly scalable Z-RAM with remarkably long data retention for DRAM application', VLSI Symp. Tech. Dig., Jun. 2009, pp. 234-235.
- [Kane 61] E. O. Kane et al., 'Theory of Tunneling', Journal of Applied Physics, vol. 32, no. 1, pp. 83-91, 1961.
- [Kang 00] L. Kang et al., 'Electrical Characteristics of Highly Reliable Ultrathin Hafnium Oxide Gate Dielectric', IEEE Electron Device Letters Vol. 21, no. 4, April 2000.

BIBLIOGRAPHIE

- [KATO 85] K. Kato et al., 'Analysis of Kinif Characteristics in Silicon-on-Insulator MOSFET's Using Two-Carrier Modeling', IEEE Transactions on electron devices, Vol. ED-32, No. 2, February 1985.
- [Kawasaki 09] H. Kawasaki et al., 'Challenges and Solutions of FinFET Integration in an SRAM Cell and a Logic Circuit for 22 nm node and beyond', 2009 IEEE International Electron Devices Meeting.
- [Kim 09] K. Kim et al., 'A New Investigation of Data Retention Time in Truly Nanoscaled DRAMs', IEEE Electron Device Letters, Vol. 30, No. 8, August 2009.
- [Kurimoto 89] K. Kurimoto et al., 'Drain leakage current characteristics due to the band-to-band tunneling in LDD MOS devices', 1989 International Technical Digest on Electron Devices Meeting.
- [Lacord 12] J. Lacord, « Développement de modèles pour l'évaluation des performances circuit des technologies CMOS avancées sub-20nm », thèse, 2012.
- [Lacord 14] J. Lacord et al., 'MASTAR VA: A predictive and flexible compact model for digital performances evaluation of CMOS technology with conventional CAD tools', Solid-State Electronics, vol. 91, pp. 137-146, 2014.
- [Lacord 18] J. Lacord et al., 'MSDRAM, A2RAM and Z2-FET performance benchmark for 1T-DRAM applications', 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Austin, TX, 2018, pp. 198-201.
- [Lee 18] K. Lee et al., '22-nm FD-SOI Embedded MRAM Technology for Low-Power Automotive-Grade-1 MCU Applications', 2018 IEEE International Electron Devices Meeting.
- [Lu 10] Z. Lu et al., 'A Novel Low-Voltage Biasing Scheme for Double Gate FBC Achieving 5s Retention and 10¹⁶ Endurance at 85°C', 2010 International Electron Devices Meeting.

BIBLIOGRAPHIE

- [Lu 16] C.-M. V. Lu et al., 'Integration of Low Temperature SiGe:B Raised Sources and Drains in p-type FDSOI Field Effect Transistors', ECS Transactions, 75 (8) 51-58 (2016).
- [Makarov 11] A. Markarov et al., 'Emerging memory technologies Trends challenges and modeling methods', Microelectronics Reliability Volume 52, Issue 4, April 2012, Pages 628-634.
- [Malinge 05] P. Malinge, « Etude et modélisation d'un point mémoire eDRAM sans capacité, et conception de circuit mémoire haute densité », thèse, 2005.
- [Malinge 05] P. Malinge et al., 'An 8Mbit DRAM design using a 1TBulk cell', 2005 Symposium on VLSI Circuits Digest of Technical Papers.
- [MarketersMEDIA 18] <https://marketersmedia.com/dram-market-2018-global-growth-industry-size-segmentation-competitive-landscape-sales-revenue-development-history-regional-analysis-by-forecast-to-2023/452647>.
- [Martinie 09] S. Martinie, « Modélisation du transport quasi-balistique pour la simulation de circuits à base de nano-transistor multigrilles », thèse, 2009.
- [Martinie 17] S. Martinie et al., 'Z2-FET Compact Model: DC and Memory Operation', in ESSDERC 2017.
- [Maryline 11] A. Nazarov et al., 'Semiconductor-On-Insulator Materials for Nanoelectronics Applications', DOI 10.1007/978-3-642-15868-1.
- [Maserjian 74] J. Maserjian et al., 'Saturation capacitance of thin oxide MOS structures and the effective surface density of states of silicon', Solid-State Electron 1974;17:335-9.

BIBLIOGRAPHIE

- [Masetti 83] G. Masetti et al., 'Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-, Phosphorus-, and Boron-Doped Silicon', IEEE Transactions on electron devices, Vol. 30, no. 7, July 1983.
- [Mathieu 04] H. Mathieu « Physique des semiconducteurs et des composants électroniques », éditions Dunod, 2004.
- [Matsuoka 07] F. Matsuoka et al., 'FBC's Potential of 6F2 Single Cell Operation in Multi-Gbit Memories Confirmed by a Newly Developed Method for Measuring Signal Sense Margin', IEDM Tech. Dig., 2007, p. 39.
- [Mohapatra 09] N. R. Mohapatra et al., 'Effect of Source-Drain Asymmetry on the Performance of zram devices' 2009 IEEE International SOI Conference.
- [Nanver 10] L. K. Nanver et al., 'Pure-Boron Chemical-Vapor-Deposited Layers: a New Material for Silicon Device Processing', 18th IEEE Conference on Advanced Thermal Processing of Semiconductors - RTP 2010.
- [Navarro 14] C. Navarro, 'Design, simulation and electrical characterization of A2RAM memory cells', thèse 2014.
- [Navarro 15] C. Navarro et al., 'Supercoupling effect in short-channel ultrathin fully depleted silicon-on-insulator Transistors', Journal of Applied Physics 118, 184504 (2015).
- [Navarro 15] C. Navarro et al., 'Channel-Length Impact on Supercoupling Effect in FD-MOSFETs', EUROSOI 2014, January 28-29, 2014 Tarragona, Spain.
- [Okhonin 01] S. Okhonin et al., 'A SOI capacitor-less 1T-DRAM concept', 2001 IEEE international SOI Conference.
- [Okhonin 02] S. Okhonin et al., 'A Capacitor-Less 1T-DRAM Cell', IEEE Electron Device Letters, Vol. 23, No. 2, February 2002.

BIBLIOGRAPHIE

- [Okhonin 07] S. Okhonin et al., 'New Generation of Z-RAM', 2007 IEEE International Electron Devices Meeting.
- [Okhonin 08] S. Okhonin et al., 'Ultra-scaled Z-RAM cell', 2008 IEEE International SOI Conference Proceedings.
- [Overstraeten 70] R. V. Overstraeten et al., 'Measurement of the ionization rates in diffused silicon p-n junctions', Solid-State Electronics Pergamon Press 1970. Vol. 13, pp. 583-608.
- [Park 05] S. G. Park et al., 'The New Technology for DRAM Cell Transistor with S-RCAT and its Size Effect', Extended Abstracts of the 2005 International Conference on Solid State Devices and Materials, Kobe, 2005, pp.624-625.
- [Park 06] D. Park 06 et al., 'Stack DRAM Technologies for the Future', IEEE 2006 International Symposium on VLSI Technology, Systems, and Applications.
- [Planes 12] N. Planes et al. "28nm FDSOI technology platform for high-speed low voltage digital applications" in VLSI Technology (2012), pp. 133-134.
- [Puget 09] S. Puget et al., 'FDSOI Floating Body Cell eDRAM Using Gate-Induced Drain-Leakage (GIDL) Write Current for High Speed and Low Power Applications', 2009 IEEE International Memory Workshop.
- [Rafhay 12] Q. Rafhay et al., 'Revisited approach for the characterization of Gate Induced Drain Leakage', Solid-State Electronics Volume 71, May 2012, Pages 37-41.
- [Ranica 04] R. Ranica et al., 'A one transistor cell on bulk substrate (1T-Bulk) for low-cost and high density eDRAM', Digest of Technical Papers. 2004 Symposium on VLSI Technology.
- [REM20] <http://www.reminder2020.eu/deliverables-2/>

BIBLIOGRAPHIE

- [Rodriguez (1) 11] N. Rodriguez et al., 'NEW CONCEPTS FOR 1T-DRAMs: OVERCOMING THE SCALING LIMITS', 978-1-61284-172-4/11/\$26.00 © 2011 IEEE.
- [Rodriguez (2) 11] N. Rodriguez et al., 'Novel Capacitorless 1T-DRAM Cell for 22-nm Node Compatible With Bulk and SOI Substrates', IEEE Transactions on electron devices, Vol. 58, no. 8, August 2011.
- [Rodriguez 10] N. Rodriguez et al., 'A-RAM Memory Cell: Concept and Operation', IEEE Electron Device Letters, Vol. 31, no. 9, September 2010.
- [Rodriguez 11] N. Rodriguez et al., 'Capacitor-less A-RAM SOI memory: Principles, scaling and expected performance', Solid-State Electronics 59 (2011) 44–49.
- [Rodriguez 12] N. Rodriguez et al. 'Experimental Demonstration of Capacitorless A2RAM Cells on Silicon-on-Insulator', IEEE Electron Device Letters, Vol. 33, No. 12, December 2012.
- [shenck 09] A. shenck, 'Suppression of Gate-Induced Drain Leakage by Optimization of Junction Profiles in 22 nm and 32 nm SOI nFETs', Preprint submitted to Elsevier 21 August 2009.
- [Shin 11] J. S. Shin et al, 'Vertical-Gate Si/SiGe Double-HBT-Based Capacitorless 1T DRAM Cell for Extended Retention Time at Low Latch Voltage', IEEE Electron Device Letters Vol. 33, no. 2, February 2012.
- [SILVACO] <https://www.silvaco.com/>
- [Song 08] K.-W. Song et al., '55 nm capacitor-less 1T DRAM cell transistor with non-overlap structure', IEDM Technical Dig., Dec. 2008, pp. 1-4.
- [SPICE] <https://www.spice-space.org/spice-user-manual.html>
- [SYNOPSYS] TCAD Sentaurus Device Manual, Synopsys, Inc.: J-2014.09.

BIBLIOGRAPHIE

- [SZE 07] S. M. Sze "Physics of Semiconductor Devices" 2nd Edition, 1981.
- [TAUR 09] Y. Taur, "Fundamentals of Modern" VLSI Devices " 2nd Edition, 2009.
- [Tihanyi 75] J. Tihanyi et al, 'Properties of ESFI MOS transistors due to the floating substrate and the finite volume', IEEE Transactions on electron devices, Vol. ED-22, no. 11, November 1975.
- [Villaret 04] A. Villaret, 'Nouvelles architectures de mémoires embarquées compatibles CMOS', thèse 2004.
- [Wan 13] J. Wan et al, 'A systematic study of the sharp-switching Z2-FET device: From mechanism to modeling and compact memory applications', Solid. State. Electron., Vol. 90, Dec 2013, pp. 2–11.
- [Wann 93] H.-J. Wann et al, 'A Capacitorless DRAM Cell on SOI Substrate', 1993 IEEE International Electron Devices Meeting.
- [Weber 08] O. Weber et al, 'High Immunity to Threshold Voltage Variability in Undoped Ultra-Thin FDSOI MOSFETs and its Physical Understanding', In 2008 International Electron Devices Meeting. Technical Digest, pages 245 – 248.
- [Wendell 85] Wendell et al, "Fundamental Limitations on DRAM Storage Capacitors", IEEE Circuits and devices magazine ; January 1985.
- [Yoshida 03] E. Yoshida et.al, 'A design of a capacitorless 1T-DRAM cell using gate-induced drain leakage (GIDL) current for low-power and high-speed embedded memory', IEDM 2003, pp. 913.

PUBLICATIONS DE L'AUTEUR

Articles de journaux

- [Tchame Wakam 19] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, G. Ghibaudo, J.-Ch. Barbe, 'Doping profile extraction in thin SOI films: Application to A2RAM' in Solide State Electronics 2019.
- [Tchame Wakam 19] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, and T. Poiroux, 'A2RAM Compact Modeling: From DC to 1T-DRAM Memory Operation', in progress for Solide State Electronics 2019.

Conférences internationales

- [Tchame Wakam 17] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, and J.-Ch. Barbe, 'Optimization guidelines of A2RAM cell performance through TCAD simulations', 2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Kamakura, 2017, pp. 329-332.
- [Tchame Wakam 18] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, G. Ghibaudo, J.-Ch. Barbe, 'Doping profile extraction in thin SOI films: Application to A2RAM' 2018 Joint International EUROSIO Workshop and International Conference on Ultimate Integration on Silicon (EUROSIO-ULIS).
- [Tchame Wakam 18] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, and J.-Ch. Barbe, 'Evidence of fast and low-voltage A2RAM '1' state programming', 2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), Austin, TX, 2018, pp. 115-118.
- [Tchame Wakam 19] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, and T. Poiroux, 'Modeling of the bridge threshold voltage in A2RAM cell', 2019 Joint International EUROSIO Workshop and International Conference on Ultimate Integration on Silicon (EUROSIO-ULIS).

Contribution

- [Lacord 18] Joris Lacord; Mukta Singh Parihar ; Carlos Navarro ; **François Tchame Wakam** ; Maryline Bawedin ; Sorin Cristoloveanu ; Fransisco Gamiz ; Jean-Charles Barbé, 'MSDRAM, A2RAM and Z2-FET performance benchmark for 1T-DRAM applications', 2018 International Conference on Simulation

PUBLICATIONS DE L'AUTEUR

of Semiconductor Processes and Devices (SISPAD), Austin, TX, 2018, pp. 198-201.

Poster

[Tchame Wakam 19] F. Tchame Wakam, J. Lacord, M. Bawedin, S. Martinie, S. Cristoloveanu, and T. Poiroux, 'Pure boron monolayer to boost A2RAM performance', 2019 Joint International EUROSIO Workshop and International Conference on Ultimate Integration on Silicon (EUROSIO-ULIS).

Brevet

[Lacord, Tchame Wakam] Method for programming a memory cell has a dram memory transistor and device

TITRE: Modélisation, simulation et caractérisation électrique de cellule mémoire DRAM 1T: A2RAM

RESUME

Avec la croissance de transfert de données, principalement à cause des applications de type internet des objets, il y a un besoin accru de système de stockage (mémoires). L'idéal est d'avoir une mémoire spécifique qui sera facilement intégrée dans ces applications. Cette mémoire doit respecter des exigences spécifiques telles que : une simplicité du mode de fonctionnement, une grande densité d'intégration, une faible consommation électrique, et un bas coût de fabrication. Une mémoire capable de répondre à toutes ces exigences n'existe pas, mais une architecture mémoire qui se rapproche de ces critères est la cellule mémoire dynamique (DRAM) intégrée (eDRAM). La DRAM a été proposée pour la première fois en 1968 dans son architecture traditionnelle 1T/1C-DRAM où le transistor sert de point d'accès et la capacité est le point de stockage à l'information. Mais le principal problème de cette architecture est sa faible densité d'intégration causée par la limitation de la miniaturisation de sa capacité de stockage. Une solution pour contourner cette limitation serait l'utilisation d'architectures DRAM sans capacité de stockage: on parle de 1T-DRAM. Ici, le transistor est utilisé pour stocker et lire l'information. On trouve dans la littérature de nombreuses architectures DRAM 1T, mais le but de cette thèse est d'étudier l'A2RAM par simulation à éléments finis (TCAD), caractérisation électrique et modélisation compacte afin de voir si elle peut être utilisée en tant que DRAM intégrée.

Mots clés: IOT, CMOS, DRAM, TCAD, SPICE, modèle compact, 1T-DRAM, A2RAM, électrostatique, caractérisation électrique.

TITLE: Modeling, simulation and electrical characterization of 1T-DRAM cell: A2RAM

ABSTRACT

The growth of transferring data, mainly due to applications such as Internet of Things, there is an increased need for storage system (memories). Ideally, we need a specific memory which would be integrated easily in all these applications. This memory must meet specific requirements such as: a simplicity of the operating mode, high density of integration, low power consumption, and low cost of manufacturing. A memory able to answer all these requirements does not exist, but a memory that is approaching these criteria is the embedded dynamic memory cell (DRAM) (eDRAM). DRAM has been proposed for the first time in 1968 in its traditional 1T / 1C-DRAM architecture. The transistor serves as an access point and the capacity is the storage point of the information. But, the main problem of this architecture is its low density of integration caused by the limitation of the scalability of its storage capacity. One solution could be the use of DRAM architectures without storage capacity: 1T-DRAM. In this case, the transistor is used to store and read information. Many DRAM 1T architectures can be found in the literature, but the aim of this PhD is to study the A2RAM through finite elements simulations (TCAD), electrical characterization and compact modeling to see if it can be used as embedded DRAM.

Key words : IOT, CMOS, DRAM, TCAD, SPICE, compact model, 1T-DRAM, A2RAM, electrostatic, electrical characterization.