Escola Politécnica da Universidade de São Paulo (EPUSP)
Departamento de Engenharia de Sistemas Eletrônicos (PSI)
Laboratório de Sistemas Integráveis (LSI)

Kátia Regina Akemi Sasaki

Propostas de melhorias de desempenho de célula de memória dinâmica
utilizando um único transistor UTBOX SOI

São Paulo
2013

Kátia Regina Akemi Sasaki

Propostas de melhorias de desempenho de célula de memória dinâmica
utilizando um único transistor UTBOX SOI

Dissertação de mestrado apresentada à
Escola Politécnica da Universidade de
São Paulo para obtenção do título de
Mestre em Ciências.

São Paulo

2013

Kátia Regina Akemi Sasaki

Propostas de melhorias de desempenho de célula de memória dinâmica
utilizando um único transistor UTBOX SOI

Dissertação de mestrado apresentada à
Escola Politécnica da Universidade de
São Paulo para obtenção do título de
Mestre em Ciências.

Área de concentração: Microeletrônica

Orientador Profº Dr. João Antonio Martino

São Paulo

2013

Agradecimentos

Gostaria de agradecer ao Profº Dr. João Antonio Martino pela orientação, pelos ensinamentos, pelo constante incentivo transmitido ao longo de todo o trabalho e pela confiança depositada.

Agradeço também aos meus pais sem os quais não teria chegado aqui e aos meus familiares que me apoiaram e ajudaram.

Ao grande amigo Luciano Almeida que foi o responsável, não apenas por parte das medidas (realizadas no imec e aqui apresentadas) e ajustes da estrutura, mas também pelo trabalho e aprendizado conjunto, por todas e tantas discussões realizadas, pela paciência prestada e pela amizade.

À grande amiga Talitha, também responsável por parte das medidas (realizadas no imec e aqui apresentadas), pelo trabalho e aprendizado conjunto, pelas muitas discussões realizadas, paciência e amizade.

Ao também grande amigo Albert pelo trabalho e aprendizado conjunto, incluindo parte das medidas (realizadas na EPUSP) e discussões, pela paciência e amizade.

A Marc Aoulaiche, Eddy Simoen, Cor Claeys e a todos os pesquisadores do imec, pelo auxílio nas medidas, fornecimento de lâminas e pelas discussões e revisões de nossos artigos.

Aos também grandes amigos Sara, Glória, Rangel e Felipe, pelo aprendizado conjunto, pelas muitas discussões, paciência, incentivos e amizade.

Ao meu namorado Ygor pela compreensão, paciência, incentivo, companhia e amizade.

E a todos do grupo SOI que ajudaram de alguma forma nesse trabalho e que, de forma involuntária, foram aqui omitidos.

RESUMO

Neste trabalho foi analisado o comportamento de um transistor UTBOX FD SOI MOSFET (*Ultra-Thin-Buried-Oxide Fully-Depleted Silicon-on-Insulator Metal-Oxide-Semiconductor Field-Effect-Transistor*) planar do tipo N, em sua aplicação como uma célula de memória 1T-DRAM, dando ênfase no estudo das polarizações e propostas de melhorias de desempenho para viabilizar sua aplicação como uma célula de memória. Dessa forma, foram analisados os efeitos das diferentes polarizações (de porta, de dreno e de substrato), bem como a influência da concentração de uma região de extensão de fonte e dreno menos dopada (LDD – *Lightly Doped Drain*), nos principais parâmetros da referida memória. Assim, foram analisados alguns parâmetros da memória tais como tensão de disparo no dreno, margem de sensibilidade, janela de leitura e tempo de retenção, além dos mecanismos atuantes em cada estado da memória (escrita, leitura e repouso). Por fim, foram propostas algumas melhorias de desempenho para o tempo de retenção.

Foi observado que o aumento da temperatura facilita a escrita na memória diminuindo a mínima tensão no dreno (até 72% para temperatura de 25 a 300°C, ficando limitada a 0,8V) e o tempo necessários para a escrita (até 95%), porém reduz a margem de sensibilidade (até 90%) e o tempo de retenção (até 2 ordens de grandeza). Verificou-se também que, apesar da menor espessura do filme de silício e do óxido enterrado aumentar a tensão no dreno necessária para ativar o efeito BJT (efeito bipolar parasitário), um potencial positivo no substrato pode reduzir este requisito (61% para tensão de substrato variando de 0 V até 1,5 V). Além disso, foi visto que pode haver uma geração ou uma recombinação de portadores, dependendo da tensão na porta durante o repouso, degradando o bit '0' ou '1'. Já a otimização da polarização de substrato demonstrou ser limitada pelo compromisso de ser alta o suficiente para ativar o efeito de corpo flutuante durante a escrita, sem prejudicar a leitura do '0'. Os resultados também demonstraram que a margem de sensibilidade é menos dependente da tensão do substrato que o tempo de retenção, levando a este último parâmetro ser considerado mais crítico. Com relação à leitura, maiores tensões no dreno resultaram na presença do efeito BJT também neste estado, aumentando a

margem de sensibilidade (60%) e diminuindo o tempo de retenção (66%) e o número de leituras possíveis sem atualização do dado (de mais de 30 para 22 leituras).

No tópico da concentração das extensões de fonte e dreno, os dispositivos sem extensão de fonte e dreno apresentaram uma taxa de geração de lacunas menor (aproximadamente 12 ordens de grandeza), levando a um tempo de retenção muito maior (aproximadamente 3 ordens de grandeza) quando comparado ao dispositivo referência. Em seu estudo no escalamento, verificou-se uma diminuição no tempo de retenção para canais mais curtos (quase 2 ordens de grandeza), demonstrando ser um fator limitante para as futuras gerações das memórias 1T-DRAM. Apesar disso, quando comparados com os dispositivos convencionais com extensão de fonte e dreno (com extensão), seu tempo de retenção aumentou (quase 1 ordem de grandeza), permitindo a utilização de menores comprimentos de canal (30nm contra 50nm do dispositivo com extensão) e polarizações de substrato menores.

Outra proposta de melhoria no tempo de retenção apresentada foi a utilização da polarização de substrato pulsada apenas durante a escrita do nível '1', o que resultou no aumento do tempo de retenção em 17%.

Finalmente, estudou-se também a variação da banda proibida motivado pela utilização de novos materiais para o filme semicondutor. Observou-se que o aumento da banda proibida aumentou o tempo de retenção em até 5 ordens de grandeza, possibilitando retenções mais próximas das DRAMs convencionais atuais.

Palavras-chave: SOI, MOS, Microeletrônica, Transistores, Memória RAM, Memória,

ABSTRACT

In this work, it was analyzed the behavior of a planar UTBOX FD SOI NMOSFET (*Ultra-Thin-Buried-Oxide Fully-Depleted Silicon-on-Insulator Metal-Oxide-Semiconductor Field-Effect-Transistor*), as a 1T-DRAM (*Single Transistor Dynamic Random Access Memory*) cell, focusing on the best biases and other proposals for enabling the 1T-DRAM applications. Therefore, it was analyzed the effects of different biases (gate, drain and substrate), as well as the influence of the concentration of a less doped source/drain extension region on the main parameters of this kind of memory. Thus, it was analyzed some of the main memory parameters such as the trigger drain voltage, the sense margin, the read window and the retention time, as well as the mechanisms operating in each state of the memory (writing, reading and holding). Finally, it were proposed some performance enhancements for the retention time of this kind of memory.

It was observed that the increase in temperature facilitates the memory write decreasing the minimum drain bias and time required for writing, but reduces the sense margin. It was also verified that, despite the thinner silicon film and buried oxide increase the drain voltage required to activate the BJT effect (parasitic bipolar effect), a positive potential on the substrate may reduce this requirement (61% for back gate bias varying from 0 to 1,5V), being an alternative for solving the problem and allowing the use of smaller devices as a memory cell. Furthermore, it was seen that there can be a carriers generation or recombination, depending on the gate voltage during the holding state, degrading the bit '0' or '1'. Moreover, the optimization of substrate bias proved to be limited by enabling the writing state, without degrading the reading of '0'. The results also demonstrated the sense margin is less dependent on the substrate voltage than the retention time, therefore, the retention time was considered as a more critical parameter. With respect to the reading state, there was the presence of BJT effect also in this state, increasing the margin of sensitivity (60%) and reducing the retention time (66%) and the number of possible readings without updating the data (over 30 for 22 readings) in cases of higher drain bias.

On the topic of the concentration of the source and drain extensions, the devices with source and drain extensions presented a generation rate lower

(about 12 orders of magnitude), resulting in a retention time far longer than the reference one (about 3 orders of magnitude). About its downscaling, the retention time decreased for shorter channel lengths (almost 2 orders of magnitude), which is a limiting factor for 1T-DRAM future generations. Nevertheless, when it was compared to the conventional devices with source and drain extensions, their retention time increased (almost 1 order of magnitude), allowing the use of shorter channel lengths (30nm against 50nm of reference device) and lower back gate biases.

Another proposal presented to improve the retention time was the pulsed back gate only during the writing '1' state, which resulted in an increase on the retention time by 17%.

Finally, we also studied the band gap influence motivated by the use of new materials for the semiconductor film. It was observed that higher band gaps increase the retention time by up to 5 orders of magnitude, allowing a retention time closer to the current conventional DRAMs.

Keywords: SOI, MOS, Microelectronic, Transistors, RAM memory, Memory.

LISTA DE FIGURAS

Figura 1.1: Estrutura da matriz de memória 1T-DRAM. [25].....	24
Figura 2.1: Estrutura de um transistor SOI nMOSFET planar.....	27
Figura 2.2: Diagrama de faixas de energia de um dispositivo parcialmente depletado (a) e um totalmente depletado (b). [12]	29
Figura 2.3: Corrente de dreno em função da tensão de porta para diferentes tensões de substrato. [12].....	32
Figura 2.4: Transcondutância em função da tensão de porta. [12].....	33
Figura 2.5: Regiões de depleção em um transistor MOSFET convencional de canal longo e curto (a e b) e em um SOI MOSFET de canal longo e curto (c e d). [12].....	38
Figura 2.6: Redução da barreira de potencial do canal induzida pela tensão no dreno.....	39
Figura 2.7: Estrutura de um transistor MuGFET (a) e UTBOX SOI nMOSFET de forma simplificada (b) e de forma real (c) [20].	40
Figura 2.8: Classificação das memórias de acordo com as operações possíveis.	42
Figura 2.9: Histórico das memórias 1T-DRAM. [20].....	42
Figura 2.10: Mecanismo de escrita '1' na memória 1T-DRAM por ionização por impacto (a) e curva $I_D \times V_G$ nos estados '0' e '1' (b) [25].....	44
Figura 2.11: Estrutura do transistor bipolar em um SOI nMOSFET (a), disparo do efeito BJT na curva $I_D \times V_G$ [34](b) e esquema da realimentação (c).....	45
Figura 2.12: Circuito para o equacionamento da corrente gerada pelo efeito BJT [12].....	45
Figura 2.13: Curva $I_D \times V_G$ para duas tensões altas de dreno e aparecimento da histerese. [36]	46
Figura 2.14: Efeito GIDL na curva $I_D \times V_G$ (a) e mecanismo de escrita '1' por meio deste efeito (b). [38].....	47
Figura 2.15: Tipos de tunelamento (a) [39] e mecanismo de escrita '1' por meio do GIFBE (b) [20].....	48
Figura 2.16: Mecanismo de escrita '0' por polarização direta. [20]	49
Figura 2.17: Mecanismo de escrita '0' por acoplamento capacitivo. [20]	49
Figura 2.18: Disparo do efeito BJT na curva $I_D \times \text{tempo}$ [29] [30].	50
Figura 2.19: Esquema de programação de uma célula 1T-DRAM. [37].....	51
Figura 2.20: I_1 , I_0 e ΔI_S em função da tensão na porta durante a leitura. [42]	52
Figura 2.21: Exemplo de extração do tempo de retenção. [37]	53

Figura 2.22: Estrutura com as extensões (a) e sem extensão (b) de fonte e dreno.....	54
Figura 3.1: Seqüência de estados utilizada.	56
Figura 3.2: Margem de sensibilidade em função da temperatura.	59
Figura 3.3: Tempo necessário para o disparo do efeito BJT em função da temperatura.	60
Figura 3.4: Tempo mínimo de disparo em função da tensão no dreno para várias temperaturas.....	61
Figura 3.5: Mínima tensão de disparo no dreno em função da espessura do silício para várias temperaturas.	61
Figura 3.6: Mínima tensão de disparo no dreno em função da espessura do óxido enterrado para várias temperaturas.	62
Figura 3.7: Mínima tensão de disparo no dreno em função da polarização do substrato.	63
Figura 3.8: Tempo de retenção em função da temperatura.	64
Figura 3.9: Esquema de programação utilizada. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.....	66
Figura 3.10: Janela da tensão na porta durante a leitura.....	66
Figura 3.11: Margem de sensibilidade em função da tensão na porta durante o tempo de repouso.	67
Figura 3.12: Taxa de geração a 1 nm abaixo da 1ª interface durante o estado de repouso.....	67
Figura 3.13: Corrente do estado '1' e '0' simulada ao longo do tempo para diferentes tensões na porta durante o repouso. A degradação do dado pode ser devido a recombinação (a) ou geração (b).	68
Figura 3.14: Corrente do estado '1' e '0' medida ao longo do tempo para diferentes tensões na porta durante o repouso. A degradação do dado pode ser devido a recombinação (a) ou geração (b).	68
Figura 3.15: Tempo de retenção em função da tensão na porta durante o repouso obtido por simulação (a) e por medidas (b).....	68
Figura 3.16: Comparação das taxas de recombinação (a) e de geração (b) durante o repouso no estado '1' a 1nm abaixo da 1ª interface para $V_{G_Repouso}$ de -1,5V e -2,3V.....	69
Figura 3.17: Janela de tensão do substrato.....	70
Figura 3.18: Janela de programação para várias tensões do substrato. [36]	70
Figura 3.19: Concentração de lacunas ao longo da estrutura durante o estado de repouso do dado '1' com $V_{G_Repouso} = -2.5$ V e $V_B = 2.1$ V.....	71
Figura 3.20: Corrente do estado '1' e '0' ao longo do tempo para diferentes tensões de substrato. As polarizações de porta e dreno são as apresentadas na Figura 3.9.....	72

Figura 3.21: Tempo de retenção em função da tensão no substrato.....	72
Figura 3.22: Esquema de programação utilizada. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.....	73
Figura 3.23: Margem de sensibilidade em função da tensão na porta para diferentes tensões de dreno durante a leitura (dados experimentais).	74
Figura 3.24: Melhor tensão na porta durante a leitura e a margem de sensibilidade em função da tensão no dreno durante a leitura.	74
Figura 3.25: Janela de leitura para diferentes tensões no dreno durante a leitura com dados experimentais (a) e simulados (b).	75
Figura 3.26: Margem de sensibilidade simulada em função da tensão no dreno durante a leitura.	75
Figura 3.27: Corrente nos estados '0' e '1' para alta e baixa polarização de dreno durante a leitura. Resultados experimentais (a) e simulados (b).	76
Figura 3.28: Tempo de retenção em função da tensão no dreno durante a leitura.	76
Figura 3.29: Múltiplas leituras (30x) para alta (a) e baixa (b) tensão de leitura do dreno.....	77
Figura 3.30: Esquema de programação na memória 1T-FBRAM.....	78
Figura 3.31: Janela de Leitura da tensão de porta para estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (a), $1 \times 10^{19} \text{ cm}^{-3}$ (b) e sem extensão; (c), e a margem de sensibilidade para diferentes concentrações da região de extensão (d).	80
Figura 3.32: Correntes no estado '1' e '0' ao longo do tempo para estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (quadrado), $1 \times 10^{19} \text{ cm}^{-3}$ (círculo) e com <i>underlap</i> (triângulo).	81
Figura 3.33: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (quadrado), $1 \times 10^{19} \text{ cm}^{-3}$ (círculo) e com <i>underlap</i> (triângulo).	81
Figura 4.1: Esquema de polarização utilizada.	83
Figura 4.2: Tempo de retenção para várias polarizações de substrato. Dados simulados (a) e experimentais (b).	84
Figura 4.3: Tempo de retenção em função do comprimento de canal considerando diferentes polarizações de substrato. Dados simulados (a) e experimentais (b).	85
Figura 4.4: Corrente devido ao GIDL em função da tensão de substrato para vários comprimentos de canal. Dados simulados (a) e experimentais (b).	86
Figura 4.5: Corrente devido ao GIDL em função do comprimento de canal a partir de dados simulados (quadrado aberto) e experimentais (esfera fechada). Dados experimentais encontrados em [45].	86
Figura 4.6: Taxa de geração ao longo do canal a 1nm abaixo da 1ª interface para várias tensões de substrato e para canal curto (a) e longo (b).	87

Figura 4.7: Campo elétrico a 1nm abaixo da 1ª interface para os dispositivos com extensão (a) e sem extensão (b).	87
Figura 4.8: Tempo de retenção em função do comprimento do canal para dispositivos com extensão (estrela fechada) e sem extensão (círculo aberta). Resultados simulados (a) e experimentais (b). O pequeno gráfico da figura (a) são as correntes lidas no estado '0' e '1' em função do tempo.....	88
Figura 4.9: Concentração de lacunas durante o repouso dos estados '0' e '1' em função do tempo.	89
Figura 4.10: Taxa de geração durante o repouso do estado '0' ao longo da estrutura sem extensão considerando o modelo de BBT (a) e sem considerá-lo (b).	90
Figura 4.11: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com (a) e sem extensão de fonte e dreno (b). ...	91
Figura 4.12: Taxa de geração durante a leitura do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com extensão (estrela fechada) e sem extensão (círculo aberto) de fonte e dreno.....	91
Figura 5.1: Sequência de estados e polarizações utilizadas. Apenas a polarização de substrato foi alterada para esta análise. A duração da leitura foi maior devido a necessidade de confirmação da estabilização da corrente.	93
Figura 5.2: Densidade de corrente de lacunas durante a escrita do estado '0' a 1 nm acima da 2ª interface utilizando V_B constante (quadrado tracejado) e pulsado (estrela contínua).	94
Figura 5.3: Concentração de lacunas durante o repouso dos estados '0' (aberto) e '1' (fechado) a 1 nm abaixo da 1ª interface utilizando V_B constante (quadrado tracejado) e pulsado (estrela contínua).....	94
Figura 5.4: Densidade de corrente de elétrons durante a leitura do '0' (aberto) e do '1' (fechado) a 1 nm acima da 2ª interface utilizando V_B constante (quadrado tracejado à esquerda) e pulsado (estrela contínua à direita).	95
Figura 5.5: Corrente de dreno em função do tempo obtida para V_B constante (tracejada) e pulsada (contínua).	95
Figura 5.6: I_1 e I_0 ao longo do tempo para V_B constante (quadrado tracejado) e pulsada (estrela contínua) e a extração de seu tempo de retenção.	96
Figura 5.7: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para V_B constante (quadrado tracejado) e pulsada (estrela contínua).	96
Figura 5.8: Diferentes bandas proibidas do filme semiconductor em um dispositivo <i>Semiconductor-on-Insulator</i> nMOSFET	97
Figura 5.9: Sequência de estados e polarizações utilizadas. O substrato foi polarizado em 2,1V e a uma temperatura de 85°C.....	97
Figura 5.10: I_1 e I_0 ao longo do tempo as diferentes bandas proibidas do semiconductor.	98
Figura 5.11: Tempo de retenção em função da banda proibida.	98

Figura 5.12: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as diferentes bandas proibidas utilizadas. 99

Figura 5.13: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface considerando ou não, ambos: o modelo de BBT e do TAT. 100

LISTA DE TABELAS

Tabela 3.1: Sequência de programação da memória.	57
Tabela 3.2: Estrutura do dispositivo gerada no ATLAS.	57
Tabela 3.3: Estrutura do dispositivo gerada no ATHENA.	65
Tabela 3.4: Estados e polarizações utilizadas. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.	65
Tabela 3.5: Doses de implantação e a concentrações obtidas para as diferentes regiões de extensão de fonte e dreno. O dispositivo de dose 0 refere-se àquele sem extensão.....	77
Tabela 3.6: Tensão no dreno durante a escrita.	79
Tabela 3.7: Tensão de leitura na porta e máxima margem de sensibilidade para diferentes concentrações da região de extensão.	80
Tabela 3.8: Tempo de retenção para as estruturas com diferentes concentrações da região de extensão.	82
Tabela 4.1: Estados e polarizações utilizadas.	84
Tabela 5.1: Estados e polarizações utilizadas.	93
Tabela 5.2: Margem de sensibilidade e tempo de retenção para V_B constante e V_B pulsado.	95
Tabela 6.1: Estados e polarizações otimizadas. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo. Temperatura a 85°C.	102

LISTA DE ABREVIATURAS E SIGLAS

1T1C-DRAM	<i>One Transistor One Capacitor DRAM</i> ou DRAM convencional
1T-DRAM ou 1T-FBRAM	<i>One Transistor DRAM</i> ou <i>One Transistor Floating Body RAM</i>
BBT	<i>Band to Band Tunneling</i> ou tunelamento banda a banda
BJT	<i>Bipolar Junction Transistor</i> ou <i>Transistor Bipolar de Junção parasitário</i>
CC	<i>Capacitive Coupling</i> ou acoplamento capacitivo
DIBL	<i>Drain Induced Barrier Lowering</i> ou redução da barreira de potencial do dreno
DRAM	<i>Dynamic Random Access Memory</i> ou memória dinâmica de acesso aleatório
EPROM	<i>Erasable Programmable Read Only Memory</i> ou memória programável e apagável somente de leitura
FBE	<i>Floating Body Effect</i> ou efeito de corpo flutuante
FBJ	<i>Forward Biased Junction</i> ou junção diretamente polarizada
FD SOI	<i>Fully Depleted SOI</i> ou SOI totalmente depletado
GIDL	<i>Gate Induced Drain Leakage</i> ou corrente de fuga induzida pela porta
GIFBE	<i>Gate Induced Floating Body Effect</i> ou efeito de corpo flutuante induzido pela porta
GP	<i>Ground Plane</i>
HDD	<i>Highly Doped Drain</i> ou região de fonte e dreno altamente dopada
II	<i>Impact ionization</i> ou ionização por Impacto
imec	<i>Interuniversity Microelectronics Center</i>
ITRS	<i>International Technology Roadmap for Semiconductor</i>
LDD	<i>Lightly Doped Drain</i> ou região de fonte e dreno menos dopada
MuGFET	<i>Multiple Gate Field Effect Transistor</i> ou Transistor de Efeito de Campo de Múltiplas Portas
PD SOI	<i>Partially Depleted SOI</i> ou SOI parcialmente depletado
PROM	<i>Programmable read only memory</i> ou memória programável somente de leitura
RAM	<i>Random access memory</i> ou memória de acesso aleatório
ROM	<i>Read only memory</i> ou memória somente de leitura

RDF	<i>Random-Dopant Fluctuations</i> ou flutuações aleatórias de dopantes
SCE	<i>Short channel effect</i> ou efeitos de canal curto
SOI MOSFET	<i>Silicon on Insulator Metal Oxide Semiconductor Field Effect Transistor</i> ou Transistor de Efeito de Campo Metal-Óxido-Semicondutor Silício sobre Isolante
SOS	<i>Silicon on Sapphire</i> ou silício sobre safira
SRAM	<i>Static random access memory</i> ou memória estática de acesso aleatório
TAT	<i>Trap Assisted Tunneling</i> ou Tunelamento assistido por armadilha
UTBOX	<i>Ultra Thin Buried Oxide</i> ou óxido enterrado ultrafino

LISTA DE SÍMBOLOS

A	Área da seção transversal em que se passa a corrente
C_b	Capacitância entre o canal de inversão e o substrato aterrado
C_{it1} e C_{it2}	Capacitâncias de armadilhas das interfaces óxido de prota/silício e silício/óxido enterrado
C_{ox} ou C_{ox1}	Capacitância do óxido de porta
C_{ox2}	Capacitância do óxido enterrado
C_{Si}	Capacitância do filme de silício
d	Máxima profundidade da camada de depleção da junção de fonte/dreno em dispositivos SOI
D_n	Coeficiente de difusão do elétron
E_c	Nível energético da faixa de condução
E_{Fm}	Nível de Fermi do metal
E_{Fsi}	Nível de Fermi do semiconductor
E_g	Banda proibida
E_i	Nível intrínseco do semiconductor
E_{vac}	Nível de energia do vácuo
E_{val}	Nível energético da faixa de valência
g_m	Transcondutância
$g_{m_máx}$	Transcondutância máxima
I_1 e I_0	Nível de corrente nos estados '1' e '0'.
I_C	Corrente de ionização por impacto amplificada pelo efeito BJT
I_{ch}	Corrente sem considerar a ionização por impacto
$I_{Coletor}$	Corrente de coletor
I_D ou I_{DS}	Corrente de fonte e dreno
I_{Dsat}	Corrente de fonte e dreno na região de saturação
I_{ES}	Corrente de saturação reversa da junção do emissor
$I_{fuga,pn}$	Corrente de fuga em uma junção PN
I_l	Corrente apenas devido à ionização por impacto
$I_{lacunas,ger}$	Corrente de lacunas devido à ionização por impacto
k	Constante de Boltzmann
L	Comprimento do canal
L_{ef}	Comprimento efetivo do canal

L_{LDD}	Comprimento da região de extensão de fonte/dreno
L_{porta}	Comprimento de máscara do canal
L_{min}	Comprimento mínimo do canal sem SCE
M	Fator multiplicativo referente à ionização por impacto
n	Fator de corpo
N_A	Concentração de dopantes no corpo
n_i	Concentração intrínseca dos portadores
N_{it}	Densidade de armadilhas de interface
q	Carga elementar do elétron
Q_{depl}	Carga total em depleção do filme de silício
Q_{dl}	Cargas de depleção controladas pela porta
Q_{inv1}	Densidade de cargas invertidas no canal próximo da 1ª interface
Q_{ox1} e Q_{ox2}	Densidade de cargas fixas na 1ª e na 2ª interfaces
Q_{s2}	Densidade de cargas no possível canal da 2ª interface
Q_{ss}	Concentração das cargas efetivas na interface
R_{AC}	Resistência da região de acumulação
R_{ch}	Resistência do canal
R_{conA} R_{conB} ou R_c	Resistência de contato
R_{dif}	Resistência da região de difusão
r_j	Profundidade de junção
R_{LDD}	Resistência da região de extensão de fonte/dreno
R_S e R_D	Resistência de fonte e de dreno
R_{SD}	Resistência série de fonte e dreno
R_{SH}	Resistência de folha
R_{SP1} e R_{SP2}	Resistências de espalhamento
R_T	Tempo de retenção
R_{total}	Resistência total entre fonte e dreno
S	Inclinação de sublimar
T	Temperatura absoluta
t_{BOX}	Espessura do óxido enterrado
$t_{min_escrita}$	Tempo mínimo necessário para ativar o efeito BJT
t_{OX}	Espessura do óxido de porta

t_{Si}	Espessura da região do canal
V_B	Tensão aplicada no substrato
V_{BE}	Tensão entre a base e o emissor do transistor bipolar parasitário
V_D ou V_{DS}	Tensão entre fonte e dreno
$V_{D_Leitura}$	Tensão no dreno durante o estado de leitura
$V_{Dmin_escrita}$	Mínima tensão no dreno necessária para ativar o efeito BJT
V_{Dsat}	Tensão no dreno de saturação
V_{FB}	<i>Flat Band Voltage</i> ou tensão de faixa plana
V_G	Tensão aplicada na porta
$V_{G_Leitura}$	Tensão na porta durante o estado de leitura
$V_{G_Repouso}$	Tensão na porta durante o estado de repouso
V_{GT}	Tensão de <i>overdrive</i> ($V_G - V_T$)
V_S	Tensão na fonte
V_T	Tensão de limiar
V_{T0}	Tensão de limiar para o substrato aterrado
W	Largura do canal
W_d	Largura da camada de depleção
x	Profundidade da camada de depleção
$x_{dmáx}$	Máxima profundidade da camada de depleção
α	Razão entre C_b e C_{ox1} (razão entre as capacitâncias acima e abaixo da camada de inversão)
α_F	Ganho da base em comum
β	Ganho do transistor bipolar parasitário
ΔI_S	Margem de sensibilidade
ΔL	Variação do comprimento do canal ($L_{porta} - L_{eff}$)
ϵ_{ox}	Permissividade do óxido
ϵ_{Si}	Permissividade do silício
μ_{ef}	Mobilidade efetiva dos portadores do canal
τ_e	Tempo de vida efetivo
τ_n	Tempo de vida do elétron
χ_{ox}	Afinidade eletrônica do óxido de silício
χ_{Si}	Afinidade eletrônica do silício
ϕ_F	Potencial de Fermi do semiconductor

Φ_{MS1} ou Φ_{MS2} Diferença de função trabalho entre o metal e o semiconductor da 1ª e na 2ª interfaces

SUMÁRIO

1 Introdução.....	22
1.1 Objetivo.....	24
1.2 Estrutura do trabalho.....	25
2 Revisão bibliográfica	27
2.1 SOI convencional.....	27
2.1.1 Classificação dos transistores SOI	28
2.1.2 Principais parâmetros.....	30
2.1.2.1 Tensão de limiar	30
2.1.2.2 Curvas $I_D \times V_G$	31
2.1.2.3 Transcondutância	32
2.1.2.4 Inclinação de sublimiar.....	34
2.1.2.5 Constante de efeito de corpo	35
2.1.3 Efeitos de canal curto	36
2.2 UTBOX	39
2.3 1T-DRAM.....	41
2.3.1 Programação e leitura do dado '1'	43
2.3.2 Programação e leitura do dado '0'	48
2.3.3 Principais parâmetros e seus métodos de extração.....	50
2.3.3.1 Ponto de disparo.....	50
2.3.3.2 Margem de sensibilidade de corrente e janela de leitura.....	51
2.3.3.3 Tempo de retenção	52
2.4 Dispositivos não auto-alinhados ("underlap")	53
3 Influência da concentração das regiões de extensão de fonte e dreno na 1T-DRAM	56
3.1 Estudo das polarizações	56
3.1.1 Escrita do bit '1' [29] [30]	57
3.1.1.1 Comportamento em altas temperaturas	58
3.1.1.2 Tempo de retenção	63
3.1.2 Leitura e repouso.....	64
3.1.2.1 Tensão na porta [42] [43]	65
3.1.2.2 Tensão no substrato [42] [43].....	70
3.1.2.3 Tensão no dreno [41].....	73
3.2 Impacto da concentração de extensão na 1T-DRAM [49]	77
3.2.1 Resultados e análises	78
3.2.1.1 Margem de sensibilidade e janela de leitura.	79
3.2.1.2 Tempo de retenção.....	80
4 Escalamento em dispositivos UTBOX sem extensões de fonte e dreno em termos de memória	83
4.1 Dispositivos auto-alinhados [47].....	83
4.2 Dispositivos não auto-alinhados [50].....	87
5 Melhorias no tempo de retenção.....	92
5.1 Utilização da polarização de substrato pulsada [51].....	92
5.1.1 Princípio de operação e margem de sensibilidade.....	93

5.1.2	Tempo de retenção	96
5.2	Influência da banda proibida [53]	97
6	Conclusões finais e trabalhos futuros.....	101
6.1	Conclusões finais	101
6.2	Trabalhos futuros	104
	Publicações	105
	Referências Bibliográficas	108
	Anexo 1 – Estrutura gerada no ATHENA.....	114
	Anexo 2 – Estrutura gerada no ATLAS.....	117
	Anexo 3 – Simulação de uma curva dinâmica	119

1 Introdução

Na tentativa de seguir com a tendência indicada pela Lei de Moore, novas técnicas vêm sendo estudadas para permitir o avanço do escalamento, incluindo técnicas para se diminuir o tamanho dos dispositivos, simplificação de circuitos, técnicas de projeto e de processo, entre outras [1]. No entanto, à medida que essas dimensões vão diminuindo, observa-se o aparecimento de alguns efeitos parasitários, como os denominados efeitos de canal curto e estreito [2]. A redução da tensão de limiar, o aumento da inclinação de sublimiar e a redução da barreira induzida pelo dreno são exemplos de efeitos de canal curto.

Estruturas com extensões de fonte e dreno menos dopadas e, mais recentemente, estruturas sem esta extensão (denominados *extension-less*) [3] [4] [5] têm sido estudadas para minimizar os efeitos de canal curto possibilitando um maior escalamento para aplicações de baixa tensão e baixa potência [6] [7] [8] [9] [10].

Na área de novos dispositivos SOI (*Silicon-on-Insulator*), há duas principais vertentes: a que se encaminha para as estruturas tridimensionais, cujo controle do canal é melhorado pela inserção de mais de uma porta, com os transistores de múltiplas portas (MuGFETs – *Multiple Gate Field Effect Transistors*) [2] [7] [11] [12]; e a que vai em direção à diminuição da camada de silício (canal) sobre o isolante, melhorando o controle do canal pela porta e sendo muito mais compatível com a tecnologia SOI planar, com os dispositivos de óxido enterrado ultrafino (UTBOX FD SOI MOSFETs – *Ultra-Thin-Buried-Oxide Fully-Depleted-Silicon-on-Insulator Metal-Oxide-Semiconductor-Field-Effect-Transistor*) [13] [14] [15] [16]. Neste trabalho, foram utilizadas estruturas planares com óxido enterrado ultrafino (UTBOX).

Outra área bastante estudada é a de memórias. Também tentando superar o limite de escalamento das memórias DRAM (*Dinamic Random Access Memory*) convencionais (1T1C-DRAM – *One Transistor One Capacitor DRAM*), composta de um transistor e de um capacitor, em 1978, propôs-se a memória de um único transistor (1T-DRAM – *Single Transistor Dinamic Random Access Memory*), onde se utiliza de efeitos, antes tidos como parasitários, para armazenar ou extrair cargas do canal do transistor. [17] [18] [19]

Já foram reportados quatro métodos para se armazenar cargas no transistor por ionização por impacto, pela corrente de fuga induzida pela porta, pelo efeito do transistor bipolar de junção parasitário e pelo efeito de corpo flutuante induzida pela porta. Para se expelir as cargas do corpo do transistor há dois métodos já estudados: pela polarização direta do dreno e pelo acoplamento capacitivo. [20]

Durante a operação de escrita do bit '1' na memória 1T-DRAM, cargas são armazenadas no corpo do transistor, resultando em um aumento do potencial de corpo fazendo com que a corrente de dreno se eleve, possibilitando a leitura do bit '1'. Para a escrita do bit '0', estas cargas são drenadas do corpo do transistor, reduzindo seu potencial e permitindo a leitura do bit '0' por meio de um nível de corrente menor. [21]

Devido ao fato do método de escrita pelo efeito BJT apresentar maior margem de sensibilidade, maior tempo de retenção e presença na maior parte dos dispositivos da tecnologia SOI [20] [22] [23], este método tem sido considerado como o mais promissor para ser competitivo com as especificações das memórias DRAM convencionais [24]. Portanto, este foi o método de armazenamento do bit '1' mais explorado neste trabalho e, portanto, na maioria das simulações aqui apresentadas.

Para a escrita do '0', optou-se pelo acoplamento capacitivo na maioria das análises realizadas pela sua melhor eficiência (menores tensões e correntes).

A escrita do bit '1' e do bit '0', a leitura dos mesmos e o repouso são os quatro principais estados das memórias e, na 1T-DRAM, elas ocorrem aplicando-se diferentes polarizações. Desta forma, é necessário levar em conta a interferência de cada polarização em cada estado, bem como seu comportamento na matriz de memórias. A Figura 1.1 apresenta a estrutura de uma matriz de memória 1T-DRAM 2x2. [25]

Figura 1.1: Estrutura da matriz de memória 1T-DRAM. [25]

Apesar da memória apresentar vantagens como custo reduzido, menor complexidade no processo de fabricação e ganho em área, alguns aspectos ainda precisam ser melhorados para que atinjam ou superem as especificações das atuais memórias DRAM [24]. O tempo de retenção é um exemplo, uma vez que ainda são 100 vezes menores que o das convencionais. [20]

1.1 Objetivo

O objetivo deste trabalho é propor melhorias de desempenho no funcionamento das memórias 1T-DRAM em dispositivos UTBOX FD SOI para que esta se torne viável frente às atuais memórias DRAM convencionais.

Entre as melhorias a serem avaliadas destaca-se:

- melhorias na estrutura: dispositivos com e sem extensão de fonte e dreno serão estudados como forma de se obter maior tempo de retenção, além de permitir maior escalamento;
- otimização das polarizações: diversas opções de polarização de porta e dreno, incluindo a polarização de substrato serão estudadas para obtenção e melhores tempos de retenção e maiores margens de sensibilidade de corrente.

- melhorias no material: estudo do uso de semicondutores de diferentes bandas proibidas para otimização do tempo de retenção.

1.2 Estrutura do trabalho

Este trabalho está dividido em seis capítulos, cuja descrição está apresentada abaixo:

Capítulo 1 - Introdução: Este capítulo descreve a importância das características gerais do estudo do presente trabalho. Também apresenta a definição do objetivo e a estrutura deste texto.

Capítulo 2 - Revisão bibliográfica: Neste capítulo encontra-se a revisão bibliográfica, abordando os principais conceitos da tecnologia SOI e seus principais parâmetros, da estrutura de óxido enterrado ultrafino (UTBOX) e suas vantagens, das memórias 1T-DRAM e seus principais parâmetros e da estrutura de fonte e dreno com os dispositivos não-autoalinhados.

Capítulo 3 - Influência da concentração das regiões de extensão de fonte e dreno na 1T-DRAM: Já este capítulo traz o estudo das polarizações nos diferentes parâmetros da memória, necessários para o entendimento dos mecanismos responsáveis no funcionamento da memória e, conseqüentemente, justificar a escolha das tensões utilizadas. Além disso, apresenta um estudo inicial da influência da concentração das regiões de fonte e dreno em alguns dos principais parâmetros da memória 1T-DRAM fazendo a ligação com os dispositivos não-autoalinhados.

Capítulo 4 - Escalamento em dispositivos UTBOX sem extensões de fonte e dreno em termos de memória: Como primeira proposta de melhoria de desempenho no funcionamento das memórias 1T-DRAM, este capítulo aborda a influência dos dispositivos sem as regiões de extensão de fonte e dreno no escalamento considerando sua utilização como memória 1T-DRAM, comparando com as estruturas convencionais, ou seja, com essas extensões.

Capítulo 5 - Melhorias no tempo de retenção: Este capítulo apresenta mais duas propostas de melhorias de desempenho no funcionamento das memórias 1T-DRAM: a utilização da polarização de substrato pulsada, como consequência do

estudo das polarizações, e a influência da largura da banda proibida, motivada pelo estudo de novos materiais para a camada semicondutora.

Capítulo 6 - Conclusões finais e trabalhos futuros: Finalmente, este capítulo apresenta as conclusões obtidas e sugestões de novos estudos para a continuidade do trabalho.

2 Revisão bibliográfica

Para a realização deste trabalho foi necessário estudar alguns conceitos de duas principais áreas: das novas estruturas de transistores e das memórias. Dessa forma este capítulo apresenta os aspectos mais importantes destes assuntos.

2.1 SOI convencional

O transistor com estrutura SOI (*Silicon on Insulator*) difere do convencional MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) pela presença de uma camada isolante de óxido de silício, o óxido enterrado, em seu substrato (Figura 2.1). Essa camada confere vantagens ao dispositivo como melhor isolamento elétrica entre os vários transistores existentes na mesma lâmina; melhor controle do canal e, conseqüentemente, melhor escalamento; menores capacitâncias de junção, chaveamento mais rápido, maior resistência à radiação, maior simplicidade de processo entre outras. [12]

Figura 2.1: Estrutura de um transistor SOI nMOSFET planar.

Na Figura 2.1, a espessura do óxido de porta é representada por t_{OX} , a do óxido enterrado por t_{BOX} , a da região do canal, também denominada corpo, de t_{Si} , o comprimento do canal de L e sua largura, W . Para as tensões de fonte e dreno temos, respectivamente, V_S e V_D ; para a porta, V_G e para o substrato, V_B . Denominou-se também como 1ª interface aquela entre o óxido de porta e a região do canal, como 2ª interface aquela entre a região do canal e o óxido enterrado e como 3ª interface aquela entre o óxido enterrado e o substrato.

2.1.1 Classificação dos transistores SOI

A tecnologia SOI também apresenta os mesmos regimes de operação de um transistor MOS convencional, porém, podendo haver interação entre as camadas de depleção provenientes das 1ª e 2ª interfaces. Assim, sua classificação é dependente da espessura da camada de silício e da largura da camada de depleção, como se segue: [12]

- Parcialmente depletado (PD SOI – *Partially Depleted* SOI): nesse caso a espessura do corpo é grande o suficiente para que a máxima profundidade da camada de depleção de ambas as interfaces não interajam entre si, havendo sempre uma região neutra ($t_{Si} > 2x_{dmáx}$). Seu diagrama de faixas de energia é apresentado na Figura 2.2-a. [12]
- Totalmente depletado (FD SOI – *Fully Depleted* SOI): aqui a espessura do corpo é menor que a máxima profundidade da camada de depleção apresentada por uma das interfaces ($t_{Si} < x_{dmáx}$), havendo assim uma interação entre as camadas. Seu diagrama de faixas de energia é apresentado na Figura 2.2-b. [12]

Sendo

$$x_{dmáx} = \sqrt{\frac{2 * \epsilon_{Si} * 2\phi_F}{q * N_A}} \quad (2.1)$$

e

$$\phi_F = \frac{kT}{q} * \ln\left(\frac{N_A}{n_i}\right) \quad (2.2)$$

onde:

- $x_{dm\acute{a}x}$: máxima profundidade da camada de depleção;
- ϵ_{Si} : permissividade do silício;
- q : carga elementar do elétron;
- N_A : concentração de dopantes no corpo;
- ϕ_F : potencial de Fermi do semiconductor;
- k : constante de Boltzmann;
- T : temperatura absoluta;
- n_i : concentração intrínseca dos portadores.

Figura 2.2: Diagrama de faixas de energia de um dispositivo parcialmente depletado (a) e um totalmente depletado (b). [12]

Na Figura 2.2, E_C é o nível de energia da banda de condução, E_V , da banda de valência e E_F é o nível de Fermi.

2.1.2 Principais parâmetros

2.1.2.1 Tensão de limiar

A tensão de limiar (V_T) representa a tensão aplicada na porta capaz de formar uma camada de inversão no corpo do silício do transistor. Nos parcialmente depletados, como não há interação entre as duas zonas de depleção, a tensão de limiar é dada pela mesma equação que nos transistores convencionais. Sua equação é dada por [12]:

$$V_T = V_{FB} + 2\phi_F + \frac{qN_A x_{dmax}}{C_{ox}} \quad (2.3)$$

em que C_{ox} é a capacitância do óxido de porta e é dada por: $C_{ox} = \epsilon_{ox}/t_{ox}$ com ϵ_{ox} sendo a permissividade do óxido; $V_{FB} = \phi_{MS} - Q_{ss}/C_{ox}$ com ϕ_{MS} sendo a função trabalho entre o metal e o semiconductor e Q_{ss} , a concentração de cargas efetivas na interface; e t_{ox} , q , N_A , x_{dmax} e ϕ_F como definido nas seções anteriores (2.1 e 2.1.1).

Já nos transistores totalmente depletados, a expressão da tensão de limiar é obtida segundo o modelo de Lim & Fossum [26].

Considerando os parâmetros com índices 1 aqueles referentes à 1ª interface (óxido de porta/silício) e com índice 2 aqueles referentes à 2ª interface (óxido enterrado/silício), a relação entre os potenciais de superfície (ϕ_s) e a tensão de porta (V_G) é dada por [12]:

$$V_G = \Phi_{MS1} - \frac{Q_{ox1}}{C_{ox1}} + \left(1 + \frac{C_{SI}}{C_{ox1}}\right) \phi_{S1} - \frac{C_{SI}}{C_{ox1}} \phi_{S2} - \frac{0,5Q_{depl} + Q_{inv1}}{C_{ox1}} \quad (2.4)$$

e a relação entre os potenciais de superfície e a tensão do substrato por [12]:

$$V_B = \Phi_{MS2} - \frac{Q_{ox2}}{C_{ox2}} - \frac{C_{SI}}{C_{ox2}} \phi_{S1} + \left(1 + \frac{C_{SI}}{C_{ox2}}\right) \phi_{S2} - \frac{0,5Q_{depl} + Q_{s2}}{C_{ox2}} \quad (2.5)$$

onde $C_{Si} = \epsilon_{Si}/t_{Si}$ e $Q_{depl} = -q \cdot N_A \cdot t_{Si}$ são, respectivamente, a capacitância e a carga total em depleção no filme de silício; Φ_{MS1} e Φ_{MS2} são as funções trabalho das referidas interfaces; Q_{ox1} e Q_{ox2} são as densidades de cargas fixas na 1ª e na 2ª interfaces; C_{ox1} e C_{ox2} são as capacitâncias de porta e do substrato; Q_{inv1} é a

densidade de cargas invertidas no canal próximo da 1ª interface e Q_{s2} é a densidade de cargas no possível canal da 2ª interface. As equações acima representam a interação existente entre as duas interfaces nos dispositivos totalmente depletados [12].

Chegando-se assim, na expressão da tensão de limiar (V_T) para um dispositivo totalmente depletado para as três condições da segunda interface [12]:

- Para a 2ª interface em acumulação, temos $\Phi_{s2}=0$, $Q_{inv1}=0$ e $\Phi_{s1}=2\phi_F$, e portanto:

$$V_{T1,acc2} = \phi_{MS1} - \frac{Q_{ox1}}{C_{ox1}} + \left(1 + \frac{C_{SI}}{C_{ox1}}\right) 2\phi_F - \frac{0,5Q_{depl}}{C_{ox1}} \quad (2.6)$$

- Para a 2ª interface invertida, $\Phi_{s2}=2\phi_F = \Phi_{s1}$ e $Q_{inv1}=0$, resultando:

$$V_{T1,inv2} = \phi_{MS1} - \frac{Q_{ox1}}{C_{ox1}} + 2\phi_F - \frac{0,5Q_{depl}}{C_{ox1}} \quad (2.7)$$

- Por fim, para a 2ª interface depletada, $Q_{inv1}=Q_{s2}=0$ e $\Phi_{s1}=2\phi_F$, logo:

$$V_{T1,depl2} = \phi_{MS1} - \frac{Q_{ox1}}{C_{ox1}} + \left(1 + \frac{C_{SI}}{C_{ox1}}\right) 2\phi_F - \frac{0,5Q_{depl}}{C_{ox1}} - \frac{C_{SI}C_{ox2}}{C_{ox1}(C_{SI} + C_{ox2})} (V_B - V_{B,acc}) \quad (2.8)$$

2.1.2.2 Curvas $I_D \times V_G$

Como pode haver interação entre as interfaces, a curva $I_D \times V_G$ (corrente de fonte e dreno em função da tensão aplicada na porta) de um SOI nMOSFET pode ser dividida em 4 regiões (Figura 2.3). [12]

Figura 2.3: Corrente de dreno em função da tensão de porta para diferentes tensões de substrato. [12]

A Figura 2.3 refere-se à corrente de dreno em função da tensão na porta para diferentes tensões no substrato. Observa-se na região A uma corrente da 2ª interface, uma vez que a 1ª interface está em acumulação e mesmo assim, há corrente passando; além disso, esta corrente aumenta quanto mais positivo for a tensão no substrato, indicando que a 2ª interface está invertida. Na região C, não há deslocamento da curva com a variação da tensão no substrato, demonstrando que a 2ª interface está em acumulação. Já na região B, a 2ª interface está depletada e a tensão de limiar depende linearmente da polarização do substrato. Em D, ambas estão depletadas também, porém a 2ª está próxima da inversão, fazendo com que qualquer aumento na tensão de porta leve à inversão da 2ª interface. [12]

2.1.2.3 Transcondutância

A transcondutância é a taxa de variação da corrente de dreno em relação à tensão de porta, ou seja, esse parâmetro determina o quanto a tensão de porta consegue controlar a corrente de dreno. Dessa forma, seu valor é obtido por [12]:

$$g_m = \frac{dI_D}{dV_G} \quad (2.9)$$

resultando na expressão:

$$g_m = \frac{\mu_{ef} C_{ox} W}{nL} (V_G - V_T), \text{ para } V_D > V_{Dsat} \text{ (região de saturação)} \quad (2.10)$$

e em:

$$gm = \frac{\mu_{ef} C_{ox} W}{nL} V_D, \text{ para } V_D < V_{Dsat} \text{ (região linear ou triodo)} \quad (2.11)$$

com μ_{ef} sendo a mobilidade efetiva no canal e $n=1+\alpha$, sendo $\alpha = C_b/C_{ox1}$ e C_b a capacitância existente entre o canal de inversão e o substrato aterrado. Podendo possuir os seguintes valores [12]:

- $C_b = C_D = \frac{\epsilon_{SI}}{x_{dmáx}}$: para dispositivos convencionais ou SOI parcialmente depletado, sendo C_D a capacitância da camada de depleção.
- $C_b = C_{SI}$: para um transistor totalmente depletado e com a 2ª interface acumulada
- $C_b = \frac{C_{SI}C_{ox2}}{C_{SI}+C_{ox2}}$: para um dispositivo totalmente depletado com a 2ª interface depletada

Um fenômeno que só ocorre nos transistores totalmente depletados é a possibilidade de formação de um “plateau” na curva da transcondutância em função da tensão de porta, dependendo da tensão no substrato (Figura 2.4). Nesse tipo de dispositivo, como o controle das cargas do canal é realizado tanto pela tensão da porta quanto pela tensão do substrato, a variação da corrente com a tensão da porta pode fazer a segunda interface inverter mais cedo que a primeira [12].

A dependência desse parâmetro (gm) com a tensão de porta (V_G) e a tensão aplicada no substrato (V_B) tem o seguinte comportamento (Figura 2.4) [12].

Figura 2.4: Transcondutância em função da tensão de porta. [12]

2.1.2.4 Inclinação de sublimiar

A inclinação de sublimiar (S) é definida como o inverso da derivada do logaritmo da corrente de dreno em relação à tensão de porta, na região de sublimiar [12]:

$$S = \frac{dV_G}{d(\log I_D)} \quad (2.12)$$

Ela representa o quão rápido o dispositivo vai do estado desligado para o ligado, ou seja, qual a variação de tensão na porta é necessária para que a corrente passe da região de corte para a condução. Quanto menor seu valor, mais rápido é essa transição e sua unidade é dada em mV/década.

Em SOI totalmente depletado, a expressão completa da inclinação de sublimiar é:

$$S = \frac{dV_G}{d \log(I_D)} = \frac{kT * \ln 10}{q} \left[\left(1 + \frac{C_{it1}}{C_{ox1}} + \frac{C_{Si}}{C_{ox1}} \right) - \frac{\frac{C_{Si}}{C_{ox2}} \frac{C_{Si}}{C_{ox1}}}{1 + \frac{C_{it2}}{C_{ox2}} + \frac{C_{Si}}{C_{ox2}}} \right] \quad (2.13)$$

onde C_{it1} e C_{it2} são as capacitâncias de armadilhas das interfaces óxido de porta/Si e Si/óxido enterrado, onde $C_{it}=q*N_{it}$ e N_{it} é a densidade de armadilhas de interface.

Desprezando as capacitâncias de armadilhas de interface, tem-se [12]:

$$S = \frac{kT * \ln(10)}{q} (1 + \alpha) = n \frac{kT}{q} \ln(10) \quad (2.14)$$

Em que, utilizando a mesma definição de n e α , explicitada na seção 2.1.2.3, resulta para um dispositivo convencional ou SOI parcialmente depletado, desconsiderando a capacitância de armadilhas de interface:

$$S = \frac{kT * \ln(10)}{q} \left(1 + \frac{C_D}{C_{ox}} \right) \quad (2.15)$$

e para o SOI totalmente depletado:

$$S = \frac{kT * \ln(10)}{q} \left[1 + \frac{1}{C_{ox2}} \left(\frac{C_{Si} C_{ox2}}{C_{Si} + C_{ox2}} \right) \right] \quad (2.16)$$

Pelas equações (2.15) e (2.16) acima, a inclinação de sublimiar de dispositivos convencionais ou parcialmente depletados é maior que a dos totalmente depletados, o que indica um melhor valor deste parâmetro para os transistores totalmente depletados. Isso permite que se usem menores valores da tensão de limiar nele do que nos convencionais, sem um aumento da corrente de fuga para $V_G=0V$. Dessa forma, o aumento da inclinação de sublimiar pela diminuição do comprimento do canal é menor nos FD SOI [12].

2.1.2.5 Constante de efeito de corpo

Esse parâmetro representa o quanto a tensão de limiar é dependente da tensão do substrato. Em transistores convencionais, ela é obtida a partir das equações (2.1), e (2.3), resultando: [12]

$$V_T = \Phi_{MS} - \frac{Q_{SS}}{C_{ox}} + 2\phi_F + \frac{\sqrt{2\varepsilon_{SI}qN_A(2\phi_F - V_B)}}{C_{ox}} \quad (2.17)$$

Considerando:

$$\gamma = \frac{\sqrt{2\varepsilon_{SI}qN_A}}{C_{ox}} \quad (2.18)$$

Pode-se chegar a:

$$V_T = \Phi_{MS} - \frac{Q_{SS}}{C_{ox}} + 2\phi_F + \gamma\sqrt{2\phi_F} - \gamma\sqrt{2\phi_F} + \gamma\sqrt{(2\phi_F - V_B)}$$

$$V_T = V_{T0} + \gamma\left(\sqrt{(2\phi_F - V_B)} - \sqrt{2\phi_F}\right) \quad (2.19)$$

Já nos transistores SOI parcialmente depletados com o corpo aterrado, como não há acoplamento entre as interfaces, a tensão de limiar não varia com a tensão no substrato. Porém, se o corpo estiver flutuando e a tensão aplicada no dreno for alta, vários efeitos podem ocorrer devido ao alto campo elétrico formado na região próxima à junção canal/dreno como pode ser encontrado mais adiante (seção 2.3.1-a, b e c).

Em transistores FD SOI ela é obtida a partir da derivada da tensão de limiar com relação à tensão no substrato. Logo, derivando a (2.6), (2.7) e (2.8), tem-se: [12]

$$\frac{dV_{T1,acc2}}{dV_B} = 0, \quad (2.20)$$

$$\frac{dV_{T1,inv2}}{dV_B} = 0 \quad (2.21)$$

e

$$\frac{dV_{T1,depl2}}{dV_B} = -\frac{C_{Si}C_{ox2}}{C_{ox1}(C_{Si} + C_{ox2})} = \gamma \quad (2.22)$$

Ou seja, quando a 2ª interface estiver acumulada ou invertida, a tensão de limiar na 1ª interface é constante, já quando a 2ª interface estiver em depleção, ela diminui linearmente com o aumento da tensão no substrato.

2.1.3 Efeitos de canal curto

À medida que as dimensões vão diminuindo para se obter uma maior integração dos dispositivos, observa-se o aparecimento de alguns efeitos parasitários decorrentes da diminuição do comprimento do canal. Estes efeitos ocorrem devido à perda do controle das cargas de depleção pela tensão de porta, uma vez que há um compartilhamento de cargas do canal com as regiões de depleção das junções fonte/canal e dreno/canal. Formas diretas de se observar este efeito nas curvas $I_D \times V_G$ é a redução da tensão de limiar e o aumento da inclinação de sublimiar.

A Figura 2.5 apresenta um esquema das regiões de depleção em um transistor MOSFET convencional de canal longo e curto e em um SOI MOSFET de canal longo e curto.

Para os transistores MOS, a Figura 2.5 mostra que o perfil de cargas foi de um formato trapezoidal, no dispositivo de canal longo para um triangular, no de canal curto. Neste transistor, as cargas de depleção controladas pela porta (Q_{dl})

podem ser modeladas como sendo uma fração das cargas totais (Q_{depl}), dessa forma: [12]

$$Q_{\text{dl}} = Q_{\text{depl}} \left(1 - \frac{r_j}{L} \left(\sqrt{1 + \frac{2x_{\text{dmax}}}{r_j}} - 1 \right) \right) \quad (2.23)$$

com

$$Q_{\text{depl}} = qN_A x_{\text{dmax}} \quad (2.24)$$

onde r_j é a profundidade de junção.

Já para os transistores da tecnologia SOI, a mesma Figura 2.5 mostra que, devido ao isolamento provido pelo óxido enterrado, a profundidade da região de depleção e a quantidade de cargas de depleção controladas pelas junções de fonte e dreno são menores, fazendo com que a fração das cargas controladas pela porta seja maior que aquela observada para a mesma diminuição do comprimento do canal apresentada pelo MOS convencional, permitindo um escalamento maior. Neste caso, as cargas de depleção controladas pela porta podem ser modeladas segundo a equação:

$$Q_{\text{dl}} = Q_{\text{depl}} \left(1 - \frac{d}{L} \right) \quad (2.25)$$

com

$$Q_{\text{depl}} = qN_A t_{\text{Si}} \quad (2.26)$$

e d é a distância definida na Figura 2.5.

Figura 2.5: Regiões de depleção em um transistor MOSFET convencional de canal longo e curto (a e b) e em um SOI MOSFET de canal longo e curto (c e d). [12]

Observa-se, dessa forma, que a tensão no dreno, responsável pelo controle das regiões de depleção da junção dreno/canal, também influencia nesse comportamento. Para um alto potencial de fonte e dreno aplicado, mais cargas da junção passarão a ser controladas também pela tensão do dreno, podendo fazer com que a região controlada pelo dreno alcance a região controlada pela fonte, reduzindo a barreira de potencial no canal (Figura 2.6), dessa forma, quando o canal é curto, menor será a tensão necessária para ocorrer este efeito, conhecido como DIBL (*Drain-Induced Barrier Lowering* ou redução da barreira de potencial induzida pelo dreno).

Nos dispositivos SOI totalmente depletados esse efeito aparece apenas para comprimentos de canal menores e tensão de dreno mais elevados. Como toda a região do corpo está depletada ($t_{Si} < x_{d\text{máx}}$), o comprimento do canal em que começa a ocorrer a diminuição da camada de depleção é menor. Nesse mesmo tipo de transistor, a presença do óxido enterrado faz com que a região das cargas controladas pela tensão de fonte e dreno se torne menor quando comparado com as cargas totais (Figura 2.5), diminuindo o efeito DIBL [12].

Figura 2.6: Redução da barreira de potencial do canal induzida pela tensão no dreno.

2.2 UTBOX

Uma das linhas de pesquisa para obtenção de transistores avançados é a que se encaminha para as estruturas tridimensionais, que é o caso dos MuGFET's (*Multiple Gates Field Effect Transistor*) [11] [12] ilustrado na Figura 2.7-a, e outra é a utilização ainda da tecnologia planar com o transistor SOI MOSFET totalmente depletado (Figura 2.1), porém com uma espessura de óxido enterrado muito fina, estrutura esta denominada UTBOX (*Ultra Thin Buried Oxide*) (Figura 2.7-b e Figura 2.7-c) [27]. Essa estrutura possui vantagens como maior acoplamento da porta e do substrato no canal e, conseqüentemente, maior influência da tensão aplicada nesse contato o que resulta em um melhor controle na indução do efeito de corpo flutuante, necessários para a aplicação nas memórias 1T-DRAM e menor efeitos de canal curto. Além disso, esta estrutura apresenta também um melhor controle da tensão de limiar e menor resistência térmica. [13] [15] [16] [28] [29] [30] [20]

Figura 2.7: Estrutura de um transistor MuGFET (a) e UTBOX SOI nMOSFET de forma simplificada (b) e de forma real (c) [20].

Entretanto, a 3ª interface (óxido enterrado/substrato) também apresenta o comportamento descrito para as 1ª e 2ª interfaces, fazendo com que a espessura efetiva do isolante enterrado seja maior no caso desta interface estar na região de depleção. Dessa forma, a região logo abaixo do óxido enterrado é dopada de forma a minimizar a máxima espessura desta camada de depleção. Esta região mais dopada é denominada GP (*Ground Plane* ou Plano de terra) [15] [16]. Além disso, o óxido enterrado fino junto com o GP (*Ground Plane*) permite o controle da tensão de limiar sem a necessidade de se aumentar a concentração da região do canal, o que pode resultar no chamado RDF (*Random-Dopant Fluctuations* ou flutuação aleatória de dopantes) e na degradação da mobilidade. [28]

De acordo com a patente do processo de obtenção do óxido enterrado fino [31], uma estrutura é considerada UTBOX quando ela apresenta uma espessura menor que 50 nm. Porém, para que a atuação da polarização do substrato seja

efetiva, sua dimensão deve ser igual ou menor que 10nm [32], uma vez que ela permite um melhor acoplamento do canal pela tensão do substrato.

O processo de obtenção do óxido enterrado fino é descrito como uma dissolução do óxido, ou seja, inicia-se com uma lâmina SOI de óxido espesso e o óxido enterrado é afinado por meio de uma decomposição em um ambiente com argônio e um recozimento em oxigênio. [31]

2.3 1T-DRAM

Uma memória pode ser classificada pelas operações que se é possível realizar nela, ou seja, se ela é de leitura e escrita (RAM – *random access memory*) ou apenas de leitura (ROM – *read only memory*). A primeira caracteriza-se pela comparável velocidade com que os dados são armazenados e recuperados, uma vez que independe da localização física da célula dentro da memória; já a segunda permite apenas a leitura com a mesma alta velocidade. [33]

Além disso, as memórias podem ser voláteis ou não-voláteis, dependendo da necessidade de uma alimentação para que o dado seja armazenado, sendo essa necessidade presente nas memórias voláteis. [33]

Dentre as memórias RAM, há basicamente dois tipos: as RAM estáticas (SRAM – *Static RAM*) e as RAM dinâmicas (*dynamic RAM*), dependendo da necessidade ou não de uma atualização (*refresh*) periódica dos dados. As memórias DRAM são as que precisam desta atualização, uma vez que o princípio de funcionamento dessa memória consiste no armazenamento de cargas em um capacitor. Isso ocorre devido, principalmente, às correntes de fuga presentes no capacitor. [33]

Neste trabalho, a memória em estudo permite tanto a escrita como a leitura, é do tipo volátil e dinâmica. A Figura 2.8 resume essas classificações contextualizando a 1T-DRAM.

Figura 2.8: Classificação das memórias de acordo com as operações possíveis.

O primeiro estudo da memória 1T-DRAM data de 1978 [17], fabricada sobre um substrato de silício sobre safira (SOS – *Silicon-On-Sapphire*), já em 1990, o imec iniciou os estudos dessa memória em um substrato SOI e, desde 2001 suas pesquisas só vem se intensificando como mostra a Figura 2.9 e as referências [20], que aborda sobre as SRAMs, as DRAMs juntamente com as 1T-DRAMs, além de novos tipos de memória como a MSDRAM, a A-RAM e a célula 2T floating-body/gate e as [34] e [35], que estudam esse tipo de memória em MuGFETs.

Figura 2.9: Histórico das memórias 1T-DRAM. [20]

A memória DRAM convencional é formada por um transistor e um capacitor armazenador de informação que requer muito espaço na lâmina, seja ela em área ou em profundidade, sendo um fator limitante no escalamento do dispositivo e apresentando um processo mais complexo. A memória 1T-DRAM vem sendo estudada como uma alternativa para avançar na evolução do escalamento das memórias, uma vez que ela é composta por apenas um transistor, levando a uma maior capacidade da memória. Além disso, apresenta vantagens como a compatibilidade com a tecnologia SOI, maior simplicidade quando comparado com o processo da célula da memória DRAM convencional já existente e menor tempo de fabricação por célula, resultando em um custo menor. [18] [19]

Este tipo de memória se utiliza de efeitos da própria estrutura SOI, antes tidos como parasitários, para armazenar cargas no corpo do dispositivo, ou seja, escrever na memória. Há 4 métodos de escrita do bit '1' já reportados: pela ionização por impacto (II - *Impact Ionization*), pelo efeito do transistor bipolar parasitário (BJT - *bipolar junction transistor*), pelo efeito da corrente de fuga induzida pela porta (GIDL - *Gate Induced Drain Leakage*) e pelo efeito de corpo flutuante induzido pela porta (GIFBE - *Gate Induced Floating Body Effect*). [20] [21]

Para a escrita do bit '0', ou seja, para se retirar as cargas do corpo, já foram apresentados 2 métodos: por polarização direta (FBJ – *forward bias junction*) e por acoplamento capacitivo (CC – *capacitive coupling*). A seguir, encontram-se as descrições de cada mecanismo. [20] [21]

2.3.1 Programação e leitura do dado '1'

A programação do dado '1' nada mais é do que o armazenamento de cargas, lacunas no caso de um dispositivo nMOS, no corpo do transistor, levando ao chamado efeito de corpo flutuante (FBE - *Floating Body Effect*), motivo de outro nome para esta memória, 1T-FBRAM (*Single Transistor Floating Body RAM*). Estas cargas são originadas por vários modos, dentre elas, estão:

a) Ionização por impacto (II)

Quando um transistor está polarizado na região de saturação um alto campo elétrico é formado na região do canal próxima ao dreno, os

elétrons se aceleram e chocam-se com a rede cristalina, gerando pares elétrons-lacunas. Os elétrons vão para a região de maior potencial, ou seja, segue seu caminho em direção ao dreno, e as lacunas vão em direção à região de menor potencial, isto é, o corpo. Esta maior concentração de lacunas no corpo eleva seu potencial, diminuindo a tensão de limiar e resultando em uma corrente de dreno maior (Figura 2.10-b). [20] [21] [25]

Figura 2.10: Mecanismo de escrita '1' na memória 1T-DRAM por ionização por impacto (a) e curva $I_D \times V_G$ nos estados '0' e '1' (b) [25].

b) Efeito do transistor bipolar parasitário (BJT)

Este método é uma evolução do anterior [18], uma vez que se inicia a partir da elevação da concentração de lacunas originadas na ionização por impacto. Para baixas tensões no dreno, a curva $I_D \times V_G$ possui seu comportamento usual. Já para altas tensões no dreno, com o aumento da tensão de porta, como o campo elétrico junto ao dreno é alto, a corrente na inversão fraca já é suficiente para iniciar o processo de ionização por impacto. Se o tempo de vida dos portadores minoritários for alto o suficiente, o transistor bipolar parasitário (formado por fonte (emissor), canal (base) e dreno (coletor), Figura 2.11-a) amplifica a corrente de lacunas geradas pela ionização por impacto, resultando em uma realimentação positiva na corrente de dreno do transistor principal podendo levar a um aumento abrupto na curva $I_D \times V_G$ (Figura 2.11-b). Consequentemente, uma inclinação de sublimiar menor que o limite teórico de 60mV/década é atingida. Este processo pode ser esquematizado como mostra a Figura 2.11-c. [29] [30] [20] [21]

Figura 2.11: Estrutura do transistor bipolar em um SOI nMOSFET (a), disparo do efeito BJT na curva $I_D \times V_G$ [34](b) e esquema da realimentação (c).

Da Figura 2.11-a, chega-se ao circuito apresentado na Figura 2.12: [12]

Figura 2.12: Circuito para o equacionamento da corrente gerada pelo efeito BJT [12].

onde se tem: [12] [22]

$$I_{DS} = M(I_{ch} + I_C) \quad (2.27)$$

$$I_C = \beta I_i \quad (2.28)$$

e [12] [22]

$$I_i = (M - 1)(I_{ch} + I_C) \quad (2.29)$$

Logo, com as equações (2.28) e (2.29) na (2.27), a corrente de dreno considerando o efeito BJT é: [12]

$$I_{DS} = M(I_{ch} + \beta I_i) = MI_{ch} + M\beta(M - 1)(I_{ch} + I_C)$$

Substituindo a (2.27) na equação acima fica:

$$I_{DS} = MI_{ch} + I_{DS}\beta(M - 1) \rightarrow I_{DS} - I_{DS}\beta(M - 1) = MI_{ch}$$

$$I_{DS} = \frac{MI_{ch}}{1 - \beta(M - 1)} \quad (2.30)$$

Nas equações acima, I_{ch} é a corrente sem considerar a ionização por impacto, I_C é a corrente de ionização por impacto amplificada pelo efeito BJT, I_i é a corrente apenas devido à ionização por impacto, I_{DS} é a corrente total no dreno, β é o ganho do transistor bipolar parasitário e M é o fator multiplicativo referente à ionização por impacto. Nela, quando o termo $\beta(M-1)$ se aproxima de 1, há um aumento súbito da corrente de dreno e uma inclinação de sublimar muito próxima de 0 é alcançada.

Como a tensão de limiar diminuiu com o aumento do potencial de corpo, a inativação do efeito ocorrerá para uma tensão de porta menor, apresentando uma curva $I_D \times V_G$ como a mostrada na Figura 2.13, onde se vê a formação de uma histerese.

Figura 2.13: Curva $I_D \times V_G$ para duas tensões altas de dreno e aparecimento da histerese. [36]

Como se observa na Figura 2.13, há um intervalo de tensões de porta em que a corrente de dreno pode apresentar dois valores distintos, dependendo de seu estado anterior, neste caso, referente à quantidade de lacunas no corpo do dispositivo. Define-se assim, o nível de corrente no estado '1' e no estado '0'.

c) Efeito da corrente de fuga induzida pela porta (GIDL)

Quando se tem uma tensão positiva no dreno e negativa na porta, forma-se uma depleção na região sobreposta da porta sobre o dreno o que leva ao tunelamento de elétrons entre as bandas de valência e condução (BBT) do próprio silício gerando pares elétrons-lacunas. Os elétrons, que vão em direção à região de maior potencial, vão para o dreno, enquanto que as lacunas vão em direção à região de menor potencial, isto é, o corpo (Figura 2.14-a). Este efeito é visto na curva $I_D \times V_G$ como mostra a Figura 2.14-b. [20] [21] [25] [37]

Figura 2.14: Efeito GIDL na curva $I_D \times V_G$ (a) e mecanismo de escrita '1' por meio deste efeito (b). [38]

O aumento da concentração de lacunas, novamente, leva a um aumento da corrente de dreno pela diminuição da tensão de limiar, havendo a identificação do dado semelhante ao modo do armazenamento do bit pela ionização por impacto (primeira geração). [37]

d) Efeito de corpo flutuante induzido pela porta (GIFBE)

Neste caso, as lacunas no corpo são provenientes do tunelamento pelo óxido de porta. Quando o óxido é posto sob efeito de um alto campo elétrico, os portadores ganham energia suficiente para transpor a barreira do óxido, que pode apresentar um formato triangular (tunelamento por Fowler Nordheim) ou um formato trapezoidal (tunelamento direto) (Figura 2.15-a) [39]. No caso das memórias, pode ocorrer de uma lacuna tunelar da porta (no caso de silício policristalino) para o canal, ou de um elétron tunelar do canal para a porta, deixando uma lacuna no canal, elevando a concentração de lacunas no corpo e, novamente, ocorrendo o efeito de corpo flutuante, como mostra a Figura 2.15-b na estrutura e em modelo de bandas de energia. [20]

Figura 2.15: Tipos de tunelamento (a) [39] e mecanismo de escrita '1' por meio do GIFBE (b) [20].

Dessa forma, o dado '1' é identificado como um nível de corrente mais alto que uma corrente de referência previamente projetada.

2.3.2 Programação e leitura do dado '0'

Analogamente, a programação do bit '0' é a retirada das cargas do corpo do transistor. Até o momento foram reportados dois mecanismos para isso:

a) Polarização direta (FBJ)

Neste método aplica-se uma tensão negativa no dreno e, por polarização direta da junção corpo-dreno, as lacunas são drenadas por ela (Figura 2.16), diminuindo o potencial no corpo e resultando numa corrente mais baixa no estado de leitura, como mostra a figura Figura 2.16. [20] [21]

Figura 2.16: Mecanismo de escrita '0' por polarização direta. [20]

b) Acoplamento capacitivo (CC)

Quando um pulso positivo é aplicado na porta, o potencial no corpo é induzido a um valor mais alto, diminuindo a barreira da junção corpo-fonte e corpo-dreno e fazendo as lacunas fluírem até que o potencial do corpo se aproxime do potencial da fonte. Então, com essa diminuição da concentração de lacunas, o potencial no corpo diminui, resultando numa corrente mais baixa no estado de leitura (Figura 2.17). [20] [21]

Figura 2.17: Mecanismo de escrita '0' por acoplamento capacitivo. [20]

Logo, o dado é identificado como um nível de corrente mais baixo que a corrente de referência.

2.3.3 Principais parâmetros e seus métodos de extração

Em todo este trabalho, foram realizadas simulações numéricas bidimensionais [40], de onde se extraiu os principais parâmetros elétricos desse novo tipo de memória, como a janela de programação, a margem de sensibilidade e o tempo de retenção. Esta seção explica os principais parâmetros estudados neste trabalho e como eles podem ser obtidos.

2.3.3.1 Ponto de disparo

No caso do método de escrita utilizando o efeito BJT, definiu-se a polarização de dreno mínima ($V_{Dmin_escrita}$) e o tempo mínimo ($t_{min_escrita}$) para que o efeito ocorra. Uma vez que altas polarizações no dreno resultam em danos ao dispositivo devido à ionização por impacto e outros efeitos decorrentes do alto campo elétrico, a polarização no dreno durante a escrita se torna um fator limitante para um bom funcionamento da memória 1T-FBRAM [20] [41]. Por outro lado, uma baixa tensão no dreno não é suficiente para ativar o efeito BJT [34] [36]. A Figura 2.11-b apresenta um meio de se extrair o $V_{Dmin_escrita}$, neste caso, seu valor seria de 3V; já a Figura 2.18 mostra a extração do $t_{min_escrita}$, neste caso, seu valor seria de aproximadamente 23ns.

Figura 2.18: Disparo do efeito BJT na curva $I_D \times tempo$ [29] [30].

2.3.3.2 Margem de sensibilidade de corrente e janela de leitura

A margem de sensibilidade (ΔI_S) é definida como a diferença entre a corrente lida após a escrita do '1' (I_1) e a lida após a escrita do '0' (I_0). A Figura 2.19 mostra graficamente a extração deste parâmetro, no caso, o seu valor é de aproximadamente $45\mu A/\mu m$. Estas curvas foram obtidas por meio de uma simulação bidimensional realizada no ATLAS [40].

Figura 2.19: Esquema de programação de uma célula 1T-DRAM. [37]

Deseja-se que seu valor seja o maior possível pois, quanto menor for este parâmetro mais preciso o circuito de leitura deverá ser e mais susceptível a ruídos a memória será. Além disso, maior margem de sensibilidade também permite um maior escalamento [27].

A Figura 2.20 mostra a corrente lida no estado '1', no estado '0' e sua margem de sensibilidade em função da tensão aplicada na porta durante a leitura ($V_{G_Leitura}$), sendo indicado também a sua janela de leitura. A janela de leitura é o intervalo de tensões em que o efeito é identificado pela memória, ou seja, há dois níveis de corrente dependendo do estado anterior, no caso a quantidade de lacunas no corpo, e é definida pelo intervalo de tensão em que a margem de sensibilidade é acima de certo valor. Busca-se aumentar este intervalo para possibilitar maior variabilidade da tensão sem que isso afete o funcionamento da memória [42] [43].

A obtenção da janela de leitura se dá a partir de sucessivas simulações dinâmicas similares às apresentadas na Figura 2.19, variando-se $V_{G_Leitura}$ e extraindo-se os valores de I_0 , I_1 e ΔI_S . Dessa forma, cada ponto do gráfico

referente a um $V_{G_Leitura}$, corresponde a uma simulação. Um exemplo dos dados de entrada necessários para se obter uma simulação dinâmica encontra-se no Anexo 3.

Figura 2.20: I_1 , I_0 e ΔI_S em função da tensão na porta durante a leitura. [42]

2.3.3.3 Tempo de retenção

Quando não se está escrevendo (armazenando ou expelindo cargas do corpo), nem lendo, diz-se que a memória está em seu estado de repouso (ou manutenção). Sua importância deve-se ao fato dele ser o estado em que se deseja manter as cargas no corpo o máximo de tempo possível de forma que o dado ainda seja lido como aquele que se deseja. Este é o chamado tempo de retenção (RT) e é definida como o máximo tempo entre uma escrita e sua leitura (estado de repouso) no qual o dado ainda pode ser lido corretamente, ou seja, representa quanto tempo o dado fica retido na memória. Aumentando este parâmetro, diminui-se o tempo de atualização que a memória DRAM necessita (tempo de *refresh*).

A Figura 2.19 mostra os dois estados de repouso relacionados com o tempo de retenção: o intervalo entre a escrita e a leitura do bit '1' ($t_{\text{Repouso},1}$) e do bit '0' ($t_{\text{Repouso},0}$). A partir de várias simulações resultando em curvas similares a esta, porém variando-se esse tempo, extrai-se I_1 e I_0 , levantando assim, o gráfico da Figura 2.21. Esta Figura 2.21 também apresenta um exemplo da extração do tempo de retenção, obtido pelo tempo no qual a corrente I_1 é degradada, se aproximando de I_0 , neste caso foi considerada como nível de referência,

$50\% \cdot \Delta I_{S_25^\circ\text{C}}$. Mais adiante será observado que, dependendo da polarização durante o repouso, pode haver a degradação do bit '0' ao invés do bit '1', em outras palavras, será o I_0 que se aproximará do I_1 .

Figura 2.21: Exemplo de extração do tempo de retenção. [37]

Seu estudo foi necessário pois, além de estar diretamente ligado à degradação do dado, em um dispositivo real, sabe-se que todos os efeitos já mencionados podem afetar o dispositivo neste estado dependendo de sua polarização [29] [30] [37] [41] [42] [43].

Este parâmetro ainda está sendo estudado para que as memórias 1T-DRAM alcancem os tempos de retenção das memórias DRAM convencionais existentes atualmente. Segundo as especificações da *International Technology Roadmap for Semiconductor* de 2011 [24], este valor é de 64ms.

2.4 Dispositivos não auto-alinhados (“underlap”)

Na fabricação de um circuito integrado, a fonte e o dreno de um transistor são obtidos pela dopagem de certa região da lâmina de silício definida como tal. Nesse processo, também ocorre uma difusão lateral, ou seja, o espalhamento dos dopantes para a região sob o óxido de porta, originando o termo em inglês “overlap”.

Na tentativa de seguir com a tendência proposta pela Lei de Moore, estruturas com regiões de extensão de fonte e dreno menos dopadas (Figura 2.22-a) [6] [7] [8] [9] [14] e, mais recentemente, estruturas sem a dopagem dessa extensão (denominadas *extension-less* ou *underlapped* ou com *underlap*) (Figura 2.22-b), tem sido estudadas para minimizar os efeitos de canal curto [3] [4] [5] [44].

Figura 2.22: Estrutura com as extensões (a) e sem extensão (b) de fonte e dreno.

Com essa região menos dopada, também denominada LDD (*Lightly Doped Drain*), a difusão lateral e o potencial efetivamente aplicado na região do canal diminuem e, conseqüentemente, o campo elétrico e a quantidade de cargas cujo controle é compartilhado entre a tensão de porta e a de dreno também segue a mesma tendência. Dessa forma, os efeitos de canal curto são minimizados. [6] [7] [8] [9]

Por outro lado, a partir do mesmo processo de obtenção dessas regiões menos dopadas por meio de espaçadores (*spacer*), a não implantação dessa região faz com que a difusão lateral possa não atingir a região do canal, resultando em um dispositivo com “*underlap*” (em oposição à “*overlap*”) ou *extension-less* (sem a região de extensão), quando ainda resta uma região do canal sem o controle da porta.

Esta região entre a porta e o dreno (*underlap*) funciona como uma extensão de fonte e dreno sem dopagem adicional, dessa forma, apresentará um potencial e um campo elétrico lateral ainda menor atuando sobre o canal, resultando em uma melhora nos efeitos de canal curto [3] [4] [5] [45]. Além disso, já foram reportados alguns estudos deste dispositivo para aplicações analógicas e em seu uso como uma célula de memória, apresentando vantagens como: [3] [4] [5] [45]

- melhor tensão Early,
- melhor razão I_{on}/I_{off} ,
- melhor razão gm/I_D (que representa a amplificação obtida no dispositivo por energia fornecida [12]),
- menor corrente de fuga,

- promissor em termos de escalamento de seu uso como uma célula 1T-DRAM.

De acordo com [44], quando a tensão na porta é pequena, ou seja, o dispositivo está na inversão fraca, o comprimento efetivo do canal pode ser considerado como a soma do comprimento de porta e os comprimentos das extensões de fonte e de dreno.

$$L_{ef} \cong L_{porta} + 2L_{LDD} \quad (2.31)$$

Já na inversão forte, onde o canal já está formado, o comprimento efetivo se aproxima do comprimento de porta.

$$L_{ef} \cong L_{porta} \quad (2.32)$$

3 Influência da concentração das regiões de extensão de fonte e dreno na 1T-DRAM

Na primeira etapa do trabalho, foi realizado um estudo das polarizações a serem utilizadas nas simulações para que o funcionamento da 1T-DRAM fosse corretamente observado e otimizado na estrutura considerada referência e, em seguida, o estudo da influência da concentração das regiões de extensão de fonte e dreno propriamente dita. Dessa forma, este capítulo é dividido em duas partes: o estudo das polarizações e o impacto da concentração das regiões de extensão de fonte e dreno na 1T-DRAM.

3.1 Estudo das polarizações

Este estudo foi dividido entre os diferentes estados da memória, seguindo a sequência de estados, obtidas por simulação [40], indicada pela Figura 3.1.

Figura 3.1: Sequência de estados utilizada.

As polarizações iniciam-se com a operação de escrita (primeiro conjunto de pulsos) e sua subsequente leitura (segundo conjunto de pulsos), confirmando a escrita do bit '1'. Em seguida, foram aplicadas os terceiro e quarto conjunto de pulsos como polarizações de escrita '0' e sua subsequente leitura, respectivamente, garantindo a saída das lacunas no corpo. Entre cada estado foi inserido uma condição de repouso.

Foram estudadas as polarizações nos estados de escrita '1', leitura e repouso, bem como a polarização do substrato, por serem os mais importantes para o presente trabalho.

3.1.1 Escrita do bit '1' [29] [30]

Os dados apresentados nesta seção são os obtidos por simulação com a estrutura gerada no ATLAS (Anexo 2 e Tabela 3.2) e por meio de simulações dinâmicas, gerando os resultados presentes na Figura 3.1 e na Tabela 3.1, que mostra a sequência de estados e polarizações e a corrente de dreno resultante ao longo do tempo das simulações utilizadas como referência nesta etapa do trabalho. O substrato foi aterrado e a temperatura variada de 25°C a 300°C. A sequência de polarizações da Tabela 3.1 se encontra na mesma sequência que a mostrada na Figura 3.1.

Tabela 3.1: Sequência de programação da memória.

Duração [ns]	V_G [V]	V_D [V]	Estados
10	-0,8	0	Repouso
30	0	1,2	Escrita '1'
10	-0,8	0	Repouso
5	-0,4	1,2	Leitura '1'
10	-0,8	0	Repouso
5	1,0	-1,0	Escrita '0'
10	-0,8	0	Repouso
5	-0,4	1,2	Leitura '0'
10	-0,8	0	Repouso

Tabela 3.2: Estrutura do dispositivo gerada no ATLAS.

L	150 nm
W	1 μm
t_{ox}	5 nm
t_{BOX}	20 a 50 nm
t_{Si}	8 a 50 nm
N_A	$1 \times 10^{15} \text{ cm}^{-3}$
N_D	$1 \times 10^{20} \text{ cm}^{-3}$
N_{LDD}	$1 \times 10^{19} \text{ cm}^{-3}$

Inicialmente, essas polarizações foram escolhidas baseadas na literatura [21] [25] da seguinte maneira:

- para a escrita do bit '1' foi necessário um alto V_D para a ativação do efeito BJT, juntamente com um V_G tal que permita essa ativação, que ocorre na região de sublimiar;
- para a leitura, também se necessita de um alto V_D para que o efeito possa ser ativado, porém com um V_G mais baixo que o da escrita do dado '1', para que se verifique seus 2 estados (um nível de corrente mais baixo para a leitura do bit '1' e outro mais alto para a leitura do dado '0') a partir da mesma polarização de leitura.
- para a escrita do '0' foram utilizados o método da polarização direta no dreno, portanto uma tensão negativa de V_D com um V_G maior ou igual à tensão de limiar da primeira interface.
- para o estado de repouso foi utilizado um V_D aterrado, para manter as cargas armazenadas no corpo do transistor, juntamente com um V_G mais negativo que o V_G da leitura para que não cause distúrbios na célula de memória.

Foram estudadas as polarizações nos estados de escrita '1', leitura e repouso, bem como a polarização de substrato por serem os mais determinantes para o presente trabalho, como será explicado ao longo das análises a seguir.

3.1.1.1 Comportamento em altas temperaturas

A Figura 3.2 mostra a margem de sensibilidade em função da temperatura, em que se observa um comportamento decrescente deste parâmetro com o aumento da temperatura.

Figura 3.2: Margem de sensibilidade em função da temperatura.

A equação (3.1) [38] representa a dependência da corrente de coletor ($I_{Coletor}$) com a temperatura.

$$I_{Coletor} = \alpha_F \cdot I_{ES} \cdot \left[e^{\left(\frac{q \cdot V_{BE}}{k \cdot T} \right)} - 1 \right] \quad (3.1)$$

em que α_F é o ganho da base em comum, I_{ES} é a corrente de saturação reversa da junção do emissor, V_{BE} é a tensão entre a base e o emissor e k é a constante de Boltzmann.

A equação (3.1) descreve o comportamento do BJT inerente à estrutura MOS. Apesar da corrente I_{ES} aumentar com a temperatura, a parcela exponencial reduz. Isso pode levar à diminuição da corrente de dreno quando o efeito está ativo (estado '1'), uma vez que a corrente de coletor é somada à corrente oriunda do transistor MOS formando a corrente total. Além disso, a menor mobilidade em altas temperaturas também pode levar à redução da corrente do estado '1'.

A Figura 3.3 mostra o tempo necessário para a escrita do nível '1' (disparo do efeito BJT) em função da temperatura.

Figura 3.3: Tempo necessário para o disparo do efeito BJT em função da temperatura.

A corrente de lacunas ($I_{lacunas,ger}$) é originada da ionização por impacto, como em altas temperaturas há uma redução da mobilidade devido a uma maior vibração da rede cristalina, isso faz com que a geração dos pares elétrons-lacunas por ionização por impacto diminua e a corrente de lacunas também.

Logo, em altas temperaturas, ambas as correntes diminuem, porém, a redução da corrente de lacunas (corrente de base) é maior que da corrente de coletor, indicando um maior ganho (β), como se pode observar pela equação (3.2). [38]

$$\beta = \frac{I_{Coletor}}{I_{lacunas,ger}} \quad (3.2)$$

Como já foi deduzido na seção 2.3.1, a equação 2.30 mostra a corrente de dreno considerando o disparo do efeito BJT. A partir dela percebe-se a relação entre a corrente e o ganho (β).

Em altas temperaturas, o ganho aumenta e a tensão de limiar diminui, antecipando a realimentação positiva e, conseqüentemente, o tempo para se ativar o efeito BJT torna-se menor (Figura 3.3).

A Figura 3.4 mostra o tempo mínimo de disparo em função da tensão de dreno para várias temperaturas. Na Figura 3.5 encontra-se a mínima tensão de disparo no dreno em função da espessura da camada de silício para várias temperaturas, enquanto que a Figura 3.6 apresenta a mínima tensão de disparo no dreno em função da espessura do óxido enterrado para várias temperaturas.

Figura 3.4: Tempo mínimo de disparo em função da tensão no dreno para várias temperaturas.

Observa-se pela Figura 3.4, que para maiores tensões de dreno, a temperatura atua menos no tempo de escrita.

Figura 3.5: Mínima tensão de disparo no dreno em função da espessura do silício para várias temperaturas.

Figura 3.6: Mínima tensão de disparo no dreno em função da espessura do óxido enterrado para várias temperaturas.

Com relação à diminuição da espessura da camada de silício, a Figura 3.5 indica um aumento da tensão no dreno necessária para ativar o efeito BJT. Isso ocorre devido ao melhor acoplamento do canal que leva a uma menor influência do controle do campo elétrico lateral sobre as cargas do canal, aumentando a tensão de disparo.

O mesmo comportamento é observado com a variação da espessura do óxido enterrado (Figura 3.6), ou seja, à medida que essa dimensão diminui, o potencial do substrato se torna mais próxima do corpo e seu controle sobre as cargas do canal aumenta, dificultando a atuação da tensão aplicada no dreno no disparo do efeito BJT.

Apesar dessas tendências mostrarem uma maior dificuldade para se ativar o efeito, observa-se na Figura 3.7 que a tensão de disparo pode ser substancialmente reduzida por meio da polarização positiva do substrato.

Figura 3.7: Mínima tensão de disparo no dreno em função da polarização do substrato.

De acordo com [36], isso ocorre porque a tensão positiva no substrato diminui a tensão de limiar, aumentando a corrente, facilitando a ionização por impacto que ativa o efeito BJT. Em outras palavras, há um adiantamento da realimentação (Figura 2.11-c), pois já está se aplicando um potencial no corpo que, originalmente, seria resultado dos primeiros ciclos da ionização por impacto.

3.1.1.2 Tempo de retenção

A Figura 3.8 mostra o tempo de retenção com o aumento da temperatura, onde se verifica sua diminuição com a temperatura. Em [25] considerou-se como diferença entre o nível '1' e '0', uma diferença de corrente de $20\mu\text{A}$. Como a extração do tempo de retenção está muito relacionada com o circuito de leitura, um nível de corrente de $40\mu\text{A}/\mu\text{m}$ foi utilizada nessa etapa do trabalho, garantindo os $20\mu\text{A}$ da referência citada. Sua extração seria algo parecido com a Figura 2.21.

Figura 3.8: Tempo de retenção em função da temperatura.

A equação (3.3) descreve a corrente reversa de fuga ($I_{\text{fuga,pn}}$) em uma junção PN. [46]

$$I_{\text{fuga,pn}} = q \cdot A \cdot \frac{n_i \cdot W_d}{\tau_e} + q \cdot A \cdot \left(\frac{D_n}{\tau_n} \right)^{\frac{1}{2}} \cdot \frac{n_i^2}{N_A} \quad (3.3)$$

onde a concentração intrínseca (n_i) é dada pela equação (3.4)

$$n_i = 3,9 \cdot 10^{16} \cdot T^{\frac{3}{2}} \cdot e^{-\left(\frac{E_g}{2 \cdot k \cdot T}\right)} \quad (3.4)$$

Nessas equações A é a área da seção transversal da junção, W_d é a largura da camada de depleção, τ_n é o tempo de vida efetivo, τ_e é o tempo de vida do elétron, D_n é o coeficiente de difusão do elétron e E_g é a largura da banda proibida.

A partir das equações (3.3) e (3.4) conclui-se que a corrente de fuga aumenta com a temperatura. Isso significa que as lacunas não removidas mais rapidamente do canal com o aumento da temperatura do dispositivo, diminuindo o tempo de retenção.

3.1.2 Leitura e repouso

Para estes estados foram estudados a dependência das polarizações de porta, substrato e dreno principalmente sobre a margem de sensibilidade e o

tempo de retenção. Utilizou-se o método de escrita do nível '1' pelo efeito BJT, por sua maior margem de sensibilidade e maior tempo de retenção, como dito anteriormente, e o acoplamento capacitivo para a escrita do '0' pela sua melhor eficiência (menores tensões e correntes).

Nesta etapa do trabalho as diferentes polarizações dos estados da memória foram otimizadas por meio de simulações, cujas tendências foram confirmadas com dados experimentais. A partir daqui, principalmente nos trabalhos onde não foi possível se obter dados experimentais, esta foi a base das polarizações consideradas.

A estrutura utilizada é aquela gerada no simulador ATHENA [40] (comandos de entrada no Anexo 1 e características resultantes na Tabela 3.3). Maiores informações sobre esta estrutura pode ser encontrada na referência [27].

Tabela 3.3: Estrutura do dispositivo gerada no ATHENA.

L	170 nm
L _{ef}	105 nm
W	1 μm
t _{ox}	5 nm
t _{BOX}	10 nm
t _{Si}	20 nm
N _A	$1 \times 10^{15} \text{ cm}^{-3}$

3.1.2.1 Tensão na porta [42] [43]

A Tabela 3.4 e a Figura 3.9 mostram a sequência de estados e as polarizações utilizadas como referência nas simulações [40] desta etapa do trabalho. Aplicou-se a tensão de 2,1 V no substrato e a temperatura foi mantida a 85°C.

Tabela 3.4: Estados e polarizações utilizadas. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.

Duração [ns]	V _G [V]	V _D [V]	Estado
10	0	2	Escrita '1'
10	0	0,5	Escrita '0'
10	-1,5	2	Leitura
-	-2,5	0	Repouso

Figura 3.9: Esquema de programação utilizada. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.

Na Figura 3.10 encontram-se os níveis de corrente dos estados '1' e '0', bem como a margem de sensibilidade para uma tensão no substrato de 2,1V.

Figura 3.10: Janela da tensão na porta durante a leitura.

Observa-se na Figura 3.10 que a janela de leitura está localizada entre as tensões de porta -2,5V e -1V e a melhor tensão de porta para a leitura é de -1,5 V com uma margem de sensibilidade de $203 \mu\text{A}/\mu\text{m}$. Verifica-se também um aumento da corrente nos estados '1' e '0' para polarizações mais negativas na porta devido à geração originada pelo GIDL.

A Figura 3.11 indica a margem de sensibilidade em função da tensão na porta durante o repouso, onde se vê que a margem de sensibilidade aumenta para valores mais negativos da tensão de porta nesse estado até que se torna constante a partir de -2,5V devido à maior geração pelo tunelamento de banda a banda (BBT), confirmada pela Figura 3.12 que mostra a taxa de geração para tensões de porta de -1,9V e -2,9V ao longo do canal em um corte realizado 1nm abaixo da 1ª interface.

Figura 3.11: Margem de sensibilidade em função da tensão na porta durante o tempo de repouso.

Figura 3.12: Taxa de geração a 1 nm abaixo da 1ª interface durante o estado de repouso.

A influência da tensão de porta durante o repouso no tempo de retenção pode ser vista na Figura 3.13 com dados simulados e na Figura 3.14 com dados

experimentais. Ambas apresentaram a mesma tendência, indicando que, dependendo da tensão de porta durante o repouso, pode haver uma degradação do '1' (gráficos da esquerda) ou do '0' (gráficos da direita). Esses gráficos possibilitaram o levantamento das curvas da Figura 3.15, considerando como a corrente de referência $50\% \Delta I_S$.

Figura 3.13: Corrente do estado '1' e '0' simulada ao longo do tempo para diferentes tensões na porta durante o repouso. A degradação do dado pode ser devido a recombinação (a) ou geração (b).

Figura 3.14: Corrente do estado '1' e '0' medida ao longo do tempo para diferentes tensões na porta durante o repouso. A degradação do dado pode ser devido a recombinação (a) ou geração (b).

Figura 3.15: Tempo de retenção em função da tensão na porta durante o repouso obtido por simulação (a) e por medidas (b).

Esse comportamento deve-se ao fato pela maior geração, para valores mais negativos de tensão na porta durante o repouso, ou maior recombinação para valores menos negativos de $V_{G_Repouso}$, podendo ser confirmado a partir da Figura 3.16.

A Figura 3.16-a mostra a taxa de recombinação na estrutura em um corte próximo à 1ª interface para um $V_{G_Repouso}$ menos (-1,5V) e mais (-2,3V) negativo. Verifica-se uma maior recombinação para a polarização menos negativa.

Figura 3.16: Comparação das taxas de recombinação (a) e de geração (b) durante o repouso no estado '1' a 1nm abaixo da 1ª interface para $V_{G_Repouso}$ de -1,5V e -2,3V.

Com a tensão de porta durante a leitura mantida em -1,5V, esta condição de repouso implica em uma leitura de cargas fora do equilíbrio armazenadas no corpo. Estas cargas se recombinam rapidamente, levando a uma menor corrente de leitura. Por isso, polarizar o $V_{G_Repouso}$ muito próximo do $V_{G_Leitura}$ não é adequado e leva a menores tempos de retenção para o estado '1' devido ao decaimento das cargas fora do equilíbrio no corpo. Abaixo desse valor, o equilíbrio das cargas do estado '1' é lido normalmente e elas não são expostas à recombinação durante o repouso.

Porém, para valores mais negativos de $V_{G_Repouso}$, a memória entra na polarização em que ocorre o GIDL, havendo a degradação do estado '0' devido à geração de lacunas que vão para o corpo, como mostra a Figura 3.16-b, em que se observa uma baixa taxa de geração para uma tensão de porta durante o repouso de -1,5V enquanto uma taxa muito maior é verificada para um $V_{G_Repouso}$ de -2,3V.

A fim de otimizar a margem de sensibilidade (menor que -2,5V para obter uma maior margem de sensibilidade), evitando distúrbios na matriz de memória

(menor que -2,5V para sair da polarização referente à janela de leitura) e para se obter um maior tempo de retenção, o melhor $V_{G_Repouso}$ é de -2,5V.

3.1.2.2 Tensão no substrato [42] [43]

Por meio das correntes no estado '0' e '1' e da margem de sensibilidade para diferentes tensões no substrato, obteve-se a janela de leitura em função da tensão no substrato (Figura 3.17). Da Figura 3.17 verifica-se uma janela de leitura compreendendo tensões de substrato de 2,1 V a 2,6 V. Abaixo de 2,1 V, o V_B não é suficiente para ativar o efeito de corpo flutuante para $V_D = 2$ V, ou seja, a tensão no dreno deveria ser maior para haver o efeito de corpo flutuante, permitindo a escrita do bit '1'. Já altos valores de V_B (no caso da Figura 3.17, para valores maiores que 2,6 V) facilita a ativação do efeito BJT devido ao maior efeito de corpo flutuante (resultado do maior potencial de corpo proporcionado pela tensão no substrato) e a uma menor tensão de limiar da primeira interface, levando ao deslocamento da janela de programação para valores menores de V_G (Figura 3.18) e, conseqüentemente, à ativação do efeito BJT ocorre mesmo durante a leitura do bit '0' [36] [47].

Figura 3.17: Janela de tensão do substrato.

Figura 3.18: Janela de programação para várias tensões do substrato. [36]

Dessa forma, as condições ideais para o repouso da memória FBRAM é com a 1ª interface em acumulação e a 2ª interface em depleção. Figura 3.19 mostra a concentração de lacunas ao longo da estrutura durante o repouso do dado '1'. As

lacunas são acumuladas na 1ª interface, o que demonstra que esta interface está em acumulação e a 2ª na depleção, uma vez que há lacunas próximas a esta última.

Figura 3.19: Concentração de lacunas ao longo da estrutura durante o estado de repouso do dado '1' com $V_{G_Repouso} = -2.5\text{ V}$ e $V_B = 2.1\text{ V}$.

Com relação ao tempo de retenção, observa-se que a melhor polarização no substrato para se obter a melhor margem de sensibilidade não é a mesma para se obter o tempo de retenção mais longo. A Figura 3.20 mostra a corrente do estado '1' e do '0' ao longo do tempo para diferentes tensões no substrato e a Figura 3.21 apresenta os tempos de retenção extraídos a $50\%\Delta I_S$ em função da polarização do substrato, indicando uma tensão ótima de 2.1V, porém, considerando certa tolerância para variações em sua polarização, já que neste ponto qualquer variação pode ser determinante para a escrita da memória, a melhor polarização seria em torno de 2,2V.

Confrontando com o gráfico da Figura 3.17, verifica-se que um $V_B = 2,2\text{V}$ não é a polarização de maior ΔI_S , entretanto, seu valor ainda é aceitável. Dessa forma, considerando que o tempo de retenção é um parâmetro mais crítico que a margem de sensibilidade, o melhor valor para a tensão do substrato é de 2,2V.

De acordo com [48], o comportamento decrescente do tempo de retenção para tensões de substrato maiores pode ser devido à maior fuga das junções, aumentando a geração de lacunas.

Figura 3.20: Corrente do estado '1' e '0' ao longo do tempo para diferentes tensões de substrato. As polarizações de porta e dreno são as apresentadas na Figura 3.9.

Figura 3.21: Tempo de retenção em função da tensão no substrato.

3.1.2.3 Tensão no dreno [41]

A Tabela 3.4 e a Figura 3.22 mostram a sequência de estados e as polarizações utilizadas como referência para esta etapa do trabalho (simulações [40] e medidas). Aplicou-se a tensão de 2,1 V no substrato e a temperatura foi mantida a 85°C.

Tabela 3.4: Estados e polarizações utilizadas. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.

Duração [ns]	V_G [V]	V_D [V]	Estado
10	0	2	Escrita '1'
10	0	0,5	Escrita '0'
10	-1,5	2	Leitura
-	-2,5	0	Repouso

Figura 3.22: Esquema de programação utilizada. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo.

A Figura 3.23 apresenta a margem de sensibilidade em função da tensão na porta para diferentes tensões no dreno durante a leitura. Percebe-se que à medida que a tensão no dreno durante a leitura aumenta, a tensão na porta com a maior margem de sensibilidade se desloca para a esquerda, além disso, essa maior margem de sensibilidade também aumenta. Em outras palavras, para cada tensão no dreno durante a leitura, há uma tensão na porta ótima.

Figura 3.23: Margem de sensibilidade em função da tensão na porta para diferentes tensões de dreno durante a leitura (dados experimentais).

Da figura anterior, extraiu-se a tensão na porta que otimiza a margem de sensibilidade. Esses dois parâmetros estão apresentados na Figura 3.24 em função da tensão no dreno durante a leitura. Nota-se que ambos os parâmetros apresentam um comportamento bem diferente a partir de $V_{D_Leitura} = 1,2$ V. Para valores menores, tanto a tensão na porta ótima quanto a margem de sensibilidade são praticamente constantes, enquanto que para valores maiores, a tensão na porta ótima diminui e a margem de sensibilidade aumenta fortemente com o aumento da tensão no dreno.

Figura 3.24: Melhor tensão na porta durante a leitura e a margem de sensibilidade em função da tensão no dreno durante a leitura.

As curvas da Figura 3.25 representam a janela de leitura para diferentes tensões de leitura no dreno obtida por medidas (a) e por simulação (b). Ambas indicam uma corrente no estado '1' maior, para alta tensão no dreno.

Figura 3.25: Janela de leitura para diferentes tensões no dreno durante a leitura com dados experimentais (a) e simulados (b).

Para entender o motivo de tal comportamento foram realizadas duas simulações, uma considerando o efeito de ionização por impacto (II) e outra sem. A Figura 3.26 mostra a margem de sensibilidade em função da tensão de leitura no dreno com e sem o modelo de ionização por impacto.

Figura 3.26: Margem de sensibilidade simulada em função da tensão no dreno durante a leitura.

Percebe-se que a margem de sensibilidade diminui drasticamente quando o modelo é retirado, possuindo um nível equivalente ao daquele apresentado com a polarização de leitura no dreno menor que 1,2 V. Assim, pode-se afirmar que o aumento da margem de sensibilidade quando se aplica uma alta tensão no dreno durante a leitura está relacionada com a ionização por impacto e o disparo do efeito BJT. [18] [41]

Já a dependência da polarização no dreno durante a leitura no tempo de retenção pode ser observado na Figura 3.27 e na Figura 3.28. A primeira apresenta as correntes no estado '0' e '1' ao longo do tempo para diferentes tensões de leitura do dreno obtido por medida (a) e por simulação (b). A partir desta, extraiu-se o tempo de retenção, traçando-o em função da tensão no dreno durante a leitura na Figura 3.28.

Figura 3.27: Corrente nos estados '0' e '1' para alta e baixa polarização de dreno durante a leitura. Resultados experimentais (a) e simulados (b).

Figura 3.28: Tempo de retenção em função da tensão no dreno durante a leitura.

Apesar da margem de sensibilidade ser maior para maiores tensões no dreno, o tempo de retenção resultou menor para essas mesmas polarizações. Da Figura 3.27, vê-se que o tempo de retenção é determinado pela degradação do bit '0', o que pode ser um indício de que a razão desse menor tempo de retenção é o fato de haver ionização por impacto para altas tensões no dreno, o que levaria a

mais lacunas serem injetadas no corpo durante a leitura, ativando o estado '1' antes.

Analisando a situação de múltiplas leituras, os resultados anteriores levam a previsão de que uma tensão no dreno maior injetará lacunas no corpo durante a leitura e menor será a quantidade de leituras sem a necessidade da atualização (*refresh*). A Figura 3.29 confirma esta previsão, mostrando que para altas tensões o dreno (a) o dado é perdido após 22 leituras, enquanto para tensões mais baixas no dreno (b), mesmo após 30 leituras, o dado ainda não foi perdido.

Figura 3.29: Múltiplas leituras (30x) para alta (a) e baixa (b) tensão de leitura do dreno.

3.2 Impacto da concentração de extensão na 1T-DRAM [49]

A estrutura utilizada nesta etapa do trabalho foi a construída no simulador ATHENA (anexo 1) baseado em dados de um dispositivo do imec (*Interuniversity Microelectronics Center*). Para se obter as diferentes concentrações das extensões de fonte e dreno, sua dose de implantação foi variada de 0 até $1 \times 10^{15}\text{ cm}^{-2}$ (dose utilizada como referência), obtendo-se concentrações de $1 \times 10^{15}\text{ cm}^{-3}$ (tipo P, ou seja, referente o dispositivo sem extensão de fonte e dreno) a $6 \times 10^{19}\text{ cm}^{-3}$ (tipo N), respectivamente. A estrutura utilizada como referência foi aquela cujas polarizações foram otimizadas nas seções anteriores (seção 3.1.2).

Tabela 3.5: Doses de implantação e a concentrações obtidas para as diferentes regiões de extensão de fonte e dreno. O dispositivo de dose 0 refere-se àquele sem extensão.

Dose de implantação [cm^{-3}]	1×10^{15}	1×10^{14} (tipo n)	0
Concentração [cm^{-3}]	6×10^{19} (tipo n)	1×10^{19} (tipo n)	1×10^{15} (tipo p)

As simulações foram modeladas usando simulações numéricas bidimensionais [40], possuindo as seguintes dimensões: comprimento de canal (L) de máscara igual a 170 nm e o efetivo igual a 105 nm, espessura do óxido de porta (t_{ox}) de 5 nm, espessura do óxido enterrado (t_{BOX}) de 10 nm e espessura do filme de silício (t_{Si}) de 20 nm. A região do canal não foi dopada, possuindo apenas sua concentração natural de $1 \times 10^{15} \text{ cm}^{-3}$. A Figura 2.7-b apresenta uma visão esquemática da estrutura utilizada nessa etapa do trabalho.

Nesta parte do trabalho, embora não sejam apresentadas medidas experimentais, a estrutura utilizada foi aquela otimizada anteriormente (seção 3.1.2). Além disso, ela foi ajustada de forma a apresentar valores mais próximos de uma medida experimental.

3.2.1 Resultados e análises

Seguindo a sequência de estados apresentada em Figura 3.1, os parâmetros estudados foram obtidos por meio da curva da tensão no dreno em função do tempo apresentada na Figura 3.30, e polarizações indicadas na Tabela 3.4. Da mesma forma que na seção 3.1.2, foram considerados os métodos do efeito BJT para a escrita do '1' e acoplamento capacitivo para a escrita do '0'. A temperatura utilizada foi de 85°C (temperatura de operação da maioria das memórias), a fonte foi aterrada e o substrato foi polarizado com 2,1 V.

Figura 3.30: Esquema de programação na memória 1T-FBRAM.

Tabela 3.4: Estados e polarizações utilizadas para a estrutura referência (concentração de $6 \times 10^{19} \text{ cm}^{-3}$)

Duração [ns]	V_G [V]	V_D [V]	Estado
10	0,0	2,0	Escrita '1'
10	0,0	0,5	Escrita '0'
10	-1,5	2,0	Leitura
-	-2,5	0,0	Repouso

3.2.1.1 Margem de sensibilidade e janela de leitura.

A Tabela 3.6 mostra as mínimas tensões no dreno usadas para se ativar o efeito ($V_{Dmin_escrita}$) para diferentes concentrações das regiões de extensão de fonte e dreno, mantendo-se a tensão na porta durante a escrita em 0 V. Estas polarizações mostraram-se maiores para menores concentrações das regiões de extensão devido ao menor campo elétrico lateral, o que dificulta a ionização por impacto. Como já mencionado e explicado anteriormente (seção 3.1.1), esta tensão pode ser reduzida aplicando-se uma tensão positiva no substrato [36].

Tabela 3.6: Tensão no dreno durante a escrita.

Concentração [cm^{-3}]	6×10^{19} (tipo n)	1×10^{19} (tipo n)	1×10^{15} (tipo p)
$V_{Dmin_escrita}$ [V]	2,0	2,6	3,4

Com as polarizações de escrita definidas, a Figura 3.31 apresenta a janela de leitura da tensão na porta para diferentes concentrações da região de extensão, bem como sua margem de sensibilidade.

Figura 3.31: Janela de Leitura da tensão de porta para estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (a), $1 \times 10^{19} \text{ cm}^{-3}$ (b) e sem extensão; (c), e a margem de sensibilidade para diferentes concentrações da região de extensão (d).

Figura 3.31: Janela de Leitura da tensão de porta para estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (a), $1 \times 10^{19} \text{ cm}^{-3}$ (b) e sem extensão; (c), e a margem de sensibilidade para diferentes concentrações da região de extensão (d).

Dessas curvas, verifica-se uma menor degradação da corrente no estado ‘0’ para tensões mais negativas na porta, indicada pelo aumento nas correntes lidas da Figura 3.31-a, devido à diminuição da concentração da região de extensão, o que dificulta o tunelamento de banda a banda devido ao menor potencial efetivamente aplicado e, conseqüentemente, menor campo elétrico na região de sobreposição da porta sobre o dreno.

A tensão de leitura na porta ($V_{G_Leitura}$) foi extraída no valor de máxima margem de sensibilidade. Esses dois parâmetros estão indicados na Tabela 3.7 para diferentes concentrações da região de extensão. Observa-se que a janela de leitura é levemente deslocada para a direita com a diminuição da concentração da região de extensão (Figura 3.31-d) e, conseqüentemente, a melhor polarização de porta durante a leitura aumenta (ela se torna menos negativa), ao mesmo tempo que a margem de sensibilidade diminui.

Tabela 3.7: Tensão de leitura na porta e máxima margem de sensibilidade para diferentes concentrações da região de extensão.

Concentração [cm^{-3}]	6×10^{19} (tipo n)	1×10^{19} (tipo n)	1×10^{15} (tipo p)
$V_{G_Leitura}$ [V]	-1,5	-1,3	-1,1
ΔI_S [$\mu A/\mu m$]	203	200	170

3.2.1.2 Tempo de retenção

A Figura 3.32 apresenta as correntes nos estados ‘0’ e ‘1’ ao longo do tempo para as três concentrações da região de extensão estudadas, indicando que nos

três casos, houve a degradação do estado '0'. A partir dessa figura, percebe-se que o dispositivo sem extensão de fonte e dreno (também denominado *extension-less*) apresentou o tempo de retenção mais longo, mesmo com uma polarização mais alta de leitura, o que reduziria o tempo de retenção [41]. Esse comportamento é justificado pela maior taxa de geração de lacunas durante o repouso da estrutura referência (Figura 3.33), sendo ela doze ordens de grandeza menor para a estrutura sem extensão de fonte e dreno quando comparado com a estrutura referência.

Figura 3.32: Correntes no estado '1' e '0' ao longo do tempo para estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (quadrado), $1 \times 10^{19} \text{ cm}^{-3}$ (círculo) e com *underlap* (triângulo).

Figura 3.33: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com concentrações da região de extensão de $6 \times 10^{19} \text{ cm}^{-3}$ (quadrado), $1 \times 10^{19} \text{ cm}^{-3}$ (círculo) e com *underlap* (triângulo).

A Tabela 3.8 mostra os tempos de retenção extraídos da Figura 3.32, considerando o tempo em que a corrente de dreno se aproxima de $50\% \Delta I_S$ para as

diferentes concentrações da região de extensão, indicando uma melhora de 3 ordens de grandeza para a estrutura sem extensão.

Tabela 3.8: Tempo de retenção para as estruturas com diferentes concentrações da região de extensão.

Concentração [cm^{-3}]	6×10^{19} (tipo n)	1×10^{19} (tipo n)	1×10^{15} (tipo p)
Tempo de Retenção [ms]	0,2	56,0	167,0

4 Escalamento em dispositivos UTBOX sem extensões de fonte e dreno em termos de memória

A fim de se aprofundar no estudo dos dispositivos sem extensões de fonte e dreno, nesta seção se encontra um estudo acerca do escalamento do tempo de retenção em dispositivos sem e com as regiões de extensão de fonte e dreno. As simulações foram realizadas de forma a explicar as tendências experimentais e melhor entender os mecanismos físicos envolvidos.

Para os resultados apresentados neste capítulo, foram utilizadas a mesma estrutura presente no anexo 1, porém com a espessura do filme de silício de 14nm e do óxido enterrado de 18nm, além de variar o comprimento do canal de 50nm a 220nm. Mais detalhes dessa estrutura podem ser encontrados em [47] e [50]. Para os dispositivos sem extensão de fonte e dreno, a região em questão foi obtida da mesma forma que na seção 3.2, ou seja, retirando a dopagem da região de extensão, e o comprimento do canal variando de 30nm a 220nm. Maiores informações sobre estas estruturas estão em [47] e [50].

4.1 Dispositivos auto-alinhados [47]

A Figura 4.1 e a Tabela 4.1 mostram a sequência de estados e polarizações utilizadas para esta etapa do trabalho, escolhidas de forma a seguir com aquelas utilizadas para a obtenção dos resultados experimentais. A fonte foi aterrada e o substrato foi polarizado positivamente. Novamente, todas as análises (simulações [40] e medidas) foram realizadas a 85°C.

Figura 4.1: Esquema de polarização utilizada.

Tabela 4.1: Estados e polarizações utilizadas.

Duração [ns]	V_G [V]	V_D [V]	Estado
40	0,5	1,8	Escrita '1'
40	0,0	0,5	Escrita '0'
40	-1,8	1,8	Leitura
-	-2,5	0,0	Repouso

Da mesma forma que o apresentado na seção 3.1.2.2, a Figura 4.2 também apresenta uma redução no tempo de retenção proporcionado pelo aumento de V_B tanto por simulação (a) quanto por medidas experimentais (b). Isso pode estar relacionado com a maior geração de lacunas devido às correntes de fuga das junções com o aumento do V_B .

Figura 4.2: Tempo de retenção para várias polarizações de substrato. Dados simulados (a) e experimentais (b).

Como o tempo de retenção é um dos principais parâmetros de interesse para a sua utilização nas gerações futuras da 1T-DRAM, a Figura 4.3 apresenta o tempo de retenção em função do comprimento do canal já considerando as diferentes polarizações de substrato por simulação (a) e medidas experimentais (b).

Figura 4.3: Tempo de retenção em função do comprimento de canal considerando diferentes polarizações de substrato. Dados simulados (a) e experimentais (b).

Observa-se em ambos os gráficos, que é possível manter o tempo de retenção razoavelmente constante com a redução do comprimento de canal por meio da otimização da tensão no substrato até certo ponto. Para comprimentos de canal muito curtos, mesmo aplicando baixa tensões no substrato (que resulta em um tempo de retenção maior, Figura 4.2), não foi possível manter o dado a um tempo próximo dos apresentados para L mais longos, limitando-se assim sua utilização como memória.

Como já demonstrado até o momento, o principal mecanismo limitante do tempo de retenção pela degradação do dado '0' é a geração decorrente do GIDL. De acordo com [45], outro importante aspecto atua para comprimentos de canal menores: o fato do menor L significar também a menor base para o transistor bipolar intrínseco à estrutura MOS. Isto faz com que a pequena corrente originada pelo GIDL seja mais amplificada por este transistor BJT de base menor, resultando numa maior degradação e, conseqüentemente, menor tempo de retenção.

Desta forma, a Figura 4.4 mostra a corrente devido ao GIDL para várias tensões no substrato, por simulação (a) e medidas (b), demonstrando um aumento da corrente devido ao GIDL para tensões mais elevadas de substrato.

Figura 4.4: Corrente devido ao GIDL em função da tensão de substrato para vários comprimentos de canal. Dados simulados (a) e experimentais (b).

Já a Figura 4.5 mostra a corrente devido ao GIDL em função do comprimento do canal a partir de simulações (quadrado aberto) e medidas (esfera fechada), indicando, em ambos, um aumento abrupto da corrente para canais mais curtos.

Figura 4.5: Corrente devido ao GIDL em função do comprimento de canal a partir de dados simulados (quadrado aberto) e experimentais (esfera fechada). Dados experimentais encontrados em [45].

Finalmente a Figura 4.6 apresenta a taxa de geração ao longo do canal próxima à 1ª interface, indicando uma maior geração para o canal mais curto e V_B maior.

Figura 4.6: Taxa de geração ao longo do canal a 1nm abaixo da 1ª interface para várias tensões de substrato e para canal curto (a) e longo (b).

4.2 Dispositivos não auto-alinhados [50]

Diferente das estruturas tradicionais, em que há uma sobreposição da porta sobre as extensões de fonte e dreno, a região de extensão subposta serve como uma transição entre o potencial aplicado na fonte/no dreno e o aplicado na porta. Isso leva a uma variação mais suave do potencial nas regiões próximas às junções [44] resultando num menor campo elétrico nessas regiões. No caso da Figura 4.7, com o campo elétrico ao longo do canal próximo da 1ª interface, o campo elétrico da estrutura sem as regiões de extensão chega a ser aproximadamente duas vezes menor que a estrutura com LDD.

Figura 4.7: Campo elétrico a 1nm abaixo da 1ª interface para os dispositivos com extensão (a) e sem extensão (b).

A Figura 4.8 mostra o tempo de retenção em função do comprimento de canal já considerando as polarizações de substrato para os dispositivos sem extensão junto com os de referência obtidos tanto a partir de simulações [40] como por medidas experimentais.

O pequeno gráfico no canto da Figura 4.8-a mostra as correntes lidas no estado '0' e '1' ao longo do tempo, obtidas por simulação. Pode-se perceber que neste caso também ocorre a degradação do dado '0', uma vez que é o I_0 que se eleva após algum tempo, indicando que há uma geração de lacunas, como já apresentado anteriormente para os dispositivos referência e como também será comprovado mais adiante para os dispositivos sem extensão.

Da mesma forma que na seção anterior, também foi possível manter o tempo de retenção aproximadamente constante com a redução do comprimento do canal a partir da polarização positiva do substrato, sendo limitado pela amplificação da corrente no estado '0' pelo transistor BJT de menor base (menor comprimento do canal do transistor MOS). [45]

Figura 4.8: Tempo de retenção em função do comprimento do canal para dispositivos com extensão (estrela fechada) e sem extensão (círculo aberta). Resultados simulados (a) e experimentais (b). O pequeno gráfico da figura (a) são as correntes lidas no estado '0' e '1' em função do tempo.

Comparando-se as duas curvas, observa-se que ambas apresentam a mesma tendência, porém com tempos de retenção maiores. Além disso, verifica-se que a redução do tempo de retenção da estrutura sem extensão se dá para estruturas com comprimentos de canal menores; em outras palavras, esta estrutura é mais escalável em sua aplicação como célula de memória.

O maior tempo de retenção dos dispositivos sem extensão está relacionado com o menor campo elétrico lateral (Figura 4.7), o que diminui a taxa de geração

de lacunas e, consequentemente, a degradação do dado '0'. Já a queda do tempo de retenção dos dispositivos sem extensão para comprimentos de canal menores é devida tanto à menor taxa de geração, quanto ao maior comprimento efetivo do canal (equação 2.31), isto é, à maior base do transistor BJT, o que amplificaria menos a corrente originada pela taxa de geração pelo tunelamento banda a banda da região sob os espaçadores.

A Figura 4.9 mostra a concentração de lacunas durante o repouso do estado '0' e '1' ao longo do tempo, demonstrando que há uma geração de lacunas fazendo com que a corrente I_0 se degrade, tornando-se próxima de I_1 . De acordo com essa figura, a concentração de lacunas referente ao momento da degradação neste caso é de $5,75 \times 10^{18} \text{ cm}^{-3}$.

Figura 4.9: Concentração de lacunas durante o repouso dos estados '0' e '1' em função do tempo.

A Figura 4.10 apresenta a distribuição da taxa de geração ao longo da estrutura sem extensão durante o repouso do estado '0' obtida por simulação considerando o modelo de BBT (a) e sem ele (b). Esses cortes indicam que o principal efeito atuando na geração de pares elétron-lacunas é o GIDL, visto que as maiores taxas de geração próximas às junções estão presentes apenas quando o modelo de BBT foi considerado (a).

Figura 4.10: Taxa de geração durante o repouso do estado '0' ao longo da estrutura sem extensão considerando o modelo de BBT (a) e sem considerá-lo (b).

É importante mencionar que, embora não haja a sobreposição da porta sobre as regiões de fonte e dreno nos dispositivos sem extensão, o aumento da corrente para tensões de porta mais negativos (a corrente de GIDL para os dispositivos com extensão) nestas estruturas pode ser explicado pela depleção devido ao espalhamento do potencial na região sob os espaçadores. [44]

Na Figura 4.11 encontra-se a taxa de geração próximo da 1ª interface durante o estado de repouso para os canais mais curtos de cada estrutura, isto é, 50nm para o dispositivo referência e 30 nm para o dispositivo sem extensão. Pode-se verificar uma taxa de geração próxima às junções entre as duas tecnologias aproximadamente iguais, o que está condizente com a Figura 4.8-a, uma vez que o tempo de retenção do dispositivo referência e do dispositivo sem extensão são também próximos, 0,2ms a 0,14ms, respectivamente. Isso indica uma correlação entre o campo elétrico, a taxa de geração e o tempo de retenção. Além disso, a obtenção dos tempos de retenção próximos só foi possível graças à utilização de um comprimento de canal e de tensões de substrato menores por parte dos dispositivos sem extensão, o que demonstra um melhor escalamento para os dispositivos sem extensão, além da possibilidade de se aplicar tensões menores de substrato.

Figura 4.11: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com (a) e sem extensão de fonte e dreno (b).

É importante ressaltar que a degradação no tempo de retenção desses dispositivos não é diferente apenas durante o repouso, mas também durante a escrita. A Figura 4.12 mostra a taxa de geração durante a leitura do estado '0' próximo à 1ª interface para as estruturas com e sem extensão de fonte e dreno com mesmo comprimento de canal e tensão no substrato, em que se vê uma taxa de geração próxima às junções um pouco maior nos dispositivos referência. Isso significa que a taxa de geração de lacunas amplificada pelo BJT também pode estar presente durante a leitura, facilitando a degradação do I_0 e, portanto, diminuindo o tempo de retenção.

Figura 4.12: Taxa de geração durante a leitura do estado '0' a 1 nm abaixo da 1ª interface para as estruturas com extensão (estrela fechada) e sem extensão (círculo aberto) de fonte e dreno.

5 Melhorias no tempo de retenção

A 1T-FBRAM usando o efeito BJT para a escrita do dado '1' tem sido reportado como o melhor método para atingir maiores tempos de retenção e margem de sensibilidade [18]. No entanto, para se atingir a especificação de 64ms das DRAMs convencionais [24], seu tempo de retenção ainda precisa ser melhorado. Para isso, além da utilização das estruturas sem extensão de fonte e dreno, também foram estudadas a influência no tempo de retenção através da polarização de substrato pulsada (seção 5.1) e da banda proibida (seção 5.2).

O dispositivo usado nesta etapa do trabalho foi aquele apresentado no capítulo 3.2 com a concentração das regiões de extensão consideradas referência. Dessa forma, embora não sejam apresentados resultados experimentais, as simulações foram ajustadas de forma a obter valores próximos de um dispositivo real.

5.1 *Utilização da polarização de substrato pulsada [51]*

Como já demonstrado na seção 3.1.1, a polarização positiva do substrato facilita a escrita do dado '1', diminuindo a mínima tensão de dreno necessária para o disparo do efeito BJT. Entretanto, quanto maior esta tensão de substrato, menor será o tempo de retenção devido ao maior GIDL durante o repouso e a leitura do estado '0'. Dessa forma, uma das propostas de melhoria do tempo de retenção é a utilização da polarização de substrato pulsada para se obter maiores tempos de retenção, permitindo uma polarização mais baixa do dreno durante a escrita do dado '1'.

É importante mencionar que a tecnologia de processo da referência [52] viabiliza o mecanismo presente neste estudo, permitindo que apenas uma célula/transistor tenha seu substrato polarizado de forma independente, reduzindo distúrbios nas células vizinhas.

As simulações [40] foram realizadas seguindo o esquema de programação dos diferentes estados e suas durações presentes na Figura 5.1 e na Tabela 5.1. A temperatura foi de 85°C, com o V_B igual a 2V para o caso do V_B constante ou usando a proposta de V_B pulsado com 2V somente durante a escrita do '1', com o

substrato aterrado nos outros estado. Foi aplicado um pulso mais largo durante a leitura somente para verificar o nível de corrente lido depois da estabilização do pulso, o que não foi necessário para a leitura do estado '0'.

Figura 5.1: Sequência de estados e polarizações utilizadas. Apenas a polarização de substrato foi alterada para esta análise. A duração da leitura foi maior devido a necessidade de confirmação da estabilização da corrente.

Tabela 5.1: Estados e polarizações utilizadas.

Duração [ns]	V_G [V]	V_D [V]	V_B constante [V]	V_B pulsado [V]	Estado
40	0,5	1,8	2	2	Escrita '1'
40	0,0	0,5	2	0	Escrita '0'
40	-1,8	1,8	2	0	Leitura
-	-2,5	0,0	2	0	Repouso

5.1.1 Princípio de operação e margem de sensibilidade

Não houve diferença entre a polarização constante e a pulsada na intensidade da ionização por impacto durante a escrita do bit '1', o que era esperado visto que, em ambos os casos, a polarização de substrato era a mesma neste estado.

Já durante a escrita do '0', a Figura 5.2 mostra uma maior densidade de corrente de lacunas para a polarização de substrato pulsada ($V_B=0V$ durante a escrita do '0') do que para a polarização constante ($V_B=2V$ durante a escrita do '0'). Isso pode ser devido à menor tensão de limar da 1ª interface para $V_B=2V$, o que significaria uma maior corrente de elétrons e uma menor corrente de lacunas para o caso constante.

Figura 5.2: Densidade de corrente de lacunas durante a escrita do estado '0' a 1 nm acima da 2ª interface utilizando V_B constante (quadrado tracejado) e pulsado (estrela contínua).

A Figura 5.3 mostra altas concentrações de lacunas no estado '1' em ambos os casos, demonstrando que a escrita do '1' pelo método do BJT foi realizada. Observa-se também uma redução da diferença entre as concentrações de lacunas dos estados '1' e '0' para o caso do substrato pulsado. Dessa forma, a margem de sensibilidade será menor, como se confirma na Figura 5.4, por meio da densidade de corrente de elétrons ao longo do canal durante a leitura do '1' e do '0' para a polarização de substrato constante (a) e pulsada (b).

Figura 5.3: Concentração de lacunas durante o repouso dos estados '0' (aberto) e '1' (fechado) a 1 nm abaixo da 1ª interface utilizando V_B constante (quadrado tracejado) e pulsado (estrela contínua).

Figura 5.4: Densidade de corrente de elétrons durante a leitura do '0' (aberto) e do '1' (fechado) a 1 nm acima da 2ª interface utilizando V_B constante (quadrado tracejado à esquerda) e pulsado (estrela contínua à direita).

A Figura 5.5 e a Tabela 5.2 também confirma essa redução, porém a partir da extração dos valores e o cálculo da margem de sensibilidade, observa-se uma redução de apenas 4%. Além disso, a polarização de substrato aterrada reduz o efeito BJT durante a leitura, como se vê na Figura 5.5, onde a corrente lida no caso pulsado demorou mais para se estabilizar. Isso permite um maior número de leituras sem a necessidade de atualização do dado, como visto na seção 3.1.2.3.

Figura 5.5: Corrente de dreno em função do tempo obtida para V_B constante (tracejada) e pulsada (contínua).

Tabela 5.2: Margem de sensibilidade e tempo de retenção para V_B constante e V_B pulsado.

	Margem de sensibilidade	Tempo de retenção
V_B constante	233 μ A	637 μ s
V_B pulsado	223 μ A	745 μ s
Melhora	- 4 %	17 %

5.1.2 Tempo de retenção

A Figura 5.6 mostra as correntes lidas no estado '1' e '0' em função do tempo, demonstrando a extração do tempo de retenção. A Tabela 5.2 também apresenta os valores extraídos. Verifica-se que, em ambos os casos, houve uma degradação do dado '0' (devido à geração de lacunas) e que a polarização de substrato pulsada apresentou uma melhora no tempo de retenção de aproximadamente 17%. Esta melhora no tempo de retenção pode ser explicada a partir da Figura 5.7, de onde se vê uma menor taxa de geração durante o repouso do estado '0' para o caso pulsado, atrasando a degradação.

Figura 5.6: I_1 e I_0 ao longo do tempo para V_B constante (quadrado tracejado) e pulsada (estrela contínua) e a extração de seu tempo de retenção.

Figura 5.7: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para V_B constante (quadrado tracejado) e pulsada (estrela contínua).

5.2 Influência da banda proibida [53]

Motivado pelas ligas Si-Ge e Si-C que têm sido estudadas, por exemplo, para aplicações opto-eletrônicas e de memórias [54] [55] [56] [57] [58], esta seção analisa a influência da banda proibida (E_g) do filme semiconductor sobre a camada isolante no desempenho da 1T-FBRAM em dispositivos UTBOX SOI a fim de obter tempos de retenção mais longos e, portanto, mais próximos dos apresentados pelas DRAMs atuais [24]. A Figura 5.8 mostra como a banda proibida foi alterada neste trabalho.

Figura 5.8: Diferentes bandas proibidas do filme semiconductor em um dispositivo *Semiconductor-on-Insulator* nMOSFET

O esquema de polarização utilizado segue a sequência apresentada na Figura 5.9. As simulações [40] foram realizadas a 85°C com o substrato positivamente polarizado.

Figura 5.9: Sequência de estados e polarizações utilizadas. O substrato foi polarizado em 2,1V e a uma temperatura de 85°C.

A Figura 5.10 mostra as correntes lidas nos estados ‘1’ e ‘0’ em função do tempo para diferentes bandas proibidas e a Figura 5.11, os respectivos tempos de retenção extraídos.

A principal e mais significativa vantagem obtida no aumento da banda proibida de filme semiconductor sobre o óxido enterrado é o maior tempo de retenção como se pode ver nas Figura 5.10 e Figura 5.11.

Figura 5.10: I_1 e I_0 ao longo do tempo as diferentes bandas proibidas do semiconductor.

Figura 5.11: Tempo de retenção em função da banda proibida.

Da Figura 5.10, um deslocamento no tempo para a degradação do ‘0’ é observado quando a banda proibida é aumentada de 0,7eV para 1,4eV, resultando em uma melhora no tempo de retenção de 5 ordens de grandeza (Figura 5.11).

Isso está relacionado com a menor geração de lacunas para as bandas proibidas maiores. De acordo com a Figura 5.11, a banda proibida deve ser maior que 1,33eV para que se atinja um tempo de retenção maior que as memórias DRAM convencionais [24].

A Figura 5.12 mostra que uma menor taxa de geração é observada para maiores bandas proibidas, confirmando a razão para a menor degradação e, portanto, o maior tempo de retenção.

Figura 5.12: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface para as diferentes bandas proibidas utilizadas.

Verifica-se também que a diferença entre as bandas maiores na taxa de geração diminui. Provavelmente, nesses casos, a polarização necessária para haver o tunelamento banda-a-banda é maior que 2,5V que foi aplicado durante o estado de repouso, havendo outro efeito junto ao BBT. A Figura 5.13 mostra a taxa de geração ao longo do canal considerando ou não os modelos de BBT e TAT (*Trap Assisted Tunneling*). A partir dela, nota-se que, para bandas menores, a taxa de geração se altera em sua maior parte pelo modelo do BBT, enquanto que, para bandas maiores a geração se modifica em sua maioria pela atuação do TAT.

Figura 5.13: Taxa de geração durante o repouso do estado '0' a 1 nm abaixo da 1ª interface considerando ou não, ambos: o modelo de BBT e do TAT.

O TAT é predominante quando o campo elétrico na região de sobreposição da porta sobre o dreno não é tão alto quanto aquele necessário para o BBT, mas é suficiente para ocorrer o tunelamento da banda de valência para a de condução com a ajuda dos defeitos e armadilhas da interface. [40]

6 Conclusões finais e trabalhos futuros

6.1 *Conclusões finais*

Neste trabalho foram propostas várias melhorias de desempenho das células de memória 1T-DRAM em dispositivos UTBOX, bem como a influência de diferentes polarizações e os mecanismos responsáveis por tal comportamento.

Com relação às tendências no estado de leitura, observou-se que a tensão de disparo no dreno, a margem de sensibilidade, o tempo necessário para o disparo e o tempo de retenção diminuem (até 72%, 90%, 95% e 2 ordens de grandeza, respectivamente) com o aumento da temperatura (de 25 a 300°C) devido à redução da mobilidade, ao aumento do ganho ($\beta(M-1)$) e ao aumento da corrente de fuga. Contudo, o dispositivo apresentou desempenho satisfatório mesmo operando em altas temperaturas.

Percebeu-se também a forte influência da espessura da camada de silício e do óxido enterrado na tensão de disparo no dreno. Apesar da menor espessura do filme de silício e do óxido enterrado aumentar a tensão no dreno necessária para ativar o efeito, devido ao maior acoplamento do canal em detrimento da influência do campo elétrico lateral, um potencial positivo no substrato pode reduzir este requisito (61% para tensão do substrato variando de 0 V a 1,5 V), sendo uma boa alternativa para solucionar o problema e permitindo o uso de dispositivos menores como uma célula de memória.

Acerca das polarizações de leitura e de repouso foi visto que, dependendo da polarização de porta utilizada, pode ocorrer uma geração de portadores, degradando o bit '0', ou uma recombinação de portadores, degradando o bit '1'.

Também se verificou que há uma janela de leitura para a polarização do substrato, sendo limitada pelo compromisso da polarização de substrato ser alta o suficiente para ativar o efeito de corpo flutuante, possibilitando a escrita do '1' sem no entanto, ativar o efeito BJT durante a leitura do '0', permitindo o aparecimento dos dois estados.

Além disso, observando-se o comportamento da margem de sensibilidade e do tempo de retenção, concluiu-se que o primeiro parâmetro é menos dependente da tensão de substrato que o segundo, o que nos levou a dar maior atenção na otimização do tempo de retenção.

Já o impacto da tensão no dreno na leitura está bem relacionado com o mecanismo do efeito BJT e o princípio envolvido em seu disparo. Observou-se uma maior margem de sensibilidade (60%) para altas tensões no dreno, resultado da ionização por impacto, ativando o efeito BJT. Por outro lado, isso faz com que mais lacunas sejam injetadas no corpo durante a leitura, diminuindo o tempo de retenção (66%) e limitando o número de leituras possíveis sem atualização do dado (de mais de 30 para 22). Quando uma baixa tensão é aplicada no dreno, verifica-se uma menor margem de sensibilidade, porém, um tempo de retenção maior é observado e mais de 30 leituras pôde ser feita sem necessidade de uma atualização do dado.

Concluiu-se então, que as polarizações ótimas nos diferentes contatos para os diferentes estados são as apresentadas na Tabela 6.1.

Tabela 6.1: Estados e polarizações otimizadas. Escrita '1' por BJT e escrita '0' por acoplamento capacitivo. Temperatura a 85°C.

Estado	V_G [V]	V_D [V]	V_B [V]
Escrita '1'	0	2	2
Escrita '0'	0	0,5	0
Leitura	-1,5	2	0
Repouso	-2,5	0	0

Com respeito à concentração das extensões de fonte e dreno, os resultados indicaram uma diminuição na margem de sensibilidade (16%) e uma maior tensão no dreno necessária para ativar o efeito (aumento de 70%) quando a concentração das extensões é reduzida (de $6 \times 10^{19} \text{cm}^{-3}$ tipo N para $1 \times 10^{15} \text{cm}^{-3}$ tipo P), devido ao menor campo elétrico lateral, resultante do menor potencial efetivamente aplicado ao canal ou ao maior espalhamento deste potencial (no caso do dispositivo sem extensão de fonte e dreno); apesar disso, a tensão de substrato pode ser utilizada para se diminuir esta tensão de dreno.

Os dispositivos sem as extensões de fonte e dreno apresentaram uma taxa de geração de lacunas menor (aproximadamente 12 ordens de grandeza), levando

a um tempo de retenção muito maior (aproximadamente 3 ordens de grandeza) quando comparado ao dispositivo referência. Também se observou um deslocamento da janela de leitura para maiores valores de tensão de porta com a diminuição da concentração da extensão, indicando que para cada concentração, a melhor polarização da porta durante a leitura é diferente de uma estrutura para a outra.

Ainda nos estudos dos dispositivos sem extensão de fonte e dreno, verificou-se a influência da tensão do substrato e do comprimento de canal no tempo de retenção, na janela de leitura e na taxa de geração. O tempo de retenção diminuiu (quase 2 ordens de grandeza) com o comprimento do canal, mesmo com sua otimização por meio da tensão do substrato devido ao maior GIDL e a sua amplificação pela menor base do transistor BJT inerente à estrutura MOS, demonstrando ser um fator limitante para as futuras gerações das memórias 1T-DRAM. Apesar disso, quando comparados com os dispositivos sem extensão de fonte e dreno, seu tempo de retenção aumentou (quase 1 ordem de grandeza) devido à menor taxa de geração, permitindo a utilização de menores comprimentos de canal (redução de 50nm para 30nm) e polarizações de substrato menores (de 2,2V para 1,5V no caso de menor comprimento de canal).

Outra proposta de melhoria no tempo de retenção apresentada foi a utilização vantajosa da polarização de substrato pulsada apenas na escrita pois, mesmo resultando em menor margem de sensibilidade (redução de apenas 4%), o tempo de retenção aumentou em 17% devido ao menor GIDL. Além disso, a menor polarização de substrato durante a leitura reduz o efeito BJT durante a leitura, permitindo um maior número de leituras sem a atualização do dado.

Finalmente estudou-se também a variação da banda proibida motivado pela utilização de novos materiais para a camada semicondutora. Observou-se que o aumento da banda proibida aumentou o tempo de retenção em até 5 ordens de grandeza, possibilitando retenções mais próximas das DRAMs convencionais atuais. Este aumento do tempo de retenção mostrou ser dependente do GIDL e, para as bandas maiores, dependente do TAT.

6.2 *Trabalhos futuros*

Seguindo-se com o tema de se melhorar o tempo de retenção, pode-se ainda estudar o impacto no tempo de retenção dos três temas abordados neste trabalho de forma conjunta, ou seja, estudar o tempo de retenção de uma memória 1T-DRAM em um dispositivo sem extensão de fonte e dreno, cujo filme semiconductor é formado por diferentes materiais (Si-C ou Si-Ge) e utilizando-se o substrato também com polarizações pulsadas.

Publicações

Durante este trabalho foram publicados os seguintes artigos de minha autoria ou coautoria:

Artigos em congresso (12):

- **K.R.A.Sasaki**, T.Nicoletti, J.A.Martino, “Optimized LDD Length in SOI NMOSFET Devices”, Seminatec2011 – VI Workshop on Semiconductors and Micro & Nano Technology;
- L.M.Almeida, **K.R.A.Sasaki**, M.Aoulaiche, E.Simoen, C.Claeys, J.A.Martino, “Analysis of UTBOX 1T-DRAM Memory Cell at High Temperatures”, SBMicro2011 – 26th Symposium on Microelectronics Technology and Devices;
- **K.R.A.Sasaki**, L.M.Almeida, J.A.Martino, M.Aoulaiche, E.Simoen, C.Claeys, “Temperature Influence on UTBOX 1T-DRAM Using GIDL for Writing Operation”, ICCDCS2012 – 8th IEEE International Caribbean Conference on Devices, Circuits and Systems;
- L.M.Almeida, **K.R.A.Sasaki**, M.Aoulaiche, N.Collaert, E.Simoen, C.claeys, J.A.Martino, M.Jurczak, “The Dependence of Sense Margin and Retention Time on Front and Back Gate Bias in UTBOX FBRAM”, EUROSOI2012 – 8th Workshop of the Thematic Network on Silicon on Insulator Technology, Devices and Circuits;
- L.M.Almeida, M.Aoulaiche, **K.R.A.Sasaki**, T.Nicoletti, M.G.C.de Andrade, N.Collaert, E.Simoen, C.Claeys, J.A.Martino, M.Jurczak, “Comparison between high and low read bias in FBRAM on UTBOX device”, ULIS2012 – Ultimate Integration on Silicon;

- **K.R.A.Sasaki**, L.M.Almeida, M.Aoulaiche, E.Simoen, C.Claeys, J.A.Martino, “Comparison of GIDL and BJT effect for Writing Operation in UTBOX 1T-DRAM at high temperatures”, Seminatec2012 – VII Workshop on Semiconductors and Micro & Nano Technology;
- **K.R.A.Sasaki**, L.M.Almeida, J.A.Martino, “Impact of the Extension Region Concentration on the UTBOX 1T-FBRAM”, SBMicro2012 – 27th Symposium on Microelectronics Technology and Devices;
- L.A.Pereira, **K.R.A.Sasaki**, T.Nicoletti, J.A. Martino, “Influência da espessura do filme de silício na tensão de limiar em transistores UTBOX FDSOI”, SICT2012 – 14th Simpósio de Iniciação Científica e Tecnológica;
- T.Nicoletti, **K.R.A.Sasaki**, M. Aoulaiche, E. Simoen, C.Claeys, J.A.Martino, “Experimental and Simulation of 1T-DRAM Trend with the Gate Length on UTBOX Devices”, EUROSOI 2013 – 9th Workshop of the Thematic Network on Silicon on Insulator Technology, Devices and Circuits; (aprovado)
- T.Nicoletti, **K.R.A.Sasaki**, S.D. dos Santos, J.A.Martino, M. Aoulaiche, E. Simoen, C.Claeys, “The Generation Rate Analysis of Different S/D Junction Engineering in Scaled UTBOX 1T-DRAM”, 223rd ECS Meeting 2013 (Canadá) – The Electrochemical Society Meeting (aprovado);
- **K.R.A.Sasaki**, A.Nissimoff, L.M.Almeida, M.Aoulaiche, E.Simoen, C.Claeys, J.A.Martino, “Improvement of Retention Time Using Pulsed Back Gate Bias on UTBOX SOI Memory Cell”, EUROSOI 2013 – 9th Workshop of the Thematic Network on Silicon on Insulator Technology, Devices and Circuits; (aprovado)
- **K.R.A.Sasaki**, A.Nissimoff, L.M.Almeida, M.Aoulaiche, E.Simoen, C.Claeys, J.A.Martino, “Semiconductor Film Bandgap Influence on Retention Time of UTBOX SOI 1T-FBRAM”, 223rd ECS Meeting 2013 (Canadá) – The Electrochemical Society Meeting (aprovado);

Artigos em revista (2):

- L.M.Almeida, **K.R.A.Sasaki**, M.Aoulaiche, E.Simoen, C.Claeys, J.A.Martino, “One Transistor Floating Body RAM Performances on UTBOX Devices Using the BJT Effect”, JICS – Journal of Integrated Circuits and Systems;
- L.M.Almeida, **K.R.A.Sasaki**, M.Aoulaiche, N.Collaert, E.Simoen, C.Claeys, J.A.Martino, M.Jurczak, “Optimizing the front and back biases for the best sense margin and retention time in UTBOX FBRAM”, Solid State Electronics (aprovado);

Referências Bibliográficas

1. MOORE, G. E. Progress in Digital Integrated Electronics. **Intel**, 1975.
Disponível em: <download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore/1975_Speech.pdf>. Acesso em: 26 Abril 2012.

2. CHANG, L. et al. Moore's law lives on CMOS Transistor. **IEEE Circuits & Devices Magazine**, p. 35-42, 2003.

3. NICOLETTI, T. et al. **The Impact of the Gate Length Scaling on UTBOX FDSOI Devices**: the Digital/Analog Performance of Extension-less Structures. Proceedings of 13th Ultimate Integration on Silicon. Grenoble: [s.n.]. 2012.

4. SANTOS, S.; NICOLETTI, T.; MARTINO, J. **Analog performance of Gate-Source/Drain Underlap Triple Gate SOI nMOSFET**. Microelectronics Technology and Devices - SBMicro 2011. Pennington, NJ, USA: [s.n.]. 2011.

5. SONG, K.-W. et al. **55nm Capacitorless 1T-DRAM Cell Transistor with Non-overlap Structure**. IEEE International Electron Device Meeting. Washington, DC: [s.n.]. 2008.

6. NICOLETT, A. S. et al. Extraction of the lightly doped drain concentration of fully depleted SOI NMOSFETs using the back gate bias effect. **Solid-State electronics**, v. 44, p. 677-684, 2000.

7. NICOLETTI, T. **Estudo da Resistência Série de Fonte e Dreno de Transistores SOI FinFETs de Porta Tripla e com Canal Tensionado (Dissertação de Mestrado)**. São Paulo. 2009.

8. DIXIT, A. et al. Analysis of the Parasitic S/D Resistance in Multiple-Gate FETs. **IEEE Transactions on Electron Devices**, v. 52, n. 6, p. 1132-1140, 2005.

9. HU, G. J.; CHANG, C.; CHIA, Y.-T. Gate-Voltage-Dependent Effective Channel Length and Series Resistance of LDD MOSFET's. **IEEE Transactions on Electron Devices**, v. 34, n. 12, p. 2469-2475, 1987.

10. TERADA, K.; MUTA, H. A New Method to Determine the Effective MOSFET Channel Length. **Japanese Journal of Applied Physics**, v. 18, n. 5, p. 953-

- 959, 1979.
11. COLINGE, J. P. **FinFETs and Other Multi-Gate Transistors**. New York: Springer, 2008.
 12. COLINGE, J.-P. **Silicon-on-Insulator Technology: Materials to VLSI**. New York: Springer, 2004.
 13. FUJIWARA, M. et al. **Impact of BOX Scaling on 30 nm Gate Length FD SOI MOSFETs**. Proceedings of 2005 IEEE International SOI Conference. Honolulu, Hawaii: [s.n.]. 2005.
 14. YAN, R. et al. **LDD Depletion Effects in Thin-BOX FDSOI Devices with a Ground Plane**. Proceedings of 2009 IEEE International SOI Conference. Foster City, California: [s.n.]. 2009. p. 1-2.
 15. FENOUILLET-BERANGER, C. et al. **Impact of a 10nm Ultra-Thin BOX (UTBOX) and Gound Plane on FDSOI Devices for 32nm node and below**. Proceedings of the European Solid State Device Research Conference, 2009. Athens: [s.n.]. 2009.
 16. FENOUILLET-BERANGER, C. et al. FDSOI devices with thin BOX and ground plane integration for 32 nm node and below. **Soilid State Electronics**, v. 53, p. 730-734, 2009.
 17. SASAKI, N. et al. **Charge Pumping SOS-MOS Transistor Memory**. International Electron Devices Meeting Technical Digest. Washington, D.C.: [s.n.]. 1978.
 18. OKHONIN, S. et al. **New Generation of Z-RAM**. Electron Devices Meeting, 2007. IEDM 2007. IEEE International. Washington, DC: [s.n.]. 10-12 Dezembro 2007. p. 925-928.
 19. FISCH, D. **Innovative Approach to Drive Floating Body Z-RAM (R) Embedded Memory to 32 nm and Beyond**. 2007 IEEE international SOI Conference Proceedings. Indian Wells, CA: [s.n.]. 2007.
 20. JURCZAK, M. Memories on SOI: Floating Body Cell Memory. **Site da EUROSIOI**, 17 janeiro 2011. Disponivel em: <http://www.eurosoi.org/public/T5_EUROSIOI2011_MJurczak.pdf>. Acesso em:

26 Setembro 2011.

21. BAWEDIN, M. et al. **Floating-Body SOI Memory: Concepts, Physics and Challenges.** ECS Transactions. San Francisco, California: The Electrochemical Society. 2009. p. 243-256.
22. CHOI, J.-Y.; FOSSUM, J. G. Analysis and Control of Floating-Body Bipolar Effects in Fully Depleted Submicrometer SOI MOSFET's. **IEEE Transactions on Electron Devices**, Junho 1991. 1384-1391.
23. AVCI, U. E. et al. **Floating Body Cell (FBC) Memory for 16 nm technology with low variation on thin silicon and 10 nm BOX.** 2008 IEEE International SOI Conference. New York: [s.n.]. 2008.
24. ITRS. International Technology Roadmap for Semiconductor, 2011. Disponível em: <www.itrs.net>. Acesso em: 30 outubro 2012.
25. YOSHIDA, E.; TANAKA, T. A Capacitorless 1T-DRAM technology using gate induced drain leakage (GIDL) current for low power and high speed embedded memory. **IEEE Transactions on Electron Devices**, 4 Abril 2006. 692-697.
26. LIM, H.-K.; FOSSUM, J. G. Threshold Voltage of Thin Film Silicon on Insulator (SOI) MOSFET's. **IEEE Transactions on Electron Devices**, v. 30, n. 10, p. 1244-1251, Outubro 1983.
27. COLLAERT, N. et al. **Analysis of sense margin and reliability of 1T-DRAM fabricated on thin-film UTBOX substrates.** IEEE International SOI Conference. Foster City, CA: [s.n.]. 2009.
28. OHTOU, T.; SARAYA, T.; HIRAMOTO, T. Variable-Body-Factor SOI MOSFET with Ultrathin Buried Oxide for Adaptive Threshold Voltage and Leakage Control. **IEEE Transactions on Electron Devices**, Janeiro 2008. 40-47.
29. ALMEIDA, L. M. et al. **Analysis of UTBOX 1T-DRAM Memory Cell at High Temperatures.** Microelectronics Technology and Devices. [S.l.]: ECS Transactions. 2011. p. 61-68.
30. ALMEIDA, L. M. et al. One Transistor Floating Body RAM Performances on UTBOX Devices Using the BJT Effect. **Journal of Integrated Circuits and**

- Systems**, São Paulo, 2012. 113-119.
31. KONONCHUK, O. **Precise Oxide Dissolution**. US2010/0193899 A1, 5 Agosto 2010.
 32. MONFRAY, S. et al. Emerging silicon-on-nothing (SON) devices technology. **Solid-State Electronics**, v. 48, p. 887-895, 2004.
 33. SEDRA, A. S.; SMITH, K. C. **Microeletrônica**. 5. ed. São Paulo: Pearson Prentice Hall, 2007.
 34. MOON, D.-I. et al. Fin-Width Dependence of BJT-Based 1T-DRAM Implemented on FinFET. **IEEE Electron Device Letters**, v. 31, n. 9, p. 909-911, 2010.
 35. COLLAERT, N. et al. **Comparison of scaled floating body RAM architectures**. 2008 IEEE International SOI Conference Proceedings. Tempe, Arizona: [s.n.]. 2008. p. 35-36.
 36. LU, Z. et al. **A Novel Low-Voltage Biasing Scheme for Double Gate FBC Achieving 5s Retention and 10^{16} Endurance at 85°C**. Electrical Devices Meeting (IEDM). San Francisco, CA: [s.n.]. 2010.
 37. SASAKI, K. R. A. et al. **Temperature Influence on UTBOX 1T-DRAM Using GIDL for Writing Operation**. Proceeding of 8th International Caribbean Conference on Devices, Circuits and Systems. Playa del Carmen, Mexico: [s.n.]. 2012.
 38. COLINGE, J. P.; COLINGE, C. A. **Physics of Semiconductor Devices**. New York: Springer, 2006.
 39. STREETMAN, B. G.; BANERJEE, S. K. **Solid State Electronic Devices - Sixth Edition**. New Jersey: Prentice Hall Inc., 2006.
 40. SILVACO INTERNATIONAL. **ATLAS Device Simulation - User's Manual, version 5.14.0.R**. Santa Clara. 2010.
 41. ALMEIDA, L. M. et al. **Comparison between low and high read bias in FB-RAM on UTBOX FDSOI Devices**. Proceedings of 13rd Ultimate Integration on

- Silicon. Grenoble: [s.n.]. 2012.
42. ALMEIDA, L. M. et al. **The Dependence of Sense Margin and Retention Time on Front and Back Gate Bias in UTBOX FBRAM**. Proceedings of 8th Workshop of the Thematic Network on Silicon on Insulator Technology, Devices and Circuits. Montpellier: [s.n.]. 2012. p. 23-24.
 43. ALMEIDA, L. M. et al. Optimizing the front and back biases for the best sense margin and retention time in UTBOX FBRAM. **Solid State Electronics**, 2013.
 44. FOSSUM, J. G. et al. **Physical Insights on Design and Modeling of Nanoscale FinFETs**. IEEE International Electron Devices Meeting. Washington, DC: [s.n.]. 2003.
 45. NICOLETTI, T. et al. The Dependence of Retention Time on Gate Length in UTBOX FBRAM with Different Source/Drain Junction Engineering. **IEEE Electron Device Letters**, Julho 2012. 940-942.
 46. SZE, S. M. **Physics of Semiconductor Devices**. New Jersey: John Wiley & Sons Eds., 1981.
 47. NICOLETTI, T. et al. **Experimental and Simulation of 1T-DRAM Trend with the Gate Length on UTBOX Devices**. Proceedings of 9th Workshop of the Thematic Network on Silicon on Insulator Technology, Devices and Circuits. Paris, França: [s.n.]. 2013.
 48. NICOLETTI, T. et al. **Gate Length Impact on UTBOX FBRAM Devices**. Proceedings of IEEE International SOI Conference 2012. Napa, California: [s.n.]. 2012. p. 66-67.
 49. SASAKI, K. R. A.; ALMEIDA, L. M.; MARTINO, J. A. **Impact of the Extension Region Concentration on the UTBOX 1T-FBRAM**. Microelectronics Technology and Devices. [S.l.]: ECS Transactions. 2012. p. 281-287.
 50. NICOLETTI, T. et al. **The Generation Rate Analysis of Different S/D Junction Engineering in Scaled UTBOX 1T-DRAM**. ECS Transactions. [S.l.]: [s.n.]. 2013.
 51. SASAKI, K. R. A. et al. **Improvement of Retention Time Using Pulsed Back Gate Bias on UTBOX SOI Memory Cell**. Proceedings of 9th Workshop of the

Thematic Network on Silicon on Insulator Technology, Devices and Circuits.
Paris, França: [s.n.]. 2013.

52. IKEDA, K. et al. **First Demonstration of Threshold Voltage Control by Sub-1V Back-gate Biasing for Thin Body and Buried-Oxide (TBB) Ge-on-Insulator (GOI) MOSFETs for Low Power Operation**. Proceedings of IEEE International SOI Conference. Napa, California: [s.n.]. 2012. p. 162-163.

53. SASAKI, K. R. A. et al. **Semiconductor Film Band Gap Influence on Retention Time of UTBOX SOI 1T-FBRAM**. ECS Transactions. [S.l.]: [s.n.]. 2013.

54. SOREF, R. A. **Silicon-Based Optoelectronics**. Proceedings of IEEE. [S.l.]: [s.n.]. 1993. p. 1687-1706.

55. GIORGIS, F. et al. Wide band gap a-SiC:H films for optoelectronic applications. **Journal of Non-Crystalline Solids**, 1998. 465-469.

56. TANG, P.; HUANG, R.; WU, D. Performance Improvement of Capacitorless Dynamic Random Access Memory Cell with Band Gap Engineered Source and Drain. **Japanese Journal of Applied Physics**, 2010. 04DD02-1 - 04DD02-3.

57. HAN, J.-W. et al. **Energy Band Engineered Unified-RAM (URAM) for Multi-Functioning 1T-DRAM and NVM**. IEEE International Electron Device Meeting. San Francisco, CA: [s.n.]. 2008. p. 227-230.

58. SHIN, J. S. et al. A Novel Double HBT-Based Capacitorless 1T-DRAM Cell with Si/SiGe Heterojunctions. **IEEE Electron Device Letters**, 2011. 850-852.

Anexo 1 – Estrutura gerada no ATHENA

```

"go athena

#####
# Definição da grade#
#####

line x loc=0 spac=0.02 tag=esq
line x loc=0.05 spac=0.007
line x loc=0.09 spac=0.005
line x loc=0.13 spac=0.003
line x loc=0.23 spac=0.003
line x loc=0.27 spac=0.005
line x loc=0.31 spac=0.007
line x loc=0.36 spac=0.02 tag=dir

line y loc=0 spac=0.001 tag=cima
line y loc=0.01 spac=0.005
line y loc=0.02 spac=0.001 tag=cimaux
line y loc=0.025 spac=0.05
line y loc=0.03 spac=0.001 tag=baixox
line y loc=0.37 spac=0.01
line y loc=0.53 spac=0.05 tag=baixo

#####
#Definição da lâmina#
#####

# Regioes
region silicon xlo=esq xhi=dir ylo=cima yhi=baixo

region silicon xlo=esq xhi=dir ylo=cima yhi=baixox
region oxide xlo=esq xhi=dir ylo=cimaux yhi=baixox

# Concentracao inicial e orientacao cristalografica da lamina
init boron=1e15 orient=100

#####
#Definicao da regioa ativa - os isolantes não foram considerados#
#####

#####
#Ground Plane#
#####

implant boron energy=40 dose=1e13 tilt=0 gauss

implant boron energy=60 dose=4e13 tilt=0 gauss

diffuse temp=1000 time=0.5 seconds nitrogen

#####
#Oxido e metal de porta#
#####

deposit oxide thick=0.005 div=4

deposit polysilicon thick=0.1 div=10

# obs.: o athena não tem o TiN, o SiON nem o HfO2, então depositou-se SiO2 e
# Si-poli para depois alterar suas WFs.

#structure outfile=etapa2_gatestack_rev02.str two.dim

#####
#Litografia - porta#
#####

```

```

#abertura da fonte
etch polysilicon start x=0 y=-0.105
etch continue x=0.13 y=-0.105
etch continue x=0.13 y=-0.005
etch done x=0 y=-0.005

#abertura do dreno
etch polysilicon start x=0.23 y=-0.105
etch continue x=0.36 y=-0.105
etch continue x=0.36 y=-0.005
etch done x=0.23 y=-0.005

#structure outf=etapa3_openedSD_rev02.str two.dim

#####
#LDD + Spacer#
#####

#dopagem do LDD
deposit oxide thick=0.005 div=4
implant arsenic energy=5 dose=1e15 tilt=0 gauss

#spacer
deposit nitride thick=0.035 div=4
etch nitride start x=0 y=-0.05
etch continue x=0.09 y=-0.05
etch continue x=0.09 y=0.005
etch done x=0 y=0.005

etch nitride start x=0.27 y=-0.05
etch continue x=0.36 y=-0.05
etch continue x=0.36 y=0
etch done x=0.27 y=0

etch nitride start x=0.09 y=-0.15
etch continue x=0.27 y=-0.15
etch continue x=0.27 y=-0.105
etch done x=0.09 y=-0.105

etch oxide start x=0 y=-0.05
etch continue x=0.09 y=-0.05
etch continue x=0.09 y=0
etch done x=0 y=0

etch oxide start x=0.27 y=-0.05
etch continue x=0.36 y=-0.05
etch continue x=0.36 y=0
etch done x=0.27 y=0

etch oxide start x=0.13 y=-0.11
etch continue x=0.23 y=-0.11
etch continue x=0.23 y=-0.105
etch done x=0.13 y=-0.105

#structure outf=etapa4_spacer_rev02.str

#####
#SEG#
#####

epitaxy temp=750 time=20 thick=0.030

#####
# remoção dos excessos #
#####

etch silicon start x=0.05 y=-0.18

```

```

etch continue x=0.31 y=-0.18
etch continue x=0.31 y=-0.03
etch done x=0.05 y=-0.03

#####
#HDD#
#####

implant arsenic energy=18 dose=3e15 tilt=0 gauss
implant phosphorus energy=8 dose=2e15 tilt=0 gauss

#structure outf=etapa6_HDD_rev02.str

#####
# Spike Annealing #
#####

diffuse temp=1050 time=0.001 SECONDS inert

#####
# Final Sintering #
#####

#diffuse temp=420 time=20 MINUTES F.H2=1 F.N2=1

#####
#Definicao dos contatos - fazer no ATLAS#
#####

structure outf=UTBOX_ATHENAS_BOX10_TSI20_L100_rev03a.str

quit

”

```

Anexo 2 – Estrutura gerada no ATLAS

```

#####
# Simulacao de um Transistor SOI nMOSFET SOI UTBOX #
# L=150 nm tsi=50nm tox=5nm tbox=30nm #
#####
go atlas

#####
# Construcao da Grade #
#####

mesh    space.mult=1.0

x.mesh  loc=0.00 spac=0.05
x.mesh  loc=0.04 spac=0.01
x.mesh  loc=0.08 spac=0.001
x.mesh  loc=0.23 spac=0.001
x.mesh  loc=0.27 spac=0.01
x.mesh  loc=0.31 spac=0.05

y.mesh  loc=-0.005 spac=0.005
y.mesh  loc=0.00 spac=0.001
y.mesh  loc=0.05 spac=0.003
y.mesh  loc=0.08 spac=0.01

#####
# Definicao das Regioes da Estrutura #
#####

region      num=1 y.min=-0.005 y.max=0 oxide
region      num=2 y.min=0 y.max=0.05 silicon
region      num=3 y.min=0.05 y.max=0.08 oxide

#####
# Definicao dos Eletrodos do dispositivo #
#####

# #1-GATE #2-SOURCE #3-DRAIN #4-SUBSTRATE

electrode    name=gate x.min=0.08 x.max=0.23 y.min=-0.005 y.max=-
0.005
electrode    name=source x.min=0.0 x.max=0.0 y.min=0.0 y.max=0.05
electrode    name=drain x.min=0.31 x.max=0.31 y.min=0.0 y.max=0.05
electrode    name=substrate bottom

#####
# Definicao das Dopagens #
#####

# Source
doping       uniform n.type conc=1e20 region=2 x.l=0.0 x.r=0.04
#LDD Source
doping       uniform n.type conc=1e19 region=2 x.l=0.04 x.r=0.08
# Channel
doping       uniform p.type conc=1e15 region=2 x.l=0.08 x.r=0.23
#LDD Drain
doping       uniform n.type conc=1e19 region=2 x.l=0.23 x.r=0.27
# Drain
doping       uniform n.type conc=1e20 region=2 x.l=0.27 x.r=0.31

```

```
#####
# Especificando as Propriedades de Interface #
#####

interf qf=1e10 y.max=0.001
interf qf=1e10 y.min=0.049

#####
# Configuracao da Funcao Trabalho da Porta #
#####

contact name=gate          workfunc=4.7
contact name=substrate     workfunc=5.2

#####
# Salva a estrutura #
#####

save outf=UTBOX_BOX30_TSI50.str

#####
# FIM #
#####
quit"
```

Anexo 3 – Simulação de uma curva dinâmica

```

#####
go atlas simflags="-P 32"
#####
mesh infile=UTBOX_ATHENAS_BOX10_TSI20_L100_DOSE0.str
#####

#####
#Definicao dos eletrodos#
#####

# 1-GATE      2-SOURCE      3-DRAIN      4-SUBSTRATE
electrode      name=gate      x.min=0.13 x.max=0.23 y.min=-0.005 y.max=-
0.005
electrode      name=source x.min=0 x.max=0.05 y.min=-0.03 y.max=-0.03
electrode      name=drain x.min=0.31 x.max=0.36 y.min=-0.03 y.max=-
0.03
electrode      name=substrate bottom

#####
# Especificando as Propriedades de Interface e workfunction#
#####

interf qf=6e11 y.max=0.001
interf qf=6e11 y.min=0.019
trap region=1 e.level=0.3 acceptor density=2e15 degen=1 sign=1e-11
sigp=1e-19
trap region=1 e.level=0.3 donor density=1e15 degen=1 sign=5e-18
sigp=5e-11
contact name=gate workfunc=4.7

#####
# Modelos #
#####

MODELS region=1 srh consrh cvt bbt.kane trap.coulombic print
temp=383

MOBILITY REGION=1 MUMAXN.CVT = 360
MOBILITY REGION=1 MUMAXP.CVT = 120
MOBILITY REGION=1 GAMN.CVT = 1.0

MODEL BB.GAMMA=1.2

IMPACT region=1 SELB
IMPACT region=1 BETAN=4.5
IMPACT region=1 BETAP=4.5

IMPACT region=1 AN1 = 1406e2
IMPACT region=1 AN2 = 1406e2
IMPACT region=1 AP1 = 1342e2
IMPACT region=1 AP2 = 3160e2

#MATERIAL TAUN0=1e-7
#MATERIAL TAUP0=1e-7

```



```
#####
# Metodos + Polarizações Iniciais #
#####

METHOD  gummel newton autonr bicgst trap maxtrap=10 carriers=1
solve init

#####
solve prev
solve vgate=0
solve vdrain=0
solve vsource=0
solve vsubstrate=0

solve vsubstrate=1e-3
solve vsubstrate=10e-3
solve vsubstrate=25e-3
solve vsubstrate=50e-3
solve vsubstrate=100e-3
solve vsubstrate=200e-3
solve vsubstrate=500e-3
solve vsubstrate=800e-3
solve vsubstrate=1.2
solve vsubstrate=1.5
solve vsubstrate=2.1

solve vgate=-1e-3
solve vgate=-10e-3
solve vgate=-25e-3
solve vgate=-50e-3
solve vgate=-100e-3
solve vgate=-200e-3
solve vgate=-250e-3
solve vgate=-500e-3
solve vgate=-800e-3
solve vgate=-1
solve vgate=-1.2
solve vgate=-1.5
solve vgate=-1.9
solve vgate=-2.5

#####
# Metodos + Levantamento dos Pulsos #
#####

METHOD  gummel newton autonr bicgst trap maxtrap=10 carriers=2

log outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5.log master

# 1a parte do grafico (repouso)
solve vgate=-2.5 vdrain=0 ramptime=1e-9 tstop=18e-9 timestep=1e-11

# 2a parte do grafico (escrita '1')
solve vgate=0 vdrain=3.355 ramptime=1e-9 tstop=26e-9 timestep=1e-11
```

```

save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_W1.str
master

# 3a parte do grafico (escrita '1' continuacao)
solve vgate=0 vdrain=3.355 tstop=28e-9 timestep=1e-11

# 4a parte do grafico (repouso)
solve vgate=-2.5 vdrain=0 ramptime=1e-9 tstop=30e-9 timestep=1e-11
save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_pW1.str
master

# 5a parte do grafico (repouso continuacao)
solve vgate=-2.5 vdrain=0 tstop=48e-9 timestep=1e-11

# 6a parte do grafico (leitura)
solve vgate=-1.5 vdrain=3.355 ramptime=1e-9 tstop=56e-9 timestep=1e-11
save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_R1.str
master

# 7a parte do grafico (leitura continuacao)
solve vgate=-1.5 vdrain=3.355 tstop=58e-9 timestep=1e-11

# 8a parte do grafico (repouso)
solve vgate=-2.5 vdrain=0 ramptime=1e-9 tstop=60e-9 timestep=1e-11
#save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_pR1.str
master

# 9a parte do grafico (repouso continuacao)
solve vgate=-2.5 vdrain=0 tstop=78e-9 timestep=1e-11

# 10a parte do grafico (escrita '0')
solve vdrain=0.5 vgate=0 ramptime=1e-9 tstop=86e-9 timestep=1e-11
save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_W0.str
master

# 11a parte do grafico (escrita '0' continuacao)
solve vdrain=0.5 vgate=0 tstop=88e-9 timestep=1e-11

# 12a parte do grafico (repouso)
solve vgate=-2.5 vdrain=0 ramptime=1e-9 tstop=90e-9 timestep=1e-11
save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_pW0.str
master

# 13a parte do grafico (repouso continuacao)
solve vgate=-2.5 vdrain=0 tstop=108e-9 timestep=1e-11

# 14a parte do grafico (leitura)
solve vgate=-1.5 vdrain=3.355 ramptime=1e-9 tstop=116e-9 timestep=1e-11
save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_R0.str
master

# 15a parte do grafico (leitura continuacao)
solve vgate=-1.5 vdrain=3.355 tstop=118e-9 timestep=1e-11

# 16a parte do grafico (repouso)

```

```
solve vgate=-2.5 vdrain=0 ramptime=1e-9 tstop=120e-9 tstep=1e-11
#save outf= UTBOX_ATHENA_BOX10_TSI20_L100_DOSE0_VGr-1v5_pR0.str
master
```

```
# 17a parte do grafico (repouso continuacao)
solve vgate=-2.5 vdrain=0 tstop=128e-9 tstep=1e-11
```

```
#####
###                                FIM                                ###
#####
quit
```

```
”
```