

Arthur Denicol Ceratti

*Desenvolvimento de um sensor "On-Chip" para
Monitoramento do Envelhecimento de SRAMs*

Porto Alegre

2012

Arthur Denicol Ceratti

*Desenvolvimento de um sensor "On-Chip" para
Monitoramento do Envelhecimento de SRAMs*

Dissertação apresentada ao Programa de Pós-Graduação de Engenharia Elétrica, da Faculdade de Engenharia da Pontifícia Universidade Católica do Rio Grande do Sul, como requisito parcial obtenção do título de Mestre em Engenharia Elétrica.

Orientador:

Prof. Dr. Fabian Luis Vargas

Co-orientador:

Prof.^a Dr.^a Letícia Maria Bolzani Poehls

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Porto Alegre

2012



DESENVOLVIMENTO DE UM SENSOR "ON-CHIP" PARA MONITORAMENTO DO ENVELHECIMENTO DE SRAMS

CANDIDATO: ARTHUR DENICOL CERATTI

Esta Dissertação de Mestrado foi julgada para obtenção do título de MESTRE EM ENGENHARIA ELÉTRICA e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Pontifícia Universidade Católica do Rio Grande do Sul.



DR. FABIAN LUIS VARGAS - ORIENTADOR



DRA. LETÍCIA MARIA BOLZANI POEHLS - CO-ORIENTADORA

BANCA EXAMINADORA



**DR. EDUARDO AUGUSTO BEZERRA - DEPARTAMENTO DE ENGENHARIA ELÉTRICA -
UFSC**



DR. RUBEM DUTRA RIBEIRO FAGUNDES - PPGE - FENG - PUCRS

RESUMO

A miniaturização da tecnologia *Complementary Metal-Oxide Semiconductor* (CMOS) tornou possível a integração de milhões de transistores em um único Circuito Integrado (CI) aumentando assim, a densidade dos mesmos. Em mais detalhes, essa miniaturização resultou em significativos avanços tecnológicos devido fundamentalmente à diminuição do *delay* do transistor o que, por sua vez, acarretou no aumento da performance dos CIs devido ao aumento na frequência de operação dos mesmos. Além disso, o aumento no nível de integração dos CIs possibilitou o desenvolvimento de CIs capazes de agregarem um número cada vez maior de funções aumentando significativamente a complexidade dos mesmos. Em paralelo, o rápido aumento na necessidade de armazenar um volume cada vez maior de informação resultou no fato de que *Static Random Access Memories* (SRAMs) ocupam hoje grande parte da área de silício de um *System-on-Chip* (SoC). A *SIA Roadmap* prevê que em 10 anos cerca de 94% da área de um SoC será dedicada à memória [1]. Entretanto, essa miniaturização gerou vários problemas, relacionados à confiabilidade, que devem ser afrontados através do uso de diferentes técnicas que visam à otimização de CIs. Neste contexto, é importante salientar o fenômeno conhecido com *Negative Bias Temperature Instability* (NBTI) que afeta a confiabilidade do CI em longo prazo, ou seja, durante a sua vida útil. Especificamente em SRAMs o NBTI provoca a degradação da *Static Noise Margin* (SNM) o que, por sua vez afeta a capacidade de armazenamento das células de memória. Neste contexto, esta dissertação de mestrado tem como principal objetivo a especificação, implementação, validação e avaliação de uma metodologia baseada em hardware para o monitoramento do nível de envelhecimento de células de SRAMs a fim de garantir a confiabilidade das mesmas durante a sua vida útil. A metodologia proposta consiste na inserção de um sensor capaz de monitorar o consumo de potência dinâmica das células durante as operações de escrita a fim de compará-los com os valores definidos como padrão para uma célula não envelhecida. Finalmente, a metodologia proposta será validada funcionalmente e sua eficiência será avaliada a partir da análise da sua capacidade de monitoramento e detecção bem como, a partir dos *overheads* de área, performance e imunidade a variabilidade do processo de fabricação.

ABSTRACT

Advances in Complementary Metal-Oxide Semiconductor (CMOS) technology have made possible the integration of millions of transistors into a small area, allowing the increase of circuits' density. In more detail, technology scaling caused the reduction of the transistors' delay, which has resulted in a significantly performance improvement of Integrated Circuits (ICs). Furthermore, the increase in the integration level of ICs allowed the development of ICs able to include an increasing number of functions, which in turn increased significantly their complexity. In parallel, the rapidly increasing need to store more information results in the fact that the Static Random Access Memory (SRAM) can occupy great part of the System-on-Chip (SoC) silicon area. This is confirmed by the SIA Roadmap which forecasts a memory density approaching 94% of the SoC area in about 10 years [1]. Consequently, memory has become the main responsible of the overall SoC area.

However, the reduction of transistor size has introduced several reliability concerns that need to be affronted by the adoption of different optimization techniques. In this context it is important to highlight the phenomenon known as Negative Bias Temperature Instability (NBTI), which affects the reliability of the ICs along their lifes. Specifically in the SRAMs NBTI causes degradation of the Static Noise Margim(SNM) which affects the storage capacity of the memory cells.

In this context, the main goal of this thesis is to specify, implement, validate and evaluate a hardware-based technique able to monitor the aging of SRAM cells in order to guarantee their reliability of during the lifetime. The proposed technique is based on an on-chip sensor capable of monitoring dynamic power consumption of the cells during write operations in order to compare them with the value set as default to a new cell. Finally, the proposed methodology has been functionally validated and its efficiency has been evaluated based on the analysis of its monitoring and detection capabilities and from the analysis of the introduced overheads as well as its immunity to the manufacturing process variation.

LISTA DE FIGURAS

1	Matriz de células.	20
2	Célula SRAM tipo 6T.	21
3	Processo de envelhecimento [2].	23
4	Processo de recuperação [3].	24
5	Esquema simplificado da SNM [4].	27
6	Modelamento gráfico da SNM [4].	28
7	Modelo elétrico da SNM [4].	29
8	SNM de armazenamento para uma célula nova.	30
9	SNM de leitura para uma célula nova.	30
10	Célula de memória assimétrica [5].	32
11	Teste da capacidade de condução dos pMOS [6].	34
12	Trasnsformação da tensão dos <i>bitlines</i> em corrente [6].	35
13	Comparação da corrente encontrada com a corrente de referência [6].	35
14	Corrente de <i>leakage</i> de uma SRAM x idade [7].	36
15	Fluxograma da Técnica Apresentada em [8]	38
16	Redução das correntes de alimentação com o envelhecimento	42
17	Colunas de células com transistores nas linhas de alimentação.	43

18	Tensão na linha de alimentação antes e durante a fase de detecção do envelhecimento.	44
19	Curvas de tensão da capacitância de alimentação vdd0 vs. idade.	44
20	SNM vs. idade.	45
21	Esquemático do sensor.	46
22	Diagrama de blocos da proposta.	47
23	Circuito comparador de tensão.	48
24	Circuito para determinação do limiar de envelhecimento.	51
25	Modelamento do envelhecimento [9]	54
26	Fluxograma de validação da técnica	56
27	Curva de tensão de uma célula em função do envelhecimento	57
28	SNM de uma célula com envelhecimento não equilibrado	59
29	Determinação da corrente média durante uma operação de escrita	61
30	Curvas de tensão que são comparadas com a referência	62
31	Saída do sensor para as curvas da figura 27	62
32	Distribuição uniforme para a entrada de diferentes idades	64
33	Distribuição gaussiana do envelhecimento entre os pMOS de cada célula	65
34	Gráfico dos resultados das simulações de Monte Carlo para todos os 5 <i>corners</i> simulados	67
35	Gráfico separando os tipos de falha	68
36	Tensão de Detecção x Distribuição do Envelhecimento.	69
37	Diferença de tensão entre idades vs. Número de operações de escrita.	71

LISTA DE TABELAS

1	Comparativo Entre as Diferentes Técnicas Apresentadas	39
2	Correntes Médias nas Operações de Escrita para Diferentes Idades	60
3	Capacidade de detecção do sensor para todos os <i>corner</i> utilizados	67
4	Funcionamento do Sensor Considerando Diferentes Variações de Processos.	68
5	Custo de Área Associado ao Sensor	70
6	Tempo de Detecção	72
7	Consumo de Energia	72

GLOSSÁRIO

BIST Built In Self Test

CI Circuitos Integrados

CMOS Complementary Metal-Oxide Semiconductor

HCI Hot Carrier Injection

I_{ddq} Stand By Leakage Current

MOSFET Metal Oxide Semiconductor Field Effect Transistor

NBTI Negative Bias Temperature Instability

SNM Signal Noise Margin

SoC System-on-Chip

SRAM Static Random Access Memory

TDD Time Dependent Dielectric Breakdown

SUMÁRIO

1	Introdução	13
1.1	Motivação	15
1.2	Objetivos	17
1.3	Apresentação dos Capítulos	18
2	Fundamentos Teóricos	19
2.1	Funcionamento de uma SRAM	20
2.2	O Fenômeno de <i>Negative Bias Temperature Instability</i> (NBTI)	22
2.3	Medida de Qualidade do Funcionamento de uma Célula	26
2.4	Estado-da-Arte: Técnicas de Detecção do Envelhecimento	31
2.4.1	Técnica Baseada em Célula Assimétrica	31
2.4.2	Técnica Baseada na Capacidade de Condução dos transistores pMOS	33
2.4.3	Técnica Baseada na Corrente de <i>Leakage</i>	35
2.4.4	Técnica Baseada na Alocação de Dados Visando Diminuição do Estresse Sob as Células	37
2.4.5	Estudo Comparativo Entre as Técnicas	38
3	Proposta	40
3.1	Princípio Fundamental	41

3.2	Arquitetura Proposta	45
3.2.1	Comparador de Tensão	47
3.2.1.1	Geração da Tensão de Referência	48
3.2.2	Determinação dos Parâmetros de Configuração do Sensor	49
3.2.2.1	Circuito para Determinação do Limiar de Envelhecimento	50
4	Validação	52
4.1	Estudo de Caso 2	52
4.2	Caracterização dos Efeitos Fenômeno de NBTI em uma Célula SRAM	53
4.3	Fluxo	55
4.3.1	Etapa 1 - Definição dos Parâmetros do Sensor	56
4.3.2	Etapa 2 - Considerações Quanto à SNM	58
4.3.3	Etapa 3 - Calibração do Sensor	58
4.3.4	Etapa 4 - Acionamento do Sensor e Detecção	61
5	Avaliação	63
5.1	Simulações por Monte Carlo	63
5.1.1	Simulações por Monte Carlo considerando variações de processos	65
5.2	Capacidade de Monitoramento	69
5.3	<i>Overhead</i>	69
5.3.1	Área	70
5.3.2	Performance	70
5.3.3	Energia	72

6 Conclusão	73
7 Trabalhos Futuros	75
Referências	76
Apêndice:	
Apêndice A – Exemplo de Estimação do Valor de SNM para Uma Célula SRAM	79
Apêndice B – Modelamento Hspice da Simulação de Monte Carlo	83
Apêndice C – Publicações Realizadas	85

1 INTRODUÇÃO

A constante miniaturização da tecnologia tornou possível o aumento significativo do número de transistores por Circuito Integrado (CI). Essa redução no tamanho dos transistores, tornou possível uma série de avanços tecnológicos dentre os quais salientam-se o aumento da frequência de operação ocasionada pela redução no *delay* dos transistores e o aumento significativo do número de funções agregadas em um mesmo CI devido fundamentalmente ao incremento do nível de integração dos mesmos. Em paralelo, o crescente aumento da necessidade de armazenar um volume cada vez maior de informações fez com que a área de silício dedicada à *Static Random Access Memories* (SRAMs) em *System-on-Chip* (SoC) passasse a representar cerca 90% do CI. Conseqüentemente, a memória passou a ser o principal responsável pela área total do SoC o que, por sua vez, aumentou ainda mais a necessidade em agregar novas técnicas capazes de garantirem o nível de confiabilidade e robustez desejado para as mesmas.

Entretanto, apesar desse contínuo aumento na performance dos CIs, uma série de desafios foram introduzidos no que diz respeito fundamentalmente a confiabilidade e a robustez dos mesmos o que, por sua vez tornou-se uma das principais preocupações durante o seu desenvolvimento. Neste contexto, fenômenos físicos como *Hot Carrier Injection* (HCI), *Time Dependent Dielectric Breakdown* (TDDB) e *Negative Bias Temperature Instability* (NBTI) destacam-se como fontes capazes de afetarem a confiabilidade de CIs [10] [11]. Esses fenômenos físico/químicos causam a degradação do óxido e resultam no deslocamento da tensão de *threshold* do transistor ao longo do tempo. Entretanto, NBTI afeta os transistores do tipo pMOS e atualmente representa um dos principal problema

de confiabilidade durante a vida útil do CI.

NBTI ocorre quando o transistor pMOS é negativamente polarizado e manifesta-se como o aumento na tensão de *threshold* no decorrer do tempo. Esse efeito resulta na redução da corrente o que, por sua vez causa a degradação do *delay* do dispositivo. Em geral, a tensão de *threshold* aumenta de 5 a 15% por ano dependendo da tecnologia e do ambiente e conseqüentemente, o *delay* segue o mesmo comportamento. Assim, o fenômeno de NBTI afeta a estabilidade das memórias e se traduz na sua incapacidade de armazenar os dados da maneira adequada. Essa propriedade é caracterizada a partir do parâmetro definido como *Static Noise Margin* (SNM) que, por sua vez, torna-se menor à medida que as células de SRAM perdem sua capacidade de armazenamento. Em mais detalhes, o principal impacto em uma célula SRAM do tipo 6T submetida ao fenômeno de NBTI é a perda a longo prazo de sua estabilidade [2], diferentemente daquilo que ocorreria no caso de um CI de lógica combinacional, por exemplo. Nesse caso, o principal problema seria relacionado ao tempo, pois este envelhecimento resultaria em um maior tempo para o transistor sair de uma condição "cortada" para conduzindo [12].

Neste contexto, este trabalho de mestrado tem como principal objetivo a especificação, implementação, validação e avaliação de uma técnica baseada em hardware capaz de monitorar o nível de envelhecimento de SRAMs. Em mais detalhes, o trabalho proposto visa o desenvolvimento de um sensor *on-chip* capaz de monitorar e detectar o envelhecimento de células de SRAM através da medição da queda da capacidade de condução dos pMOS em função do envelhecimento. Este trabalho foi realizado através da combinação de conceitos retirados de artigos e livros, bem como através de análises realizadas através do uso do simulador elétrico HSPICE. Afim de que se obtivesse uma aproximação maior dos dados obtidos com a realidade, foram utilizadas bibliotecas comerciais fornecidas pela empresa STMicroelectronics de 65nm. Através destas bibliotecas foram construídos blocos de SRAMs capazes de simular de forma bastante verossímil o funcionamento e os efeitos encontrados em memórias submetidas ao NBTI. Finalmente, a técnica proposta para monitorar o nível de envelhecimento da SRAM foi avaliada funcionalmente levando-se em

consideração os *overheads* de área e performance, a sua robustez diante da variabilidade do processo de fabricação, bem como no que diz respeito a sua capacidade de monitoramento e detecção.

1.1 Motivação

A principal motivação para a realização deste trabalho de mestrado concentra-se no fato de que foi observado através do estudo dos trabalhos realizados com o intuito de lidar com o problema do envelhecimento gerado fundamentalmente pelo fenômeno do NBIT, que alguns pontos relacionados aos mesmos poderiam ser melhorados sobremaneira. Dentre os principais pontos, destacam-se aqueles relacionados à referência utilizada para determinação de envelhecimento frente às variações do processo de fabricação e pontos como a perda da informação armazenada nas células que fossem ser submetidas aos testes para a determinação da idade. Além disso, salienta-se o problema de se determinar a idade de uma célula especificamente e não de um grupo, onde se necessita trabalhar com a questão de granularidade para ter uma "ideia" da idade de uma célula, e enfim da própria complexidade do projeto necessário para a detecção do envelhecimento.

Os pontos observados que podiam ser melhorados podem ser analisados, conforme se pode perceber, por exemplo, no trabalho apresentado em [13], onde foi apresentado o benefício de se utilizar técnicas de *power-gating* para diminuir o consumo de blocos lógicos enquanto em *idle* e ao mesmo tempo proporcionar aos transistores do tipo pMOS um prolongamento de sua vida útil em função de sua capacidade de recuperação frente ao NBTI realizada com o auxílio dos *sleeping transistors*, que ao seccionarem o referencial dos blocos inativos, permitiam que aos poucos todos os transistores das células inativas fossem se movendo para um nível lógico 1, ou seja, sem a "referência" as correntes de *leakage* acabam por equalizar cada um dos nós da célula até que todos fiquem carregados com "1 lógico", isto acaba por criar as condições necessárias para a recuperação dos pMOS [12]. Porém estas abordagens tem apenas o efeito de diminuir o avanço do envelhecimento, não

proporcionando qualquer controle relativo deterioração dos transistores, determinação de suas idades.

Em [8] foi utilizada uma função em software capaz de distribuir o uso das células de memória de forma mais homogênea, desta forma buscou-se distribuir também os efeitos de envelhecimento forçando todo o bloco de memória a envelhecer de forma igual, e assim aumentando a vida útil do bloco ao evitar que pequenos grupos de células fossem estressados de forma desigual e assim envelhecidos precocemente. Esta abordagem já garante maior controle sobre o envelhecimento e esta direcionada especificamente para uso em memórias, porém ainda não realiza a predição do envelhecimento, apenas visa um aproveitamento melhor das células de memória.

Já no trabalho apresentado em [6] através de uma alteração no *buffer* de escrita, tentou-se avaliar o envelhecimento dos transistores pMOS em função da queda de sua capacidade de condução, monitorando assim a passagem de corrente da fonte de tensão através dos pMOS até os *bitlines*. Porém neste trabalho usou-se uma referência fixa como forma de mensurar o envelhecimento, ainda que tenha sido salientado o uso de uma referência conservadora, devemos observar que o impacto das variáveis de processo e de temperatura podem acabar criando resultados algumas vezes muito diferentes daqueles obtidos nas condições de teste.

Desta forma, buscou-se então propor uma nova metodologia baseada em hardware que consiste na inserção de um sensor capaz de monitorar o nível de envelhecimento de SRAMs, célula a célula, sem que houvesse significativo incremento de área ou consumo de potencia ao bloco de memória e também se valendo de uma referência, para mensurar o envelhecimento, capaz de se ajustar automaticamente às variações do processo de fabricação de uma forma mais eficiente do que aquela que seria feita por uma referência de valor fixo.

Neste contexto, este trabalho de mestrado visa concentrar-se no fenômeno de NBTI, pois esse é tido como um dos fenômenos que mais causam falhas em memórias [12],

utilizando seus efeitos como forma de se determinar a idade de uma célula de memória e por consequência sua robustez.

1.2 Objetivos

A proposta em questão visa contornar alguns problemas que ainda não haviam sido devidamente tratados, utilizando-se de vários dos conceitos até agora observados para tentar monitorar o envelhecimento em células SRAM de forma a criar condições para se aumentar o grau de robustez de sistemas baseados nestas.

Técnicas como *power-gating*, que se baseiam em seccionar a referência ou "*ground*" dos circuitos, de forma a diminuir a corrente de *leakage*, se mostram uma forma eficiente de realizar economia de energia sobretudo em sistemas embarcados nos quais este é um problema cada dia mais difícil de ser resolvido, visto que a capacidade de armazenamento dos acumuladores de energia tem se desenvolvido muito mais vagarosamente do que a demanda crescente criada pelos novos CI's, e além da economia de energia, conforme visto em [13] também tem a capacidade de estender a vida útil dos transistores frente ao NBTI conforme será visto à seguir.

Assim, a técnica baseada em hardware proposta neste trabalho baseia-se no desenvolvimento de um sensor de baixo consumo de potência, imune a determinadas variações do processo de fabricação bem como, capaz de suportar variações na temperatura durante o seu funcionamento capaz de monitorar o envelhecimento das SRAMs.

O sensor desenvolvido é capaz de realizar o monitoramento e detecção do nível de envelhecimento da SRAM durante a vida útil do SoC. A fim de simplificar sua fabricação, optou-se por desenvolver um projeto com baixa complexidade e, para utilizá-lo em campo, também se tomou o cuidado de que o mesmo tivesse baixo consumo de potência. Por fim, tendo em vista que este sensor é apenas uma ferramenta e não a aplicação final tentou-se utilizar a menor quantidade de área possível em sua construção.

1.3 Apresentação dos Capítulos

O manuscrito desta dissertação de mestrado foi organizado em 6 grandes partes, são elas:

Capítulo 2 - Fundamentos Teóricos. Neste capítulo serão abordados os principais conceitos para o entendimento desta dissertação.

Capítulo 3 - Proposta. No capítulo 3 serão apresentados os pontos que se desejam alcançar com o desenvolvimento do trabalho.

Capítulo 4 - Implementação de Estudo de Caso. Neste momento será exposta a forma necessária para a implementação de um sensor deste tipo através do uso de um estudo de caso.

Capítulo 5 - Avaliação da Técnica Proposta. Neste capítulo serão abordados os resultados encontrados através da implementação do sensor.

Capítulo 6 - Conclusão. Aqui será feita uma análise rápida de tudo que foi visto nesta dissertação.

Capítulo 7 - Trabalhos Futuros. No capítulo 7 serão expostos alguns pontos que podem ser contornados através do prosseguimento dos estudos acerca desta dissertação.

2 *FUNDAMENTOS TEÓRICOS*

Esta dissertação de mestrado tem como principal objetivo a especificação, implementação, validação e avaliação de uma técnica baseada em hardware capaz de monitorar o nível de envelhecimento de células de SRAMs visando o aumento na confiabilidade e robustez de sistemas.

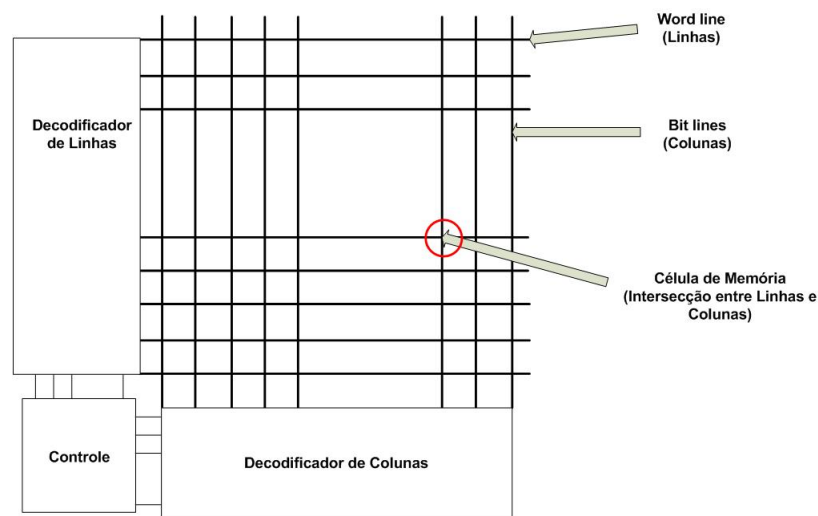
Deseja-se detectar o envelhecimento em células de SRAM ocasionado pelo fenômeno conhecido como *Negative Biased Temperature Instability* (NBTI) através da observação das consequências deste nestas células e mensurar a robustez destas células através de uma ferramenta conhecida como *Static Noise Margin* (SNM).

Tais conceitos, afim de que se possa melhor compreender a proposta do trabalho, devem ser primeiramente introduzidos e devidamente definidos. Desta forma, os próximos parágrafos apresentam uma introdução sobre Memórias Estáticas de Acesso Randômico (*Static Random Access Memories* - SRAMs), as quais são o alvo para a utilização deste sensor, logo em seguida será apresentado o fenômeno conhecido como NBTI que está relacionado ao envelhecimento sendo através do monitoramento deste a forma escolhida para detecção do envelhecimento. Ainda dentro deste capítulo, será apresentada a métrica utilizada para se avaliar a robustez das células de memória, serão citados alguns dos principais trabalhos relacionados à este problema e que serviram também de base para o desenvolvimento desta dissertação.

2.1 Funcionamento de uma SRAM

SRAMs são assim conhecidas por permitirem o acesso de qualquer bit de dado em qualquer instante de tempo e por manterem o dado armazenado indefinidamente [14]. Para tanto, SRAMs utilizam decodificadores de linha e de coluna, uma vez que os blocos de células de memória são organizados de forma matricial conforme a figura 1.

Figura 1 – Matriz de células.

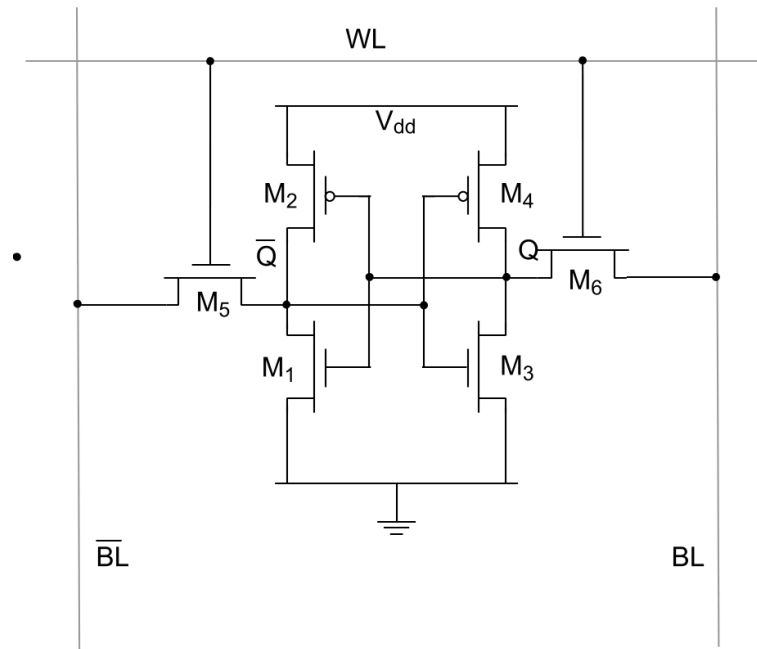


A intersecção entre as linhas e as colunas corresponde a posição de cada uma das células. Conectados aos *bitlines* (colunas), estão 4 blocos operacionais da matriz de células, são eles: (1) amplificador de sinal, (2) circuito de pré-carga, (3) circuito de escrita e (4) circuito de leitura.

Especificamente neste trabalho de mestrado, estão sendo utilizadas as do tipo 6T. Essas células recebem este nome porque são formadas por seis transistores conforme a figura 2.

Pode-se observar que uma célula 6T, conforme a figura 2, é composta por dois inversores conectados de forma cruzada, são formados pelos transistores M1, M2, M3 e M4. A célula possui ainda dois transistores que são chamados de transistores de acesso, M5 e M6, que são acionados à todo momento que se deseja realizar uma escrita ou uma leitura na célula conectando-a aos *bitlines* (BL e BL barrado). O valor armazenado pela célula é

Figura 2 – Célula SRAM tipo 6T.



a diferença de tensão entre as duas saídas dos inversores Q e Q barrado.

Uma das informações mais importantes à respeito deste tipo de célula é a relação entre o tamanho dos transistores tipo N de cada inversor (*driver*) e o respectivo transistor tipo N que o conecta ao *bitline* (acesso). Esta relação está intimamente ligada à robustez da célula. Se os transistores de acesso forem muito fracos existirá uma dificuldade em se realizar uma troca de valor na célula em uma operação de escrita, por outro lado, se eles forem muito fortes existirá primeiramente um grande aumento de área, e ao mesmo tempo como existe uma relação entre o transistor de acesso e o "transistor *driver*" (que juntos formam um divisor de tensão) existirá uma queda na robustez da célula.

A forma de organização das células é matricial, e estas matrizes de células formam bloco, que são compostos por trilhas de transferência de dados chamadas de *bitlines*, organizadas em colunas responsáveis por determinar o dado que será lido ou escrito em uma célula e *wordlines*, que são barramentos de controle de células, organizados em linhas e realizam a conexão de uma célula para escrita ou para leitura aos *bitlines*. Os blocos lógicos são conectados às colunas, cada coluna portanto possui um circuito de escrita, que recebe o dado à ser escrito do processador e cria uma diferença de tensão nos *bitlines*

da coluna onde esta a célula que será escrita. Existe ainda um circuito de leitura, que é habilitado para detectar diferenças de potencial nos *bitlines* (diferenças estas criadas pela célula que será lida). Um amplificador de sinal, que serve para reforçar tanto aquilo que se deseja escrever em uma célula, quanto aquilo que se deseja ler com a ajuda de transistores mais fortes e um circuito de pré-carga, que é responsável por "zerar" as diferenças de tensão presentes nos *bitlines* tanto para realizar as operações de escrita quanto para realizar as operações de leitura.

Para escrever na célula primeiramente é habilitado o circuito de pré-carga, que acaba com qualquer diferença de tensão existente nos *bitlines*, colocando ambos em um nível de "ativo alto". O dado é então recebido pelo circuito de escrita que cria uma pequena diferença de tensão entre um *bitline* e o outro, logo em seguida o amplificador de sinal reforça esta diferença de tensão e o sinal de *wordline* é habilitado inserindo estas diferenças de tensão na célula (escrevendo o dado).

A operação de leitura também se inicia com a habilitação do circuito de pré-carga, logo em seguida o sinal de *wordline* conecta a célula aos *bitlines* que então passam a apresentar uma pequena diferença de tensão, que é também reforçada pelo amplificador de sinal dando condições para que o circuito de leitura realize a leitura do valor que estava armazenado na célula [14].

De todas estas operações, aquela que acaba por ser considerada a mais crítica, quando se leva em conta o envelhecimento, é a de leitura. Por este motivo é através desta operação que se medirá o envelhecimento da célula [2].

2.2 O Fenômeno de *Negative Bias Temperature Instability* (NBTI)

Existem uma série de fenômenos responsáveis por causarem a deterioração dos transistores com o passar do tempo, são eles: *Channel Hot Carriers* (CHCs), que é definido como a "injeção" de elétrons ou lacunas no *gate* de *mosfets*, *Time-Dependent Dielectric Break-*

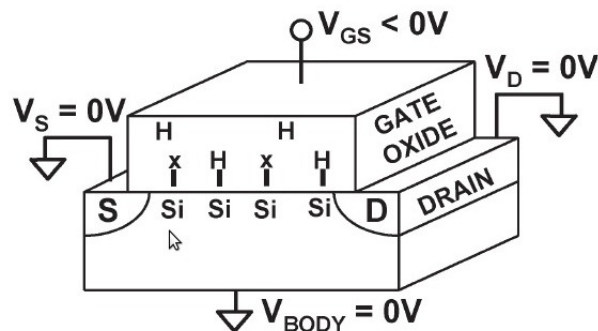
down (TDDB) que é definido como a formação de um túnel, um caminho com condução, entre o óxido do *gate* e o substrato e Negative Bias Temperature Instability (NBTI).

NBTI é conhecido desde 1966 e aparece como um dos maiores problemas ligados ao envelhecimento de memórias construídas com tecnologias que envolvem alta integração de circuitos [15].

Analisando o comportamento físico do fenômeno, verificou-se em [2] e [16] que a degradação de um transistor por NBTI é presumivelmente resultado da desestruturação das ligações químicas de Si-H durante sua fase de condução (quando é excitado com uma tensão negativa) na interface entre Si e SiO_2 (onde Si é o que compõe o corpo do transistor e SiO_2 constitui a camada isolante entre o corpo e seu gate) [17].

Convencionalmente é aplicada uma camada de hidrogênio na superfície de silício após o processo de oxidação, e são estes átomos que acabam sendo perdidos durante a fase de operação do transistor conforme figura 3, nela pode-se ver um transistor pMOS sendo polarizado negativamente (acionado) nela se percebem algumas ligações Si-H ainda intactas enquanto outras já apresentam degradação, tendo os locais onde antes haviam átomos de hidrogênio preenchidos com lacunas x .

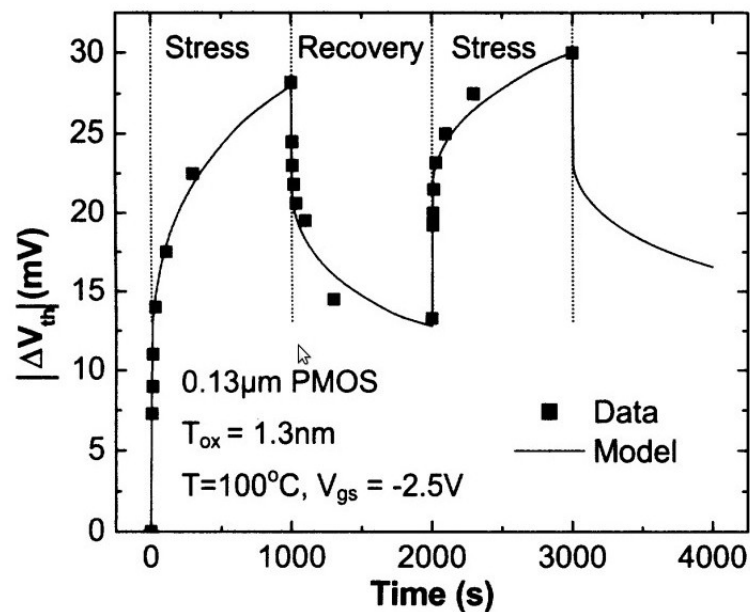
Figura 3 – Processo de envelhecimento [2].



Quando ocorre esta desestruturação, as lacunas resultantes no silício acabam por criar uma carga positiva, o problema se manifesta, então como um incremento do valor da tensão de *threshold* de transistores do tipo pMOS durante o tempo em que o transistor é estressado com um valor negativo de tensão aplicado em seu *gate* (V_{GS} negativo),

porém enquanto o transistor não está conduzindo ocorre uma fase de recuperação [18], com a volta dos átomos de hidrogênio para seus locais de origem através de um processo conhecido como *annealing* que nada mais é do que a recuperação das ligações de Si-H através da recuperação de átomos de Hidrogênios livres, fazendo com que o valor de deslocamento do *threshold* seja reduzido parcialmente, aproximando-se do valor original, conforme figura 4, cabe no entanto salientar que alguns autores percebem este processo de recuperação como sendo total, porém a ampla maioria entende que este processo de recuperação é apenas parcial criando um resíduo que pouco à pouco acaba por limitar a capacidade de condução do transistor.

Figura 4 – Processo de recuperação [3].



Conforme visto em [19], dentre vários modelos presentes na literatura que tentam explicar o NBTI, apenas o modelo conhecido como "Reação - Difusão" (R-D) apresenta consistência entre o modelo de degradação do *mosfet* e as dependências das leis de potência.

Cabe salientar ainda que este modelo é capaz de também explicar a fase de *annealing*, ou recuperação, onde ocorre a volta de parte dos átomos de hidrogênio que se soltaram durante a fase de estresse.

De acordo com o modelo, no momento que um *mosfet* tipo P é acionado (energizado com um valor negativo de tensão entre seu *gate e source*) ocorre uma interação entre as

ligações químicas de Si-H e as lacunas do Silício (N_o - densidade inicial de lacunas na interface Si/Oxido), que quando expostas à temperaturas elevadas levam a dissociação dos átomos de hidrogênio resultando na formação de novas lacunas (N_{it} - lacunas na interface Si/Óxido). Esta formação de lacunas esta ligada à duas constantes, kf (que mede como se da a dissociação das ligações Si-H) e kr (que responde pela taxa com que ocorre o *annealing*, retorno dos átomos de hidrogênio perdidos).

O modelo R-D pode ser então explicado pelas equações abaixo:

$$\frac{dN_{it}}{dT} = kf \cdot [N_o - N_{it}] - kr \cdot N_{it} \cdot NH^0 \quad (2.1)$$

$$\frac{dN_{it}}{dT} = D \frac{dN_{it}}{dx} \Big|_{x=0} + \frac{\delta}{2} \cdot \frac{dNH}{dT} \quad (2.2)$$

Onde $NH(0) \equiv NH(x = 0)$ é obtido através da integração da equação padrão de difusão do hidrogênio através da interface Si/ SiO_2 de espessura δ . Conforme [19] a equação funciona da seguinte forma, ao colocar o transistor sob estresse, a constante kf presente na equação 2.1 é finita e inicia-se desta forma a criação de lacunas e quando o transistor é desligado kf vai à zero e kr é finita iniciando-se assim o processo de recuperação (*annealing*).

A fase de estresse é caracterizada pela rápida formação de lacunas (N_{it}) na interface de Si/ SiO_2 , esta formação de lacunas é apenas limitada pela capacidade de dissociação das ligações de SiH, logo após essa fase de rápida formação de lacunas ocorre uma desaceleração na reação em função da limitação na geração de H^2 .

Já na fase de recuperação, existe uma rápida recuperação das moléculas de H^2 próximas a interface, e é esta rápida recuperação das moléculas de H^2 (*annealing*) que também é responsável pela formação de "buracos de difusão (*diffusion holes*)". Em mais detalhes são esses "buracos" os responsáveis por, no momento que for reaplicado o estresse, criarem uma perda de H^2 mais rápida que aquele que ocorreu anteriormente na próxima vez que

o transistor for estressado. Desta forma se assume que a recuperação dos transistores é apenas parcial e não total como alguns autores sugerem.

Observando-se o comportamento do fenômeno operando em DC, verifica-se uma degradação na ordem de 10% à 15% da tensão de *threshold* de um pMOS para cada ano ao qual foi submetido ao NBTI [20]. Quando se analisa uma célula de SRAM este número depende ainda do valor do dado que esta sendo armazenado pela célula, a probabilidade da célula estar armazenando 0 ou 1 é dada por um coeficiente β . Uma célula que armazena na metade do tempo o valor lógico 0 e outra metade o valor 1 possui um $\beta = 0,5$ e cada um dos pMOS é submetido à um deslocamento da tensão de *threshold* de 5% à 7,5% dependendo da tecnologia utilizada [20].

A equação 2.3 apresenta k como sendo um resumo de todas as características da tecnologia β , conforme foi explicado acima, como sendo referente à quantidade de tempo em que o transistor esta sendo estressado ou em fase de recuperação, t é a temperatura em que se esta trabalhando e o expoente $1/4$ se refere ao átomo de hidrogênio que é perdido durante a fase de estresse, diferenciando assim da perda de uma molécula de hidrogênio [20].

$$dV_{Th} = K.(\beta.t)^{1/4} \quad (2.3)$$

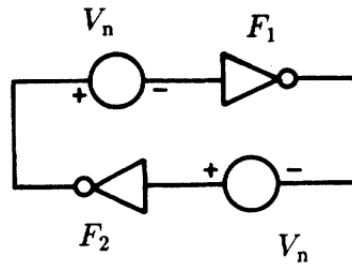
Esta relação entre o uso da célula e o envelhecimento é muito importante, pois quanto mais equilibrada for a distribuição do envelhecimento entre os pMOS da célula, maior será sua vida útil, pois sua SNM tenderá a manter seu valor por mais tempo [21].

2.3 Medida de Qualidade do Funcionamento de uma Célula

A robustez de uma célula pode ser medida através do uso de uma métrica denominada *Static Noise Margin* (SNM). A SNM é definida como o valor relativo ao máximo ruído

DC que pode ser tolerado por uma célula antes que esta tenha seu valor alterado [4]. Simplificadamente, SNM pode ser representado pela figura 5. Em mais detalhes a figura mostra dois inversores que compõem uma célula SRAM ($F1$ e $F2$) e também as duas fontes de tensão responsáveis por representar o ruído DC (V_n).

Figura 5 – Esquema simplificado da SNM [4].



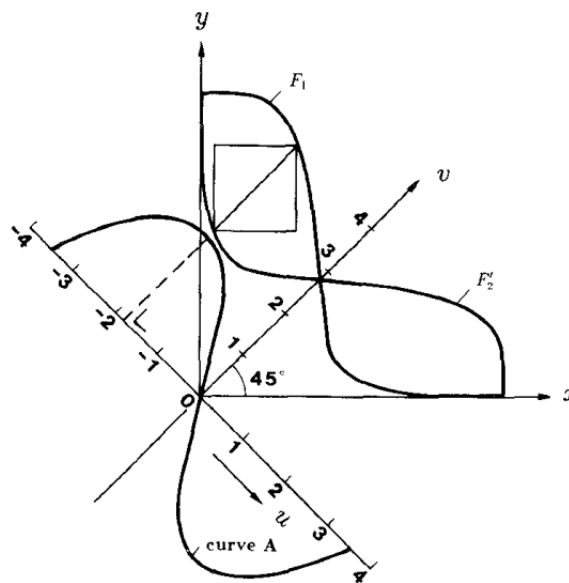
Existem três tipos de SNM, são eles: *write*, *read* e *hold* [22] [23]. A SNM de leitura (*read*) e de armazenamento (*hold*) são medidas em Volts e se referem ao máximo ruído DC que uma célula de memória é capaz de suportar antes de inverter o valor lógico por ela armazenado [4]. Já a SNM de escrita (*write*) se refere ao valor mínimo de tensão em um *bitline* capaz de inverter o dado de uma célula na fase de escrita [22]. Os tipos de SNM afetados pelo fenômeno de NBTI são o SNM de leitura e o de armazenamento, o SNM de escrita por outro lado, é melhorado a medida que a célula envelhece. Este melhoramento ocorre devido ao fato de que o pMOS envelhecido torna-se muito mais rápido durante a mudança do estado conduzindo para o estado cortado.

A SNM pode ser modelada de forma analítica ou gráfica. O modelamento através do modo gráfico pode ser obtido através das curvas características de tensão dos dois inversores. Este modelamento leva em conta uma inversão de dado realizada dentro da célula. Nesta inversão, em que um inversor vai de ativo baixo à ativo alto enquanto o outro faz o caminho contrário, o inversor que aqui será chamado de $F1$ é aquele que irá sair de ativo alto para baixo enquanto o $F2$ é o que fará o caminho contrário. A curva de cada um dos inversores então é colocada sobreposta, porém a curva do inversor $F2$ tem seus eixos trocados, ficando o eixo relativo à tensão de saída na horizontal e o de entrada na vertical conforme pode-se observar na figura 6, nesta figura, $F1$ aparece com

sua tensão de entrada em x e saída em y enquanto F_2 é rebatido tornando-se F_2' e tem suas tensões de entrada em y e saída em x .

Neste gráfico é possível observar ainda que as curvas dos inversores formam dois espaços, normalmente chamados de "olhos". Através do método gráfico deve-se encaixar um quadrado dentro de cada um destes "olhos". O menor destes quadrados terá então seu lado medido e este valor será a própria SNM.

Figura 6 – Modelamento gráfico da SNM [4].



Um modelamento elétrico baseado no modelo gráfico é proposto em [24]. Neste modelo elétrico assume-se que o inversor F_1 é modelado pela função $y = F_1(x)$ e que o inversor F_2' é modelado por $y = F_2'(x)$. Para que se possa encontrar então $F_1(x)$ no sistema de coordenadas (u,v) , primeiramente deve-se transporta-lo para este novo sistema, isto é realizado através das equações 2.4 e 2.5.

$$x = \frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v \quad (2.4)$$

$$y = -\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v \quad (2.5)$$

Logo após a substituição das equações 2.4 e 2.5 na equação $y = F_1(x)$ chega-se à

equação:

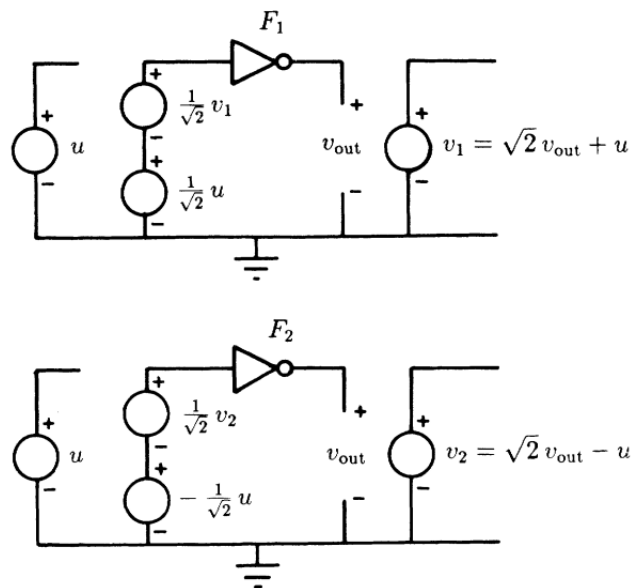
$$v = u + \sqrt{2}F_1\left(\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v\right) \quad (2.6)$$

Para encontrar F_2' , primeiramente F_2 é rebatido no eixo v (conforme a imagem 6) e então transportado para o sistema de coordenadas (u,v) . Após esta operação, realiza-se então o mesmo que foi feito com $F_1(x)$ tendo-se então a equação 2.7:

$$v = -u + \sqrt{2}F_2\left(-\frac{1}{\sqrt{2}}u + \frac{1}{\sqrt{2}}v\right) \quad (2.7)$$

O circuito presente na figura 7 é construído através destas duas equações. Neste trabalho este circuito foi simulado via HSPICE para que assim obtivéssemos a informação sobre a SNM das células que estão sendo testadas. O circuito acima no entanto se refere à uma SNM relativa a fase de armazenamento da célula, para realizar a SNM na fase de leitura foram ainda colocados dois transistores do tipo nMOS com suas "fontes" conectadas às saídas dos inversores de cada circuito com suas "portas" sempre ligadas e seus drenos em duas fontes também sempre ligadas que simulam os *bitlines*.

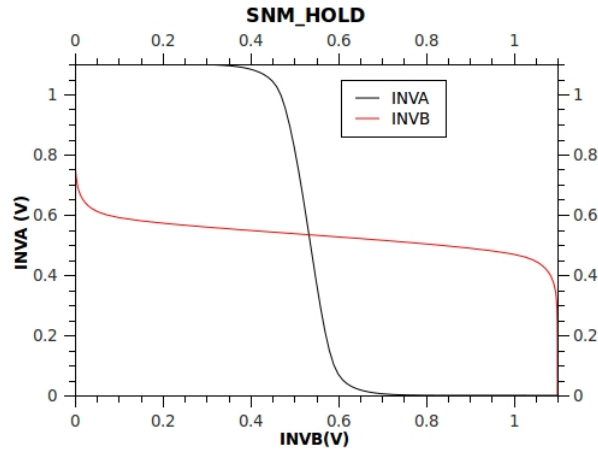
Figura 7 – Modelo elétrico da SNM [4].



A partir do circuito que foi implementado no HSPICE obteve-se então as imagens

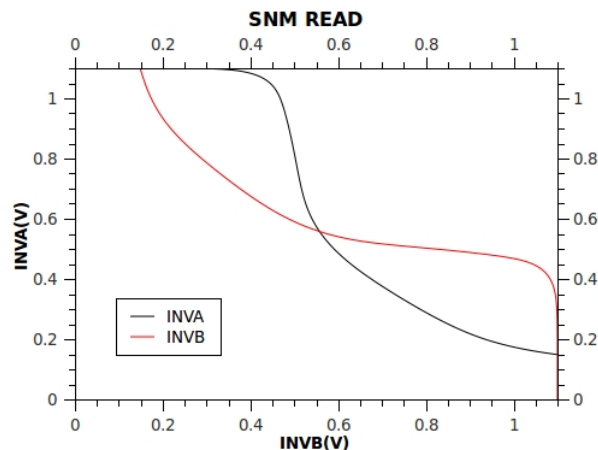
abaixo, primeiro de uma SNM de armazenamento ou *Hold*, onde é possível verificar um grande espaço que é formado entre as curvas dos inversores e logo após uma SNM de leitura onde existe a redução deste espaço, demonstrando assim sua menor capacidade de suportar ruídos.

Figura 8 – SNM de armazenamento para uma célula nova.



A SNM de leitura ou *read* é portanto aquela onde existe a maior chance de um ruído ser capaz de inverter o conteúdo armazenado em uma célula [25]. Uma falha deste tipo originaria uma leitura com valor oposto àquele armazenado.

Figura 9 – SNM de leitura para uma célula nova.



A SNM de leitura leva em consideração a descarga do *bit line* que esta no lado onde a célula esta armazenando ativo baixo, esta descarga irá criar uma corrente elétrica fluindo do *bitline* para a massa, que deverá obrigatoriamente passar através de um dos transistores de acesso e pelo transistor nMOS (*driver*) de um dos inversores. Portanto, a relação de

tamanho entre o transistor de acesso e o nMOS irá determinar qual a SNM máxima de *read* que uma determinada célula poderá ter [25].

Conforme a descrição presente em [26], quando se analisa o efeito do envelhecimento causado pelo NBTI em uma célula, devemos lembrar que é através do transistor pMOS, do braço contrário àquele onde está o nMOS que será colocado em condução que virá o nível lógico para seu acionamento e posterior descarga do *bit line*.

Logo, se o pMOS estiver envelhecido, não terá força suficiente para colocar o transistor nMOS do inversor contrário em condução da forma como colocaria em uma célula nova, criando assim uma resistência no nMOS que, por sua vez irá resultar em um aumento da tensão deste nó. Em outras palavras, a tensão entre o nó que armazena o valor lógico "um" e aquele que está armazenando "zero", irá cair. Esta queda é exatamente aquilo que se busca analisar através da SNM, pois quanto menor a diferença entre as tensões dos nós menor o ruído necessário para alterar seus estados lógicos.

Portanto, diferentemente de um processador onde o NBTI resulta em um aumento do atraso de processamento [27], para uma SRAM ele significa queda da instabilidade da célula.

2.4 Estado-da-Arte: Técnicas de Detecção do Envelhecimento

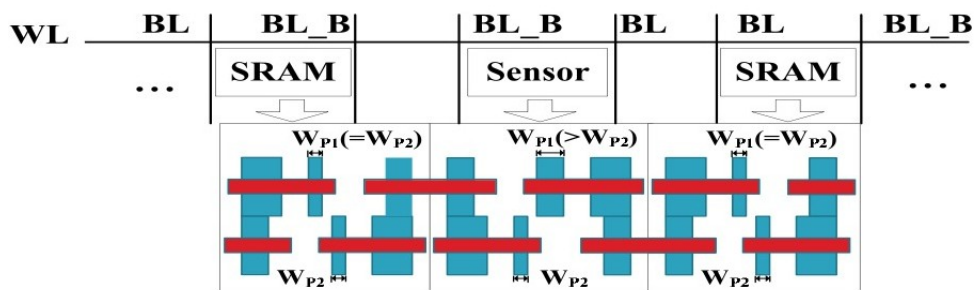
Nesta seção são apresentadas três maneiras distintas que foram utilizadas para se tentar determinar o envelhecimento de uma célula de SRAM ou pelo menos para a determinação de qual a confiabilidade que ainda se pode esperar de uma célula submetida ao envelhecimento.

2.4.1 Técnica Baseada em Célula Assimétrica

Um sensor capaz de detectar o envelhecimento de memórias submetidas ao NBTI através do uso de uma célula de memória assimétrica foi proposto em [5].

Como se sabe, o NBTI resulta em degradação do transistor pMOS quando submetido à tensões negativas entre seu *gate* e *source* enquanto opera sob temperaturas elevadas, esta degradação acaba por elevar sua tensão de *threshold* tornando-o por consequência mais resistivo, desta forma no trabalho apresentado em [5] é proposta uma célula assimétrica, com um transistor pMOS mais "forte" que o do inversor contrario conforme pode-se observar na figura 10 onde observa-se uma destas células, ou sensor, posicionada entre as células normais que compõe o bloco de memória. Também esta célula assimétrica é alimentada com uma fonte de tensão independente.

Figura 10 – Célula de memória assimétrica [5].



Seu princípio de funcionamento se baseia em, no momento que se quer determinar o quanto o envelhecimento já afetou a célula, se acionar a fonte de alimentação da célula assimétrica e se colocar em ativo baixo o sinal WL de forma a criar uma "disputa" entre os dois pMOS de cada um dos inversores. Tal disputa irá ser sempre resolvida de forma que o pMOS mais "forte" vença e por consequência passe a conduzir, porém na medida que ele vence a disputa ele também passa a envelhecer e desta forma torna-se cada vez mais fraco.

No momento então que uma nova disputa é iniciada e o pMOS, tido como mais forte, perde a disputa, automaticamente se sabe que a célula apresenta agora uma simetria e por consequência atingiu o envelhecimento limite determinado durante a construção do pMOS mais robusto. E por associação sabe-se também que as células que estão presentes ao seu redor provavelmente já atingiram uma idade semelhante.

Logo, para determinar o máximo envelhecimento que será permitido pelo sensor, deve-

se definir o tamanho do pMOS forte da célula, sua robustez inicial será a margem de idade permitida para que ocorra a detecção.

Tal abordagem apresenta pontos interessantes, como no que diz respeito à variações de processo por exemplo, pois várias destas células estão presentes em um bloco de memória e por consequência suportam tais variações de forma semelhante as outras células. Porém problemas relacionados a capacidade de detecção ficam um pouco limitados pois apenas se pode definir um valor de idade e este está ligado diretamente às características construtivas dos transistores da célula que serve como sensor. Além do problema relacionado à granularidade da proposta, pois afim de que se tenha um baixo *overhead* de área deve-se limitar o número de células que servem como sensor e desta forma perde-se em confiabilidade na informação disponibilizada, ou seja, o grau de certeza de uma detecção fica um pouco comprometido.

2.4.2 Técnica Baseada na Capacidade de Condução dos transistores pMOS

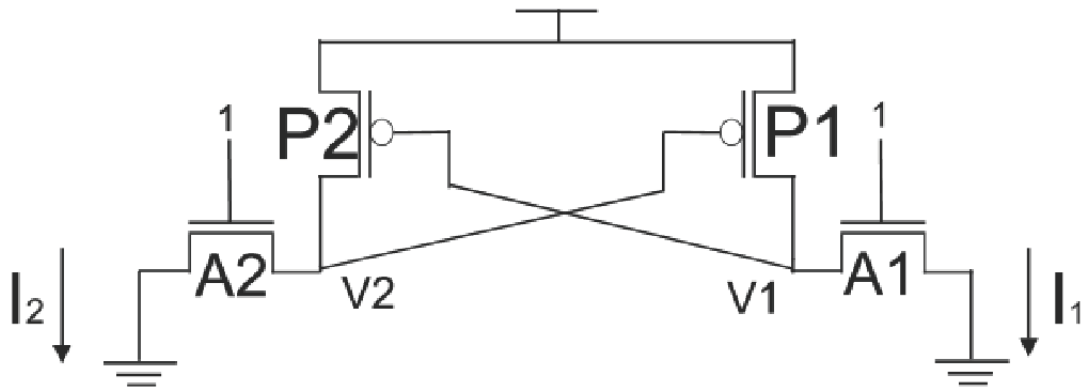
No trabalho proposto em [6] tentou-se avaliar o quanto uma célula de memória foi afetada pelo envelhecimento se utilizando da capacidade de condução dos seus pMOS.

Conforme foi visto na seção acima, o NBTI resulta em uma degradação dos transistores pMOS que resulta em elevação de sua tensão de *threshold* e por consequência em perda da sua capacidade de condução.

A técnica se utiliza então da diminuição da corrente que passa através dos pMOS como forma de determinar suas idades. Para realizar a medição desta perda de capacidade de condução, o trabalho sugere a alteração do circuito de escrita, presentes nas colunas de células, para que este seja capaz de forçar via *bitlines* a escrita de "1 lógico" em ambos os inversores da célula que será testada de forma simultânea, desta forma estes mesmo bitlines em uma próxima etapa são desconectados do circuito de escrita ficando conectados apenas à célula que está sendo testada e assim recebendo corrente que oriunda da

fonte de alimentação da célula através de seus pMOS, conforme figura 11, tal operação resulta em destruição do dado armazenado pela célula. E deve-se ainda salientar que obrigatoriamente os transistores de acesso A1 e A2 devem ser mais "fortes" que os dois pMOS que serão testados (P1 e P2).

Figura 11 – Teste da capacidade de condução dos pMOS [6].



Logo após esta fase de "detecção" surge a fase de avaliação dos resultados, quando o valor de tensão presente nos *bitlines* é analisado comparando-o com um valor de referência. Para que isto ocorra a tensão presente nos *bitlines* é então transformada em corrente C_{in} e subtraída de uma corrente de referência I_1 conforme figura 12, quando a célula que esta sendo testada é nova, a corrente resultante deverá portanto ser igual à zero, pois ambas devem ter o mesmo valor, porém com o avanço do envelhecimento da célula a corrente referente à tensão dos *bitlines* C_{in} deverá aumentar, e desta forma começará a aparecer uma diferença na subtração entre as correntes chamada C_{out} .

Esta corrente residual C_{out} é então amplificada e espelhada para o bloco de digitalização, que avalia se ela é maior ou menor que uma série de correntes de referência conforme é visto na figura 13

A técnica tenta contornar os problemas relativos à variações de processo escolhendo uma corrente de referência (que é usada no circuito que transforma a capacitância dos *bitlines* em corrente, conforme foi apresentado na figura 12) com um valor menor do que aquele verificado para uma célula nova, desta forma pretende aplicar uma abordagem mais conservadora, que identifica uma célula sempre como um pouco mais velha do que

Figura 12 – Transformação da tensão dos *bitlines* em corrente [6].

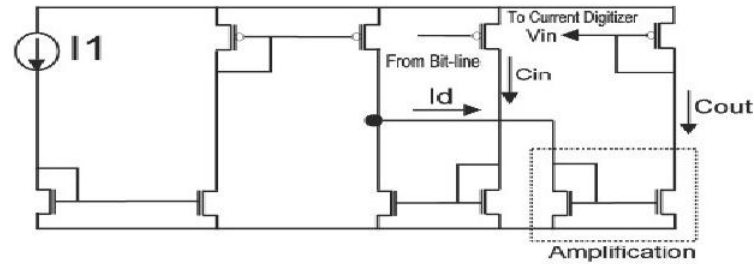
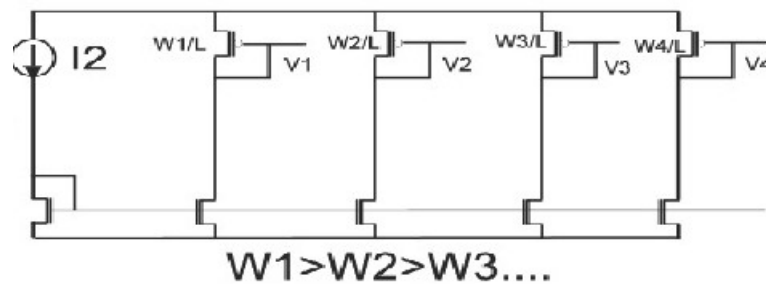


Figura 13 – Comparação da corrente encontrada com a corrente de referência [6].



realmente esta.

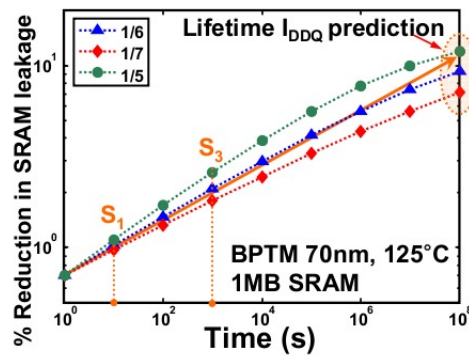
2.4.3 Técnica Baseada na Corrente de *Leakage*

No trabalho apresentado em [7] realizou-se a detecção do envelhecimento do bloco de memória através do monitoramento da corrente de *leakage* (i_{ddq}).

É apresentado em [7] que dentre os fenômenos decorrentes do envelhecimento através do NBTI podemos citar a degradação da SNM de leitura, a melhora da estabilidade na operação de escrita e a diminuição da corrente de *leakage*. Esta última se explica pelo fato de que ao avançar o envelhecimento de uma célula de memória verifica-se o aumento

do valor da tensão de *threshold* dos pMOS que ao ficarem mais "resistivos" apresentam menor corrente de *leakage*. Desta forma, o trabalho afirma que esta redução da corrente esta diretamente ligada ao envelhecimento, conforme se pode verificar na figura 14 e por consequência pode ser utilizada para determinar a idade de uma célula. É sugerido então o monitoramento da corrente total de *leakage* de um bloco de memória, de forma a se poder ter uma estimativa de sua média de idade. Tal abordagem, conforme foi explicado no decorrer do trabalho, se mostrou também tolerante aos *corners* FF, SS, FS, SF.

Figura 14 – Corrente de *leakage* de uma SRAM x idade [7].



Porém, conforme foi salientado acima, tem-se apenas uma "média" das idades das células que compõe o bloco que esta sendo monitorado. Desta forma, quanto maior o bloco de memória monitorado, menor a capacidade de acerto do sensor, pois enquanto uma célula pode estar já totalmente comprometida outra ainda pode estar "nova" e desta forma a corrente de *leakage* média destas duas células será correspondente a uma célula com uma idade mediana e desta forma capaz de funcionar com ainda alguma margem de tolerância ao ruído. Logo, tal abordagem requer um estudo prévio para definir a resolução que terá tal sensor, ou seja, o número máximo de células que será colocado em um mesmo bloco para ser monitorado. Esta decisão será tomada em função do tipo de aplicação para o qual o sistema será desenvolvido.

2.4.4 Técnica Baseada na Alocação de Dados Visando Diminuição do Estresse Sob as Células

No trabalho apresentado em [8] é apresentada uma metodologia que visa o uso homogêneo das células de SRAM. Desta forma espera-se que ao utilizar todas as células de forma igual a vida útil da memória como um todo venha a aumentar. Em [8] é apresentada então um controle através de software capaz de distribuir o uso de blocos de *Scratchpad Memory* (SPM). O projeto das SPMs dá aos programadores um maior grau de liberdade sobre como será seu uso, este novo grau de liberdade é que permite o uso de funções em software capazes de realizar o controle da forma como serão utilizados os blocos que compõe as SPMs.

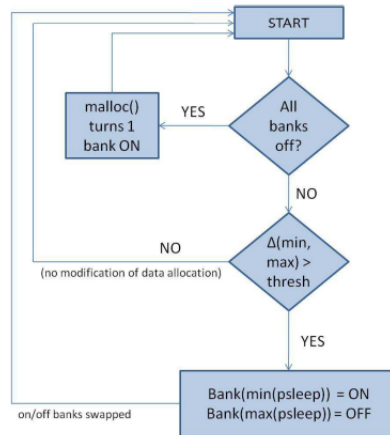
O trabalho então explica que os blocos de memória que não estão sendo utilizados acabam sendo colocados em um estado de espera. Sabe-se que uma das formas de diminuir o avanço do envelhecimento das células de SRAM é a diminuição da tensão de alimentação de células que estão em espera. Portanto, é o tempo que cada bloco da memória passa no estado de espera que acaba por ser utilizado como métrica para as funções de controle. Estas funções, ao perceberem que um bloco de memória já está a tempo suficiente na condição de espera os trocam por blocos que estavam a muito tempo em operação. Esta troca de blocos pode se dar automaticamente, sempre que um bloco na condição de espera tiver seu tempo total nesta condição superior à um dado valor de *threshold*, que responde pelo máximo envelhecimento que se deseja deixar atuar sobre o bloco.

As funções criadas para a manipulação destes blocos foram escritas na linguagem C, adicionando algumas extensões à função *malloc()*.

Na Figura 15 observa-se o fluxograma de controle da técnica apresentada em [8], neste fluxograma observa-se o registrador *psleep*, este registrador é responsável por monitorar o tempo que cada bloco permanece ativo e em espera. Através dos valores de *psleep* é que as funções tomam as ações de trocar um bloco por outro. Portanto, sempre que δ , que corresponde à relação entre o tempo que o bloco ficou ativo pelo tempo que ficou

em espera, é superior à um dado *threshold*, o bloco com maior tempo ativo é desligado (colocado em espera) e passa seu conteúdo para o bloco com menor tempo ativo.

Figura 15 – Fluxograma da Técnica Apresentada em [8]



Pode-se perceber então, que esta técnica apresenta a capacidade de realizar um melhor aproveitamento da memória como um todo ao estender sua vida útil, porém é incapaz de monitorar o envelhecimento dos blocos não tendo controle sobre este e nem sobre a confiabilidade que se pode obter de cada célula.

2.4.5 Estudo Comparativo Entre as Técnicas

Através das técnicas apresentadas anteriormente se percebe que todas possuem alguma vantagem e também alguma desvantagem quando comparadas umas às outras. Também se percebe que qualquer uma destas técnicas acaba gerando algum *overhead*, seja ele de área, de potência ou performance, que desta forma devem ser levados em consideração.

Afim de que se resumir as principais diferenças entre as técnicas vistas anteriormente que visam a detecção do envelhecimento de células de SRAM, decidiu-se pela construção da tabela 1, que busca resumir as principais características positivas e negativas de cada uma.

Tabela 1 – Comparativo Entre as Diferentes Técnicas Apresentadas

BIST [6]	Vantagens	Desvantagens
	Identifica envelhecimento célula à célula; Velocidade de detecção;	Destruição do dado armazenado durante a fase de testes; Complexidade de Implementação;
Célula Assimétrica [5]	Imunidade às variações de processos; Possibilidade de troca das células velhas.	Não Garante a Idade de cada célula, apenas uma média entre as idades de um grupo; Grande aumento de área ao projeto.
Método por IDDQ [7]	Simplicidade de implementação; Baixa utilização de área.	Não Garante a Idade de Cada Célula; Susceptível às variações de processo.
Método por Realocação em função do uso [8]	Não envolve alteração no hardware; Aumento da vida útil dos blocos.	Não Garante a Idade das Células; Atende apenas à um tipo de memória (SPM).

3 PROPOSTA

Este trabalho de mestrado propõe a especificação, implementação, validação e avaliação de uma técnica baseada em *hardware* que visa o monitoramento dos efeitos gerados pelo fenômeno do NBTI. Em mais detalhes a técnica proposta tem como principal objetivo aumentar a confiabilidade a longo prazo de SRAMs através da inserção de um sensor *On-Chip* capaz de monitorar o grau de envelhecimento das mesmas. O sensor utiliza alguns conceitos relacionados as técnicas apresentadas nos capítulos anteriores; pois visa monitorar a deterioração do transistor pMOS através de operações de escrita em uma célula.

Sabe-se que a operação de escrita, ao contrário da operação de leitura, é facilitada pelo envelhecimento da célula, pois uma mínima tensão em um *bitline* já se torna capaz de fazer o transistor pMOS envelhecido parar de conduzir e também que sob o efeito do NBTI os transistores pMOS demoram mais a conduzir assumindo assim um comportamento "resistivo". A principal ideia por traz da técnica proposta baseia-se no monitoramento do consumo da célula quando realizando sucessivas operações de escrita (tendo o cuidado de alternar os valores que estão sendo escritos afim de que exista consumo dinâmico) e na comparação de tais valores de consumo com uma referência. Desta forma o envelhecimento manifesta-se como a diminuição no consumo dinâmico da célula durante as operações de escrita a medida que os transistores pMOS envelhecem.

Assim, afim de minimizar problemas oriundos de variações no processo de fabricação, decidiu-se por utilizar uma referência que sofresse simultaneamente com a célula testada as mesmas variações. Dessa forma, foi utilizada como referência uma célula nova. Assim

espera-se que esta célula apresente um comportamento semelhante ao das células que serão testadas quanto às variações de processo, resultando assim em uma adaptação natural da referência à essas variações.

A técnica proposta mostrou-se capaz de detectar o envelhecimento de células presentes em SRAMs desconsiderando a forma como o mesmo ocorreu, ou seja, não importando qual foi a relação de envelhecimento entre o transistor pMOS de cada inversor que compõem a célula. Além disso a técnica apresentada caracteriza-se por ser facilmente implementada em um bloco de células e também de apresentar um consumo de energia desprezível em função dos poucos elementos que necessitaram ser adicionados ao bloco.

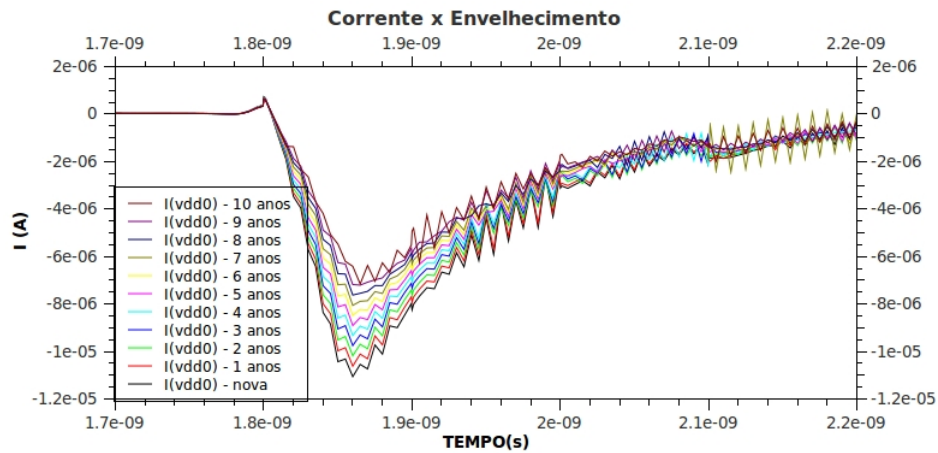
3.1 Princípio Fundamental

A técnica proposta tem como objetivo identificar o nível de envelhecimento de células de SRAM através do monitoramento da redução da capacidade de condução dos transistores pMOS em função da sua elevação da tensão de *threshold*. Para executar tal tarefa foi observado que existe uma redução na corrente drenada durante alimentação das células ao se realizarem operações de escrita conforme as mesmas são afetadas pelo fenômeno do NBTI, isto pode ser verificado na figura 16.

A figura 16 ilustra a corrente de alimentação dinâmica de uma célula durante duas operações de escrita, cada uma destas operações de escrita foi realizada trocando-se o valor à ser armazenado na célula para que assim houvesse consumo.

Assim, a figura 16 mostra uma célula que envelhece ano à ano até uma idade limite de 10 anos com ambos os seus transistores pMOS envelhecendo em paralelo (de forma equilibrada). Quando a célula envelhece apenas um dos transistores pMOS, deve-se levar em conta que existe uma mudança no fator de deslocamento do valor do *threshold*, pois conforme a equação 2.3, apresentada no capítulo 2.2, o β deve ser modificado, neste caso para o valor 1 [28].

Figura 16 – Redução das correntes de alimentação com o envelhecimento



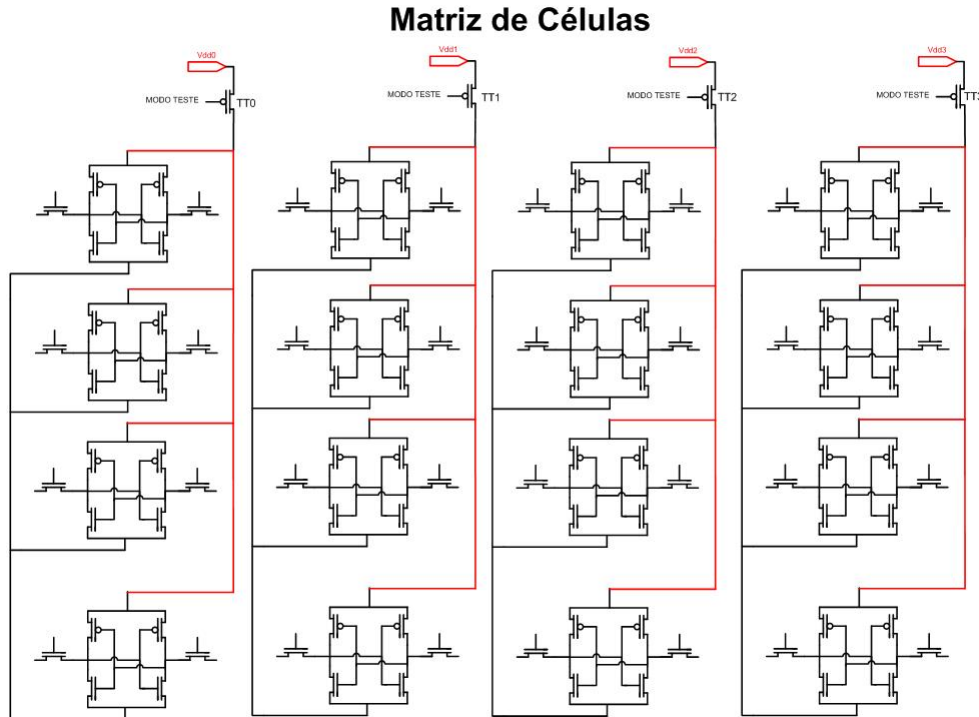
Desta forma, conforme a equação 2.3, o valor do *threshold* terá um fator de deslocamento duas vezes maior do que quando o β for de 0,5 [16]. Assim irá ocorrer uma compensação do inversor cujo pMOS está "novo", deslocando a tensão resultante na capacitância do barramento de alimentação de forma semelhante ao observado em uma célula que teve um envelhecimento equilibrado.

Em outras palavras, a realização do trabalho necessita que seja realizado o monitoramento do consumo de cada célula que é acessada no momento em que se realizam operações de escrita. Desta forma foi criada uma matriz de células cujas linhas de alimentação são separadas por colunas, ou seja, cada uma das colunas de células opera com uma linha de alimentação distinta. Juntamente com cada uma destas linhas de alimentação, foram inseridos transistores do tipo pMOS capazes de seccionar a alimentação durante as fases de teste, esta conforme foi mencionado na introdução deste capítulo, foi uma abordagem pensada em função da necessidade de se economizar energia quando realizando testes em sistemas embarcados.

Sabendo-se então, que as células de memória contribuem para a criação de uma capacitância total nas linhas de alimentação na ordem de 500fF por coluna, para uma memória de 512 células por coluna, decidiu-se por utilizar a queda de tensão nesta capacitância para realizar o monitoramento do consumo de cada célula durante as operações de escrita.

A figura 17 mostra a organização das colunas de células com a alimentação separada entre cada uma destas.

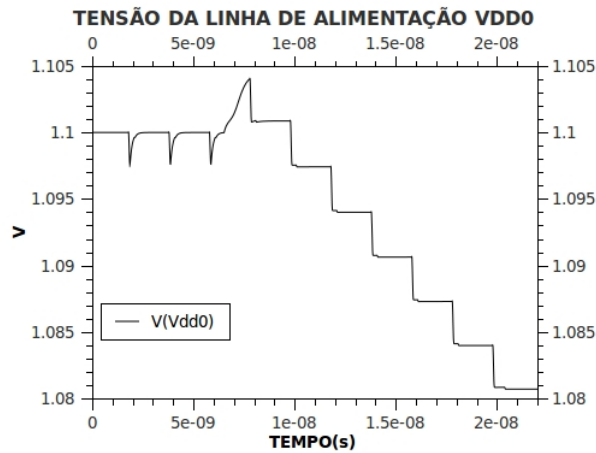
Figura 17 – Colunas de células com transistores nas linhas de alimentação.



O monitoramento do nível de envelhecimento ocorre através da observação do nível de tensão na linha de alimentação da coluna de células onde esta a células que esta sendo testada após uma sequência de escritas onde o valor à ser escrito é sempre alternado. A figura 18 ilustra um total de três operações de escrita antes do sensor ser acionado (antes da interrupção das linhas de alimentação) e logo após o início da operação de monitoramento (quando as linhas são interrompidas.) A partir da figura 18 é possível observar a queda de tensão. Note que o valor que esta sendo escrito na célula é sempre alternado entre "0" e "1" para que se tenha consumo dinâmico. Além disso, convém mencionar que neste momento não é realizada a avaliação do nível de envelhecimento da célula e sim apenas demonstrado o princípio que foi utilizado para tal.

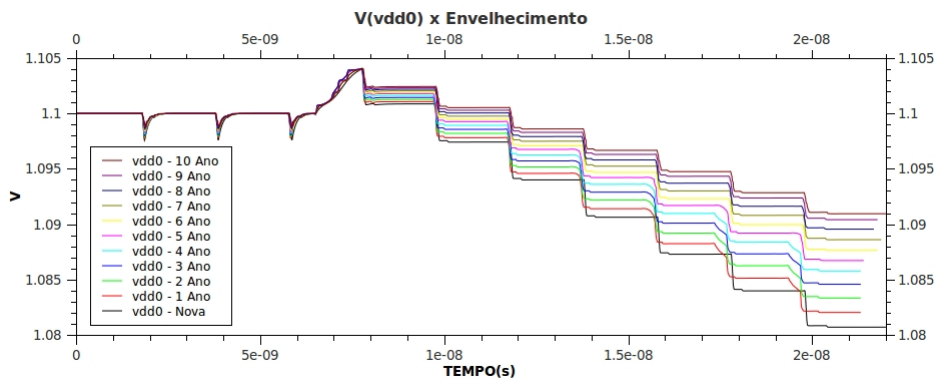
Como foi mostrado na figura 16, as células passam a consumir menos durante as operações de escrita quando envelhecidas. Assim, a tensão com valor mais baixo será aquela da célula nova e quanto mais velha for a célula avaliada maior será o nível de tensão

Figura 18 – Tensão na linha de alimentação antes e durante a fase de detecção do envelhecimento.



encontrado na capacitância da coluna de alimentação.

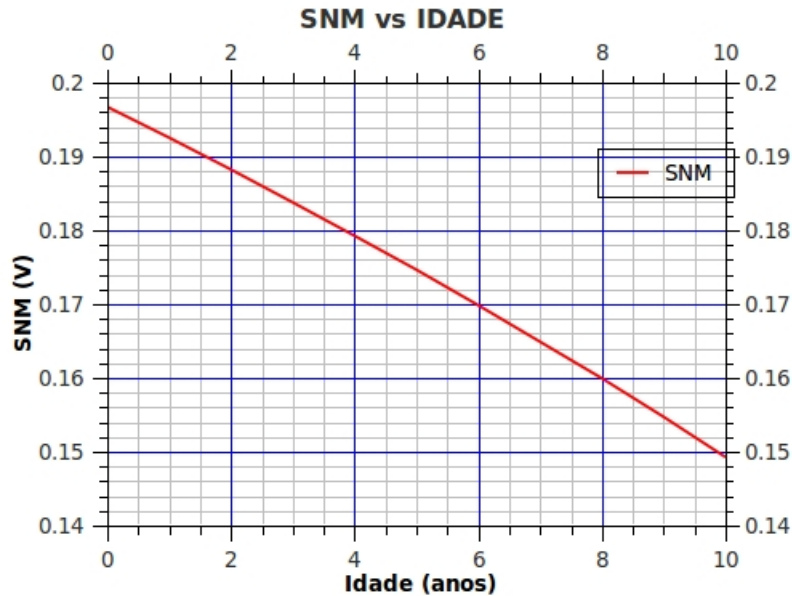
Figura 19 – Curvas de tensão da capacitância de alimentação vdd0 vs. idade.



Na figura 19 observa-se então o resultado de 11 simulações, relacionadas a mesma célula de memória, com níveis de envelhecimento distintos, esses níveis vão desde uma célula nova até uma célula com 10 anos. Torna-se portanto visível que a cada ano adicionado à célula, maior é a tensão na linha de alimentação da coluna. Além disso, quanto maior o número de operações de escritas realizadas maior acaba sendo a diferença em tensão encontrada de um ano para outro.

É importante ainda observar que conforme avança o envelhecimento menor se torna a diferença de tensão observada entre as curvas correspondentes as diferentes idades. Podemos verificar na figura 20 que o valor da SNM tende a acelerar sua queda quando a idade do transistor pMOS avança.

Figura 20 – SNM vs. idade.



Desta forma, como a SNM cai mais rapidamente para células de maior idade, da mesma forma, as curvas de tensão acabam por ocorrer com menor intervalo umas das outras, ou seja, a distância entre uma curva e outra diminui e a memória perde robustez quando exposta a ruídos com mais velocidade conforme o envelhecimento avança para idades maiores. Como o sensor utiliza como forma de detecção do nível de envelhecimento a variação de tensão destas curvas, acaba desta forma detectando um avanço do envelhecimento à uma razão semelhante ao que ocorre em células mais velhas.

Essas curvas são explicadas pela não linearidade de condução dos transistores do tipo pMOS em relação à tensão aplicada em seu *gate* que quando próximos ao valor de *threshold* passam de uma condição "aberta", em que não conduzem, para uma condição de condução quase que imediata. Logo, quanto maior o deslocamento do *threshold*, mais rapidamente o pMOS tende a parar de conduzir quando se diminui a tensão em seu *gate*.

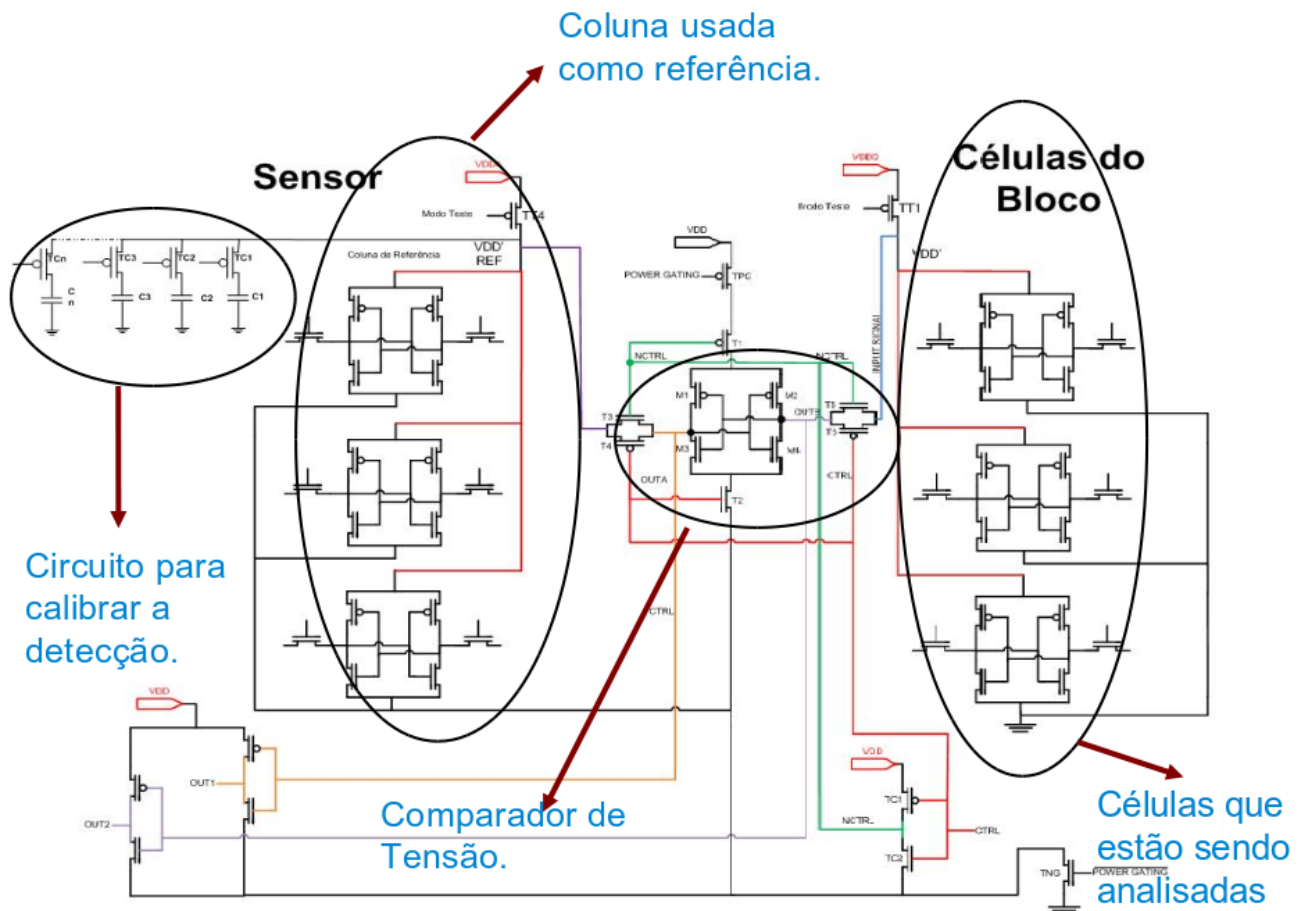
3.2 Arquitetura Proposta

O sensor utilizado para o monitoramento do nível de envelhecimento das SRAMs é, conforme esquemático da figura 21, composto por um circuito comparador de tensão, um

conjunto de chaves para o seccionamento da alimentação de cada uma das colunas do bloco de células bem como a própria matriz de células que sofreu restrições de construção para possibilitar a detecção do envelhecimento.

Ainda é adicionado à este circuito um conjunto de transistores e capacitores para que se possa assim ajustar a referência para a idade que se deseja detectar.

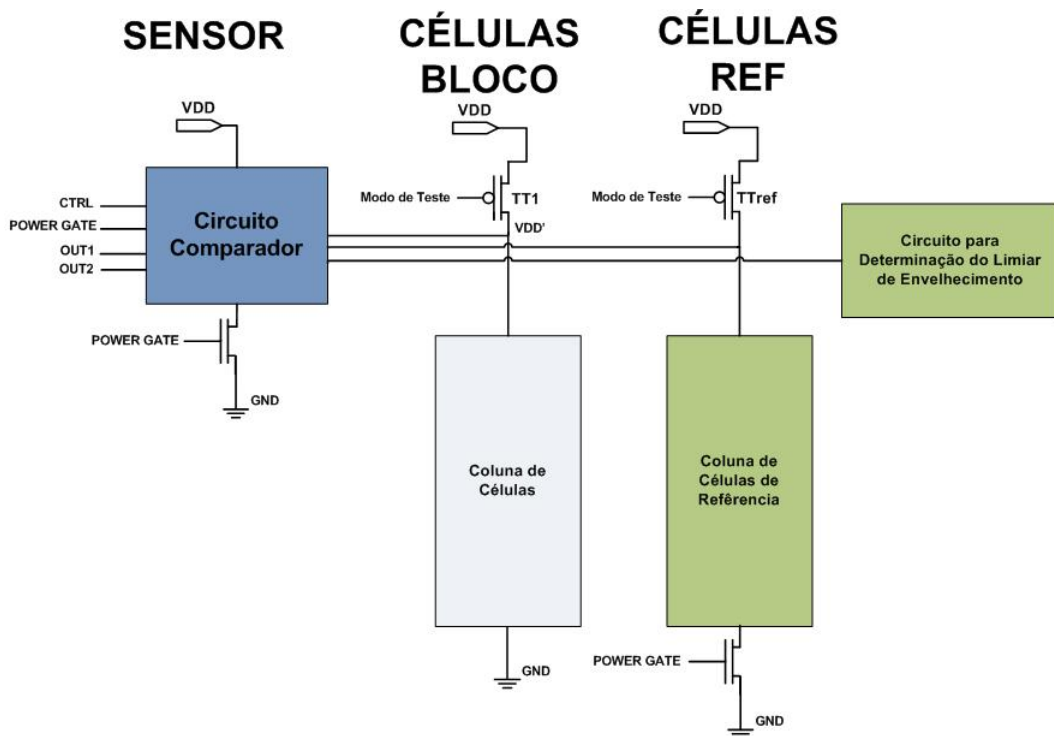
Figura 21 – Esquemático do sensor.



O diagrama de blocos que apresenta a solução proposta nesta dissertação é mostrado na figura 22. Pode-se perceber, através deste diagrama de blocos, como estão conectados os diferentes circuitos apresentados acima. Percebe-se pela figura 22 que o bloco mais importante é o do comparador de tensões, pois é este quem avalia qual o nível de envelhecimento da célula que está sendo testada. A avaliação é realizada através da referência que o comparador de tensões recebe do circuito utilizado para setar o valor de detecção desejado. Estes dois blocos (circuito comparador e circuito de calibração) são ainda conec-

tados à uma coluna de células utilizada como referência que cria um valor de consumo sempre igual ao de células jovens. Esta coluna de células utilizada como referência é protegida do envelhecimento através da técnica de *power-gating*. O sensor mede, então, o valor da célula a ser testada através de sua conexão com o barramento de alimentação da coluna ao qual esta pertence, comparando-a à coluna de células de referência.

Figura 22 – Diagrama de blocos da proposta.



A seguir os principais blocos que compõe o sensor serão descritos em mais detalhes.

3.2.1 Comparador de Tensão

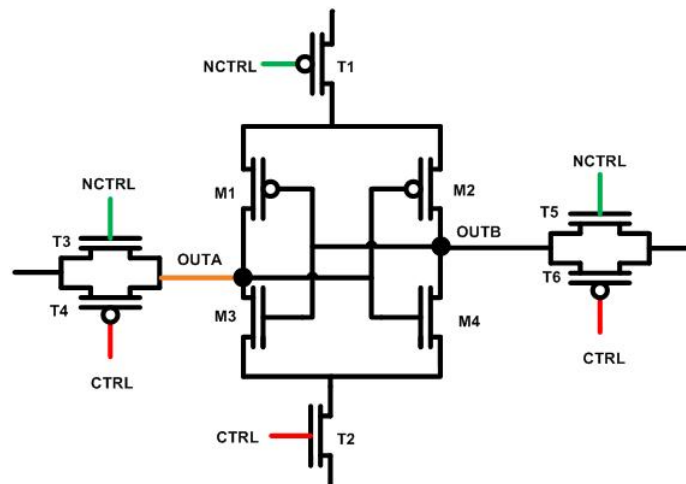
O funcionamento do circuito comparador de tensão é muito semelhante ao do amplificador de sinal apresentado no capítulo 2.1. Devido às suas características construtivas a operação deste circuito pode ser separada em duas etapas. Na primeira etapa, denominada etapa de carga, o circuito recebe os sinais que serão avaliados e na segunda etapa, denominada etapa de avaliação, o circuito se desconecta dos sinais que estão sendo medidos para disponibilizar o resultado da comparação.

De acordo com a figura 23, os transistores T1 e T2 são os responsáveis por conectar

o sensor à referência e a alimentação e desta forma, ao mesmo tempo os transistores T3, T4, T5 e T6 são os que formam os canais de acesso.

Durante a fase onde o circuito recebe os sinais que serão comparados (etapa de carga) os transistores T1 e T2 ficam em aberto enquanto os sinais entram através dos dois canais de acesso. Durante a etapa de avaliação ocorre o inverso, os canais são então fechados através dos transistores T3, T4, T5, T6 e os transistores T1 e T2 voltam a conectar o sensor à referência e à alimentação para que assim seja disponibilizado o resultado da comparação.

Figura 23 – Circuito comparador de tensão.



3.2.1.1 Geração da Tensão de Referência

Como referência, decidiu-se pelo uso de uma coluna de células, que através da técnica de *power-gating* é protegida do envelhecimento assim como o próprio sensor. Desta forma, ao não envelhecer durante a maior parte do tempo (só envelhece nos momentos em que é utilizada para realizar testes), disponibiliza o comportamento de uma célula jovem durante as rodadas de teste de toda a vida útil do bloco de memória.

Também foi testada a possibilidade de se utilizar um divisor resistivo como referência, setando através da adição ou remoção de resistores o valor desejado como *threshold*, porém esta abordagem por apresentar menor capacidade de adaptação não passou pelos testes com variações de processos e de temperatura, não conseguindo se adaptar às mudanças

de comportamento aos quais os transistores que formam as células foram submetidos.

Outra vantagem do uso de uma coluna como referência, é a possibilidade de se mudar a frequência de operação do bloco de memória sem que ocorra diferença na leitura do sensor, pois com o uso da coluna de células como referência é apenas o número de escritas realizadas nas células que serão testadas que realmente importa e não o tempo total utilizado em cada operação.

Os resultados que comprovam o uso de uma coluna de células no lugar de um divisor resistivo como sendo uma abordagem mais acertada são apresentados no capítulo 5.4 juntamente com algumas considerações sobre variações de processos.

3.2.2 Determinação dos Parâmetros de Configuração do Sensor

Através da observação de cada uma das diferentes correntes elétricas originadas das operações de escrita se conseguiu determinar a resistência equivalente das células para cada um dos diferentes níveis de envelhecimento testados. Logo, observou-se que para cada nível de envelhecimento existia uma resistência equivalente diferente, tendendo à aumentar conforme progresso do envelhecimento. Este aumento da resistência, que é originado do deslocamento do *threshold* do pMOS, segue uma forma semelhante a curva de condução do *mosfet*, ou seja, a razão de aumento da resistência tende à diminuir conforme se aproxima da sua zona de corte.

De posse dos dados relativos às resistências das células primeiramente é realizada uma sequência de operações de escrita em uma célula nova. O número de operações será de acordo com o tamanho do bloco de memória e também com a sensibilidade que se deseja obter, pois quanto maior o bloco maior será o número de operações necessárias para criar uma diferença satisfatória de tensão (uma tensão que tenha alguma tolerância à ruídos) na coluna de células. Após realizar estas operações que devem ser sempre em número par (afim de que ambos os inversores da célula sejam testados um mesmo número de vezes), é através do valor da tensão resultante na linha de alimentação da célula que será possível

obter também sua capacitância através da equação de descarga de capacitores 3.1, ou seja, o caminho contrário à partir da tensão resultante na linha nos indicará qual a sua capacitância.

$$V_c = V_i \cdot e^{-\frac{t}{RC}} \quad (3.1)$$

Onde V_c é o valor de tensão que se deseja obter ao fim da descarga de corrente na capacitância, V_i é o valor inicial de tensão, t é o tempo total de descarga, R é a resistência a descarga e C o valor da capacitância.

Então deve-se verificar se todas as colunas possuem a mesma capacitância e caso contrário realizar uma equalização.

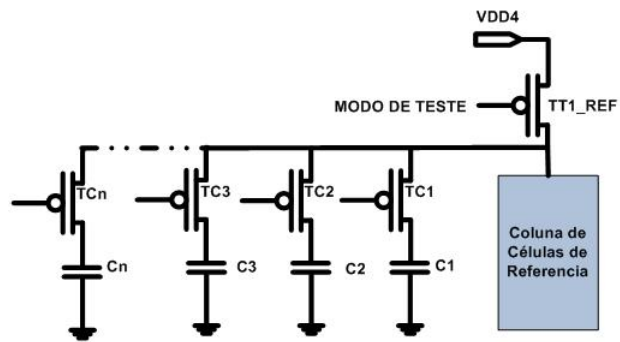
Após, utilizar a mesma equação apresentada acima, apenas tomando o cuidado de substituir a resistência equivalente à uma célula nova pela resistência equivalente à célula com a idade que se quer detectar. Com isto se obterá a capacitância que se deve adicionar a linha de alimentação da coluna de referência para que se possa realizar a detecção.

3.2.2.1 Circuito para Determinação do Limiar de Envelhecimento

Como se pode perceber pelo esquemático presente na figura 24, a importância do circuito que determina qual será o limiar do nível de envelhecimento permitido, reside em "envelhecer" a coluna de referência de forma artificial. Desta forma faremos a com que esta coluna tenha à idade que se deseja monitorar com o sensor. Para conseguir este efeito são adicionadas cargas capacitivas em paralelo com a linha de alimentação da coluna de células de referência de forma à manter sua tensão tão elevada quanto aquela que será observada na linha de alimentação da coluna de células em que houver uma célula envelhecida sendo testada. Assim, através do aumento da resistência esperado em uma célula durante a leitura frente ao envelhecimento, é então definida uma capacitância para ser colocada em paralelo com a linha de alimentação da coluna utilizada como referência, lembrando que o consumo de energia dinâmica das células que compõe esta coluna é igual

ao consumo de uma célula nova, uma vez que a coluna de referência não envelhece, pois passa todo o tempo em que não está sendo utilizada sendo protegida por um transistor que à desliga da massa.

Figura 24 – Circuito para determinação do limiar de envelhecimento.



4 VALIDAÇÃO

Este capítulo tem como objetivo demonstrar as formas utilizadas para a validação da técnica proposta. Desta forma buscou-se abordar as escolhas realizadas para a construção do sensor e quais as ferramentas utilizadas, provando assim que a técnica realmente é capaz de realizar o monitoramento do envelhecimento de células de memória.

4.1 Estudo de Caso 2

A validação da técnica proposta foi realizada a partir de um bloco de SRAM composto, conforme anteriormente mencionado no capítulo 2.1. A memória é baseada em células do tipo 6T. As características construtivas foram determinadas afim de se encontrar a melhor relação entre estabilidade e velocidade. Assim os seguintes aspectos construtivos foram observados: (1) onde a relação entre o transistor tipo pMOS (pull up) e o de acesso nMOS que conecta o inversor da célula ao *bitline* $RP = \frac{W_P}{W_a}$; (2) Relação entre o transistor tipo nMOS (pull down) e o de acesso nMOS $RC = \frac{W_N}{W_a}$. Baseado nos cálculos apresentados no livro *analysis and design of digital integrated circuits*, esses valores devem ser iguais à 1,5. Assim tentou-se manter a relação $RP = RC = 1,5$, para atender à esta exigência. A largura dos canais dos transistores tipo pMOS e nMOS de cada um dos inversores e dos transistores de acesso foram definidos, respectivamente, como W_{pMOS} (pull up) = 0,220nm W_{nMOS} (pull down) = 0,200nm e W_{nMOS} (Acesso) = 0,120nm. Chegou-se então, a $RP = 1,833\Omega$ e $RC = 1,667\Omega$.

Essa relação assegura que as células estarão equilibradas e que haverá estabilidade

frente à ruídos, na ordem de até 200mV no momento que for realizada uma leitura, segundo a análise da SNM. Além disso, essa relação indica que as células terão condições de trabalharem em frequências elevadas (até 1GHz), uma vez que os transistores de acesso são fortes o suficiente para realizar a carga e a descarga em tempo hábil.

Além da definição das características construtivas da própria célula, foram também definidas as características dos componentes dos circuitos auxiliares de leitura e de escrita (circuito de pré-carga, circuito de leitura, circuito de escrita e amplificador de sinal).

As operações de escrita e de leitura foram separadas em três estágios, pré-carga, descarga e aquisição. Na fase de pré-carga o circuito de pré-carga coloca os *bitlines* em ativo alto, já na fase de descarga ocorre a passagem da informação para os *bitlines* (operação de leitura) com a descarga dos *bitlines* sendo realizada pela célula ou a descarga dos *bitlines* pelo circuito de escrita (operação de escrita) para posterior conexão da célula aos *bitlines*.

Para cada uma das etapas anteriormente descritas foram definidos tempos para a habilitação dos sinais dos circuitos auxiliares. Desta forma, para uma operação de leitura 0,7ns ou 35% do tempo foram destinados à fase de pré-carga, 1ns ou 50% do tempo para a fase de descarga da célula com auxílio do *sense amplifier* e os restantes 0,3ns ou 15% do tempo para a fase de aquisição pelo circuito de leitura. Assim, as operações tiveram um período total de 2ns ou seja, a célula opera à 500MHz. Na fase de escrita os tempos em cada etapa são iguais. A única diferença é o fato de que na fase de descarga o circuito de escrita cria a diferença de tensão entre os *bitlines* e na fase de aquisição a célula é quem realiza a "leitura" do valor.

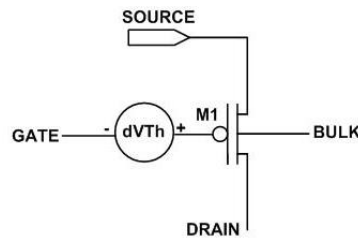
4.2 Caracterização dos Efeitos Fenômeno de NBTI em uma Célula SRAM

Para caracterizar o envelhecimento decidiu-se por utilizar a mesma técnica apresentada em [11] e [9] que, por sua vez utiliza uma forma de modelamento, conforme figura

25 capaz de ser utilizada em simuladores elétricos como HSPICE.

O modelo baseia-se na colocação de uma fonte em série com o *gate* do pMOS. Desta forma, ao se aumentar a tensão nesta fonte consegue-se deslocar o valor do *threshold*, realizando assim, a simulação do envelhecimento do transistor pMOS muito semelhantemente aquela que seria verificada em um circuito submetido ao NBTI. O circuito implementado é ilustrado na figura 25.

Figura 25 – Modelamento do envelhecimento [9]



Cada um dos transistores do tipo pMOS da célula que se deseja envelhecer recebe esta fonte dV_{Th} , proporcionado assim o modelamento do envelhecimento de forma individual. Para a realização do deslocamento do *threshold* foram utilizadas as informações apresentadas em [13] que indicam uma variação de 10% à 15% da tensão de *threshold* por ano sob estresse. De posse desse dado e sabendo-se que conforme [20] este valor ainda depende da probabilidade de se estar envelhecendo o pMOS de um inversor ou do outro [3], conforme 2.3, levou-se em consideração um deslocamento de 10% do limiar de acionamento do transistor para cada ano. Sabendo-se que o valor do *threshold* para o transistor pMOS da tecnologia utilizada é de 0,43V, tem-se então uma variação de 0,043V para cada ano de vida de um pMOS quando este é estressado ao máximo (100% do tempo). Esta abordagem foi bastante simplificada em função de que este não é o foco do trabalho (como os transistores envelhecem) e sim as consequências do envelhecimento e sua detecção, o valor escolhido foi baseado em [29], que afirma que transistores de maior valor de *threshold* envelhecem mais lentamente do que aqueles de valor menor.

Note que a degradação de células de SRAMs não depende totalmente da frequência com que ela muda o valor lógico armazenado, mas sim da quantidade de tempo que cada

valor lógico permanece armazenado na célula. Desta forma, técnicas que visam alternar o conteúdo de uma célula de tempos em tempos são muito úteis. Ao alternarem o valor armazenado pela célula acabam permitindo um envelhecimento simétrico em ambos os transistores do tipo pMOS, diminuindo a chance de que a célula venha à inverter o dado armazenado durante uma operação de leitura. Esta inversão acaba ocorrendo devido ao seu "desbalanceamento" e desta forma ocasiona uma diminuição da sua vida útil.

Sabe-se também que o envelhecimento das células pode se dar de uma forma não homogênea, principalmente em bancos de memória que são desprovidos de controle de distribuição deste envelhecimento. Nas células onde ocorre este envelhecimento não homogêneo, encontra-se uma tolerância à ruídos muito menor do que naquelas onde o envelhecimento ocorreu da mesma forma entre seus dois transistores pMOS. Esta observação indica que, considerando a configuração escolhida para o trabalho, uma célula com 5 anos de envelhecimento, em apenas um dos pMOS, possui a mesma tolerância ao ruído encontrada em uma célula com 10 anos onde o envelhecimento ocorreu de forma totalmente homogênea.

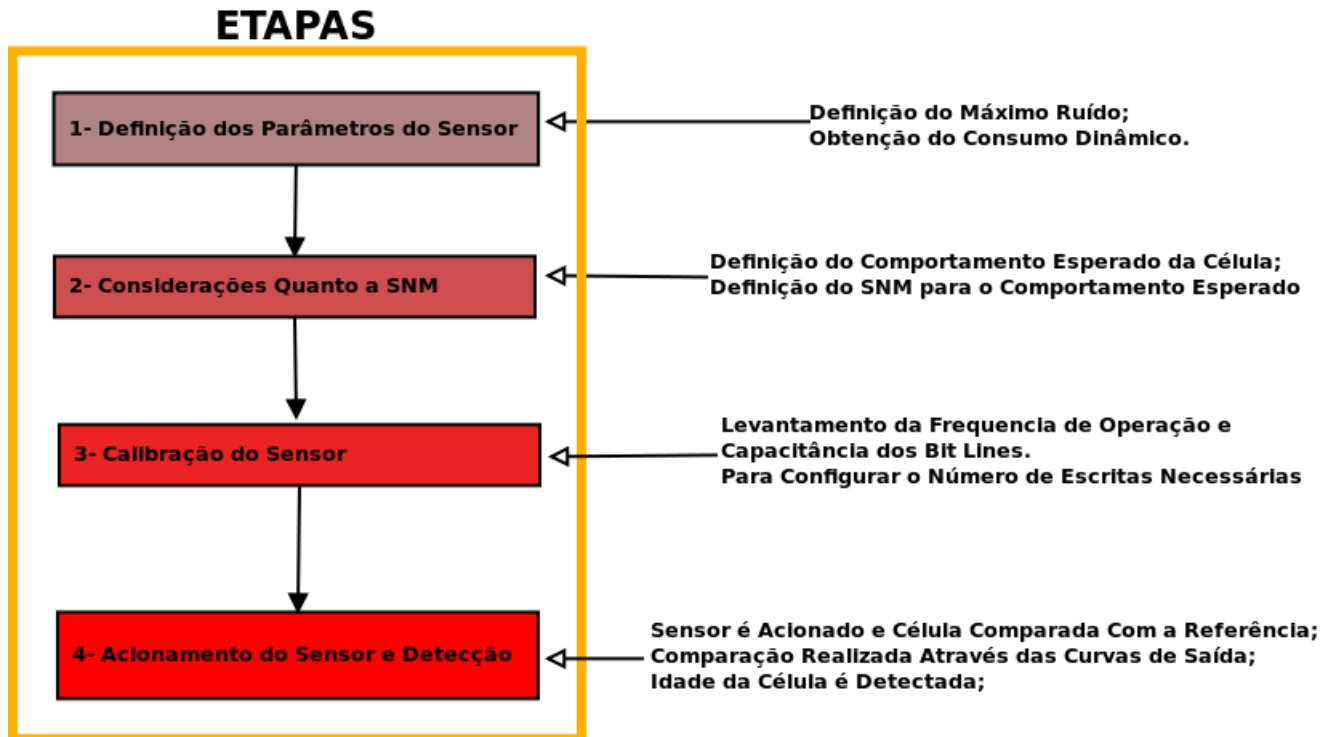
4.3 Fluxo

A validação do sensor proposto neste trabalho de mestrado foi realizada a partir de um bloco de memória de 512 linhas por 8 colunas, operando em 27°C e utilizando o *corner* TT (que será melhor explicado no capítulo 5.1.1). Conforme exemplificado na figura 17, a matriz de células possui as alimentações de cada coluna separadas umas das outras e com uma chave entre as mesmas e a alimentação do bloco. Desta forma, quando se deseja avaliar a idade de uma célula deve-se seccionar a alimentação da coluna onde ela está e realizar uma sequência "par" de escritas consecutivas tendo o cuidado de sempre alterar o valor que está sendo escrito entre cada operação. Desta forma, como existe alteração do valor que está sendo escrito, existe também consumo dinâmico, e assim o sensor é capaz de determinar o envelhecimento da célula testada em função da queda de condução dos

transistores pMOS frente ao envelhecimento.

É apresentado na figura 26 um fluxograma básico de como se dá a utilização do sensor, desde a calibragem até o monitoramento. Este fluxograma será melhor explicado através das etapas que serão descritas a seguir.

Figura 26 – Fluxograma de validação da técnica



4.3.1 Etapa 1 - Definição dos Parâmetros do Sensor

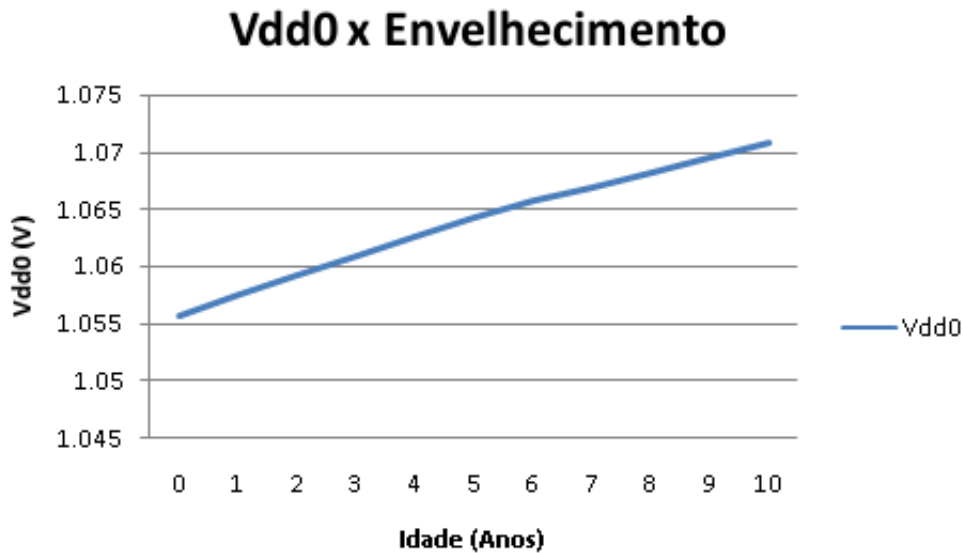
A primeira etapa da validação diz respeito a definição dos parâmetros do sensor, desta forma, primeiramente, deve-se verificar qual o máximo de confiabilidade que se deseja garantir nas células da SRAM, ou seja, qual o maior ruído que a célula deverá ser capaz de suportar.

Depois de escolhido o máximo ruído que deverá ser suportado pela célula é necessário definir o valor de *threshold* do sensor para que este detecte o envelhecimento da célula.

O valor de *threshold* do sensor é obtido através do levantamento das curvas de tensão obtidas através da simulação do envelhecimento de uma célula, segundo dados presentes

na bibliografia.

Figura 27 – Curva de tensão de uma célula em função do envelhecimento



Na figura 27, observa-se então que para cada idade que a célula atinge existe uma determinada tensão que é utilizada pelo sensor para ser comparada com a tensão de referência no momento que se deseja que este realize a detecção da célula envelhecida.

Esta curva de tensão esta diretamente ligada ao número de escritas que foram realizadas nas células e também ao consumo dinâmico da própria célula através do barramento de alimentação frente ao envelhecimento, que conforme foi observado no capítulo 3.1, tende à ser menor conforme o envelhecimento avança.

Desta forma, na figura 27 temos as tensões relativas ao consumo de uma célula que utiliza um barramento de alimentação com 500fF de capacitância. Esta capacitância esta relacionada ao número de células presentes na coluna, que neste caso são 512 células. Quanto maior o número de células em uma mesma coluna, maior a capacitância e por consequência maior terá que ser o número de escritas efetuado por célula para que se tenha uma diferença de tensão satisfatória entre uma idade e outra, ou seja, algo que tenha menor susceptibilidade de ser alterado ou modificado por um ruído.

4.3.2 Etapa 2 - Considerações Quanto à SNM

Conforme mencionado na etapa anteriormente descrita, podemos verificar através da figura 28 que existe uma grande diferença entre a SNM de uma célula que envelheceu de forma homogênea (mesmo tempo de estresse em ambos os transistores pMOS) e uma célula que teve seu envelhecimento acontecendo de uma forma desequilibrada, ou seja manutenção de um mesmo dado armazenado na célula durante a maior parte do tempo. A figura 28 nos trás duas curvas de SNM, uma para a célula onde o envelhecimento foi equilibrado e a outra onde o envelhecimento ocorreu com 80% do tempo de estresse em um dos pMOS e 20% do tempo de estresse no outro pMOS da célula.

Desta forma, para garantir um valor de SNM mínimo para uma célula, deve-se ter em mente qual será o comportamento demandado da memória, ou seja, se haverá algum controle sobre qual dado esta sendo mantido pela célula.

Decidiu-se portanto, através dos dados obtidos em [30] [31] a utilização deste valor que prova que mesmo com um valor superior à 75%, que é o valor de distribuição normalmente verificado para a distribuição do envelhecimento dentro de blocos de SRAM, a técnica se mostrou capaz de realizar a detecção corretamente.

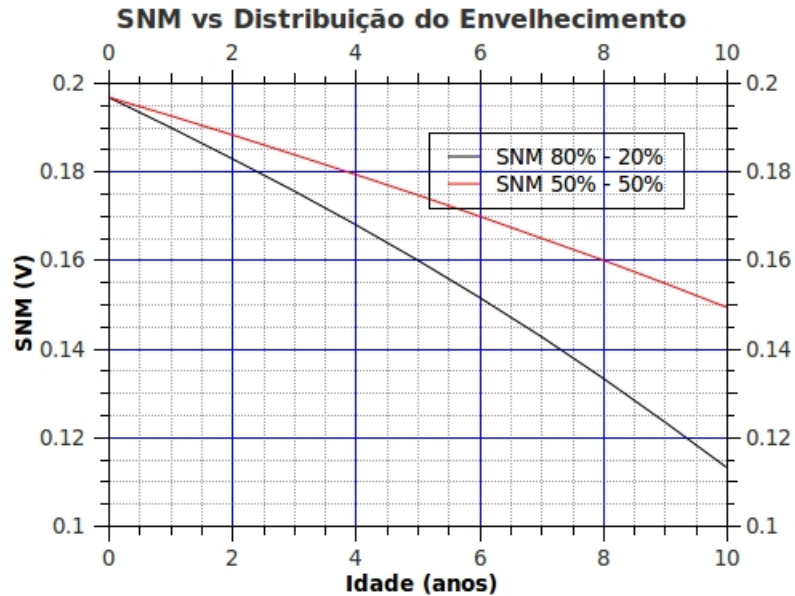
Se houver uma forma de se garantir uma distribuição de até 80% - 20% será a partir desta curva que deverá ser extraída a SNM para a idade desejada. Ou seja, uma célula com 6 anos nesta condição terá uma SNM equivalente à uma célula de 10 anos na condição em que os dois pMOS envelheceram juntos.

4.3.3 Etapa 3 - Calibração do Sensor

A calibração do sensor será realizada portanto para a idade previamente escolhida. De posse dos valores relativos à capacitância de alimentação das colunas à serem testadas e da coluna de referência, deverá ser setado o ponto de detecção do sensor.

O número de escritas necessário para realizar a comparação de tensão entre a coluna

Figura 28 – SNM de uma célula com envelhecimento não equilibrado



que possui a célula à ser testada e a célula de referência é determinado pela confiabilidade que se deseja do sensor, uma vez que quanto menor o número de escritas menor também será a diferença entre as tensões relacionadas ao envelhecimento, e estas estão igualmente sujeitas à ruídos externos. Esta informação é necessária para que se possa criar uma diferença de tensão entre uma idade e outra suficiente para a termos imunidade contra alguns ruídos.

Logo, se por exemplo, for desejado realizar uma detecção na décima quarta operação de escrita, deverá ser encontrado o valor de tensão correspondente à idade desejada para posteriormente ser configurada a referência.

Através do simulador HSPICE foi adquirida a corrente média para cada operação de escrita correspondente à cada uma das diferentes idades.

Podemos verificar na figura 29 que, mesmo que uma operação de escrita tenha um período de 2ns, quando o bloco opera em 500MHz, o tempo total em que realmente existe o consumo depende das características da própria célula. Se a célula suportar uma frequência superior à 500MHz, provavelmente o tempo em que esta estará realmente realizando consumo é menor do que os 2ns, logo, como a célula em questão suporta até

próximo de 1GHz de frequência, observamos pela figura 29 que seu período de consumo é de aproximadamente 0,85ns. De posse deste dado, foi realizado através do simulador HSPICE uma tomada das correntes médias neste período de tempo, que acabaram sendo inseridas na tabela 2, para a aquisição destes valores foi utilizado o comando:

```
.MEAS tran Out avg isub(XSrc0.vdd0) FROM=1.789ns TO=2.6315ns
```

Decidiu-se por utilizar esta abordagem (usar um período inferior aos 2ns para tomar a corrente média) para que assim tivéssemos condições de através dos mesmos valores de corrente trabalhar com frequências distintas.

Tabela 2 – Correntes Médias nas Operações de Escrita para Diferentes Idades

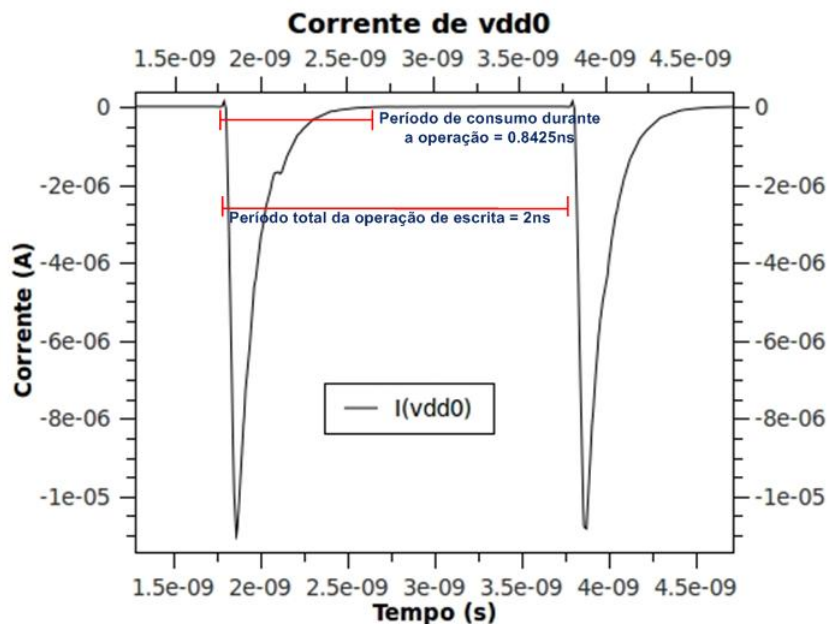
Idade	Corrente (A)	Resistência Eq. (Ω)	Tensão (V)
Nova	2,08E-06	5,31536E+05	1,05844
1 Ano	2,00E-06	5,49451E+05	1,05990
2 Ano	1,93E-06	5,69948E+05	1,06145
3 Ano	1,86E-06	5,91716E+05	1,06299
4 Ano	1,79E-06	6,13839E+05	1,06444
5 Ano	1,73E-06	6,36574E+05	1,06583
6 Ano	1,66E-06	6,61058E+05	1,06722
7 Ano	1,60E-06	6,86642E+05	1,06857
8 Ano	1,54E-06	7,13359E+05	1,06987
9 Ano	1,48E-06	7,41740E+05	1,07116
10 Ano	1,43E-06	7,71388E+05	1,07240

Ainda, conforme a tabela 2, com o uso da equação de descarga de capacitores, é obtida a tensão que será utilizada no ponto de acionamento do sensor. Desta forma, para configurar a coluna de referência de acordo com tal, é utilizada a mesma equação $V_c = V_i \cdot e^{-\frac{t}{RC}}$ para a determinação de quanta capacitância deve ser adicionada utilizando-se do circuito apresentado na figura 24. Neste caso no lugar de deixarmos como variável V_i , utilizamos o valor encontrado anteriormente e deixamos como variável a capacitância C , que terá como valor uma capacitância acima dos 500fF inseridos anteriormente. Esta diferença de capacitância será a quantidade necessária para ser adicionada em paralelo a capacitância da linha de alimentação da coluna de referência. Por exemplo, se o resultado for 600fF, deverão ser adicionados 100fF em paralelo a linha de alimentação da referência.

Para inserção do período na equação de descarga do capacitor é necessário apenas multiplicar o período de 0,85ns observado na figura 29 pelo número de operações de escritas realizadas, este será o valor de t na equação.

O valor de V , será de 1,1V que é a tensão de alimentação do circuito, e o de R será adquirido através deste valor dividido pelo valor das correntes encontradas através do HSPICE. Neste caso será utilizada uma resistência para detectar todas as células com mais de 4 anos.

Figura 29 – Determinação da corrente média durante uma operação de escrita

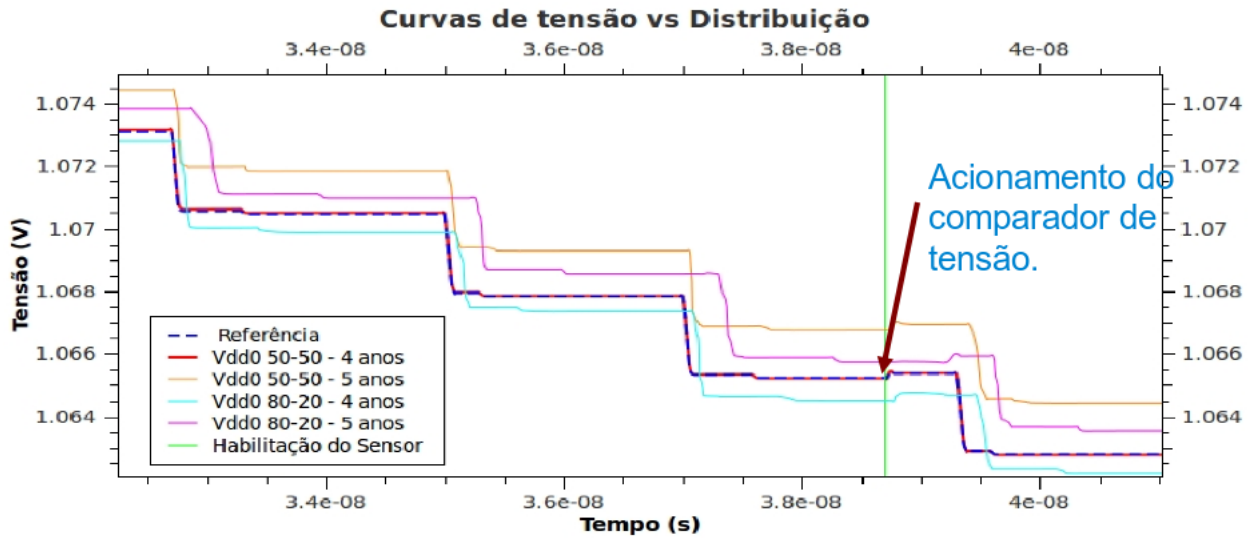


4.3.4 Etapa 4 - Acionamento do Sensor e Detecção

O acionamento do sensor é então realizado, após o número de escritas posteriormente determinado. Então a curva de tensão resultante no barramento de alimentação da coluna onde esta a célula sendo analisada é comparada com a curva de tensão da coluna de células de referência.

Na figura 30 são apresentadas as curvas de tensão para a idade de 4 e de 5 anos quando a célula apresenta um envelhecimento homogêneo (50% de envelhecimento para cada um dos pMOS da célula) e quando a célula apresenta um envelhecimento desbalanceado, neste

Figura 30 – Curvas de tensão que são comparadas com a referência

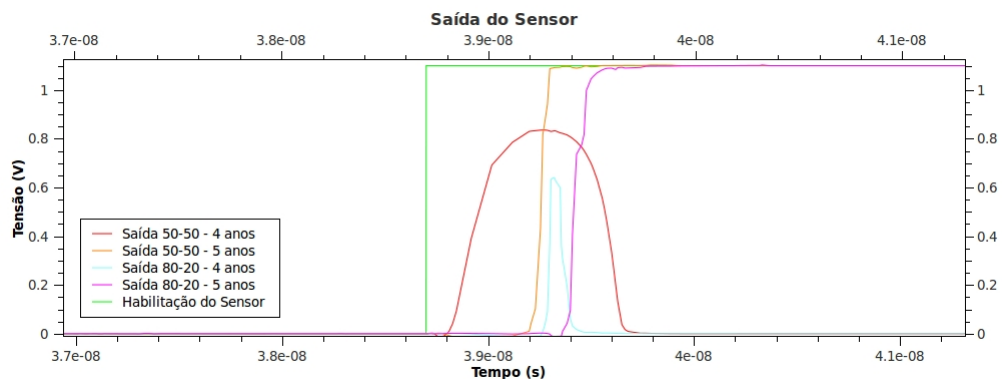


caso com uma relação de 80% - 20% entre os pMOS.

A figura 30, demonstra ainda que mesmo com a relação de envelhecimento bastante distinta, a mesma calibração para a coluna de referência foi capaz de detectar ambas as células com mais de 4 anos como envelhecidas.

Isto pode ser analisada no figura 31, onde estão presentes as diferentes saídas para cada uma das curvas da figura 30.

Figura 31 – Saída do sensor para as curvas da figura 27



Na figura 31 são apresentadas em ativo alto todas as saídas do sensor para as células mais velhas do que a referência e em ativo baixo todas as células que ainda estão de acordo com a referência, ou seja, ainda mantem uma SNM mínima desejada para o funcionamento da memória.

5 AVALIAÇÃO

Neste capítulo serão apresentados os resultados relativos ao funcionamento do sensor em um bloco de SRAM. Para a realização desta avaliação foi utilizado o método de Monte Carlo. Desta forma espera-se utilizar um número suficientemente grande de variáveis relacionadas à idade e à própria forma como se deu o envelhecimento da célula de memória que será testada, para assim avaliar com maior precisão seu funcionamento. Será dada especial importância ao comportamento relacionado ao seu funcionamento nas diferentes variações de processos de fabricação disponibilizadas pelas bibliotecas tecnológicas utilizadas.

Ao final, serão ainda levantados alguns pontos que restringem de certa forma o uso do sensor, que desta forma devem ser observados e levados em consideração durante sua utilização.

5.1 Simulações por Monte Carlo

Afim de se avaliar a capacidade de detecção do sensor, decidiu-se por utilizar o método de Monte Carlo. A forma como esta etapa foi conduzida se deu através do próprio simulador elétrico HSPICE, que possui implementada esta funcionalidade conforme pode ser observado nos anexos desta dissertação.

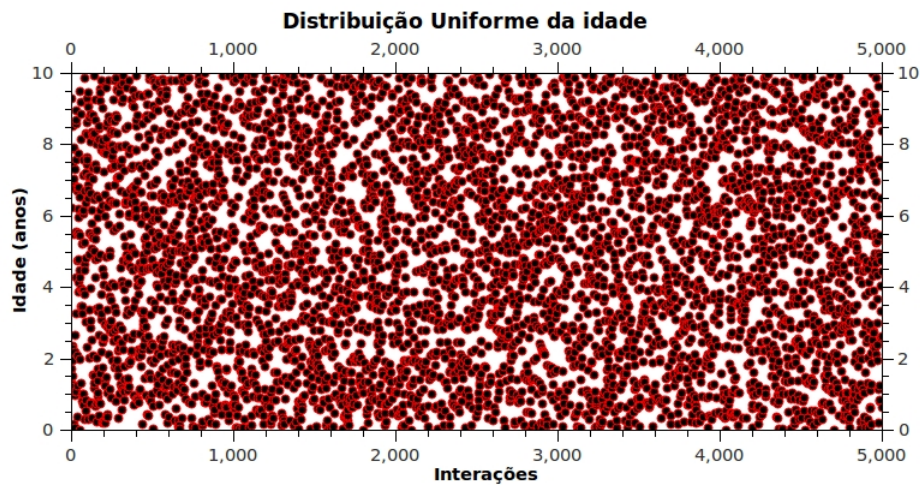
Através do simulador foram definidas as questões relativas ao tipo de entrada que seria sorteado pelo simulador para que assim fossem analisadas as respostas do sensor.

Optou-se por sortear tanto a idade que a célula poderia ter quanto a distribuição

desta idade entre os pMOS de cada inversor que compõe uma célula de memória.

Para o sorteio da idade da célula optou-se por uma distribuição do tipo uniforme. Logo todas as diferentes idades poderiam ser "sorteadas" dentro de um intervalo de 0 à 10 anos durante todas as 5000 diferentes rodadas de simulação, isto pode ser verificado pela figura 32.

Figura 32 – Distribuição uniforme para a entrada de diferentes idades

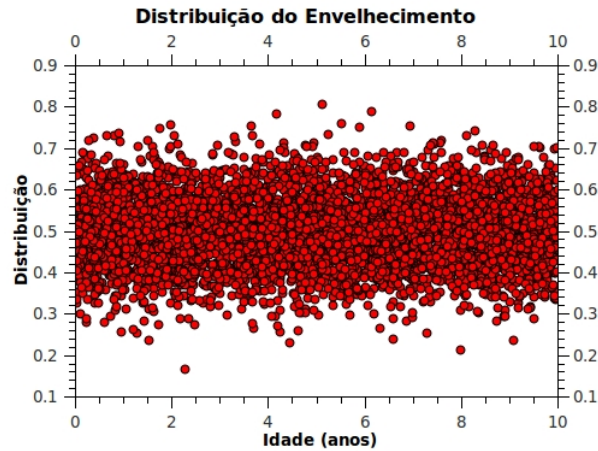


Para a distribuição do envelhecimento entre os pMOS dos inversores que compõe a célula, optou-se por uma distribuição do tipo gaussiana, centrada em 50% de estresse para cada transistor pMOS e com a possibilidade de excursionar de 0 à 100%. Esta distribuição pode ser analisada na figura 33 e foi escolhida tendo em vista resultados tipicamente observados ao se associar a técnica de *bit flipping* aos blocos de SRAM conforme [28].

Para a avaliação da capacidade de detecção foram então realizadas 5000 iterações para cada uma das possíveis variações de processos, dando condições, tendo em vista a capacidade computacional disponível, para a realização das simulações buscando-se desta forma uma resposta confiável do sensor. Para cada uma destas simulações foi sorteada uma idade e também a forma como seria distribuída entre os inversores da célula avaliada. Os resultados relativos as simulações para as diferentes variações de processão são observadas na seção seguinte (5.1.1).

A capacidade de detecção do sensor foi então avaliada em função dos diferentes re-

Figura 33 – Distribuição gaussiana do envelhecimento entre os pMOS de cada célula



sultados criados pela ferramenta de Monte-Carlo conforme as condições de probabilidade arbitradas acima, onde foram definidos os valores de variação e média para cada uma das possíveis configurações de entrada para a ferramenta. Em função da capacidade computacional e do tempo disponível foram realizadas 5000 iterações para, conforme será melhor explicado na seção a seguir (5.1.1), cada uma das variações de processos disponibilizadas pelas bibliotecas de tecnologia utilizadas.

5.1.1 Simulações por Monte Carlo considerando variações de processos

O sensor foi avaliado através da técnica de Monte Carlo para diferentes variações de processo *inter-die*, ou seja, aquelas que estão presentes na pastilha de silício utilizada para a fabricação das memórias.

As variações disponíveis pelas bibliotecas utilizadas compreendem somente a pastilha como um todo, variações *inter-die*, sem no entanto simular as diferentes variações que podem ser encontradas em uma mesma pastilha, denominadas variações *intra-die*.

Para exemplificar então os testes realizados com o sensor frente às variações *intra-die* utilizou-se como exemplo a detecção de uma célula de memória inserida em um bloco de 8 colunas por 512 linhas (4kbits) em todos os *corners* (variações de processo) de fabricação

disponíveis nas bibliotecas utilizadas. O sensor foi setado para realizar a detecção de qualquer célula com SNM de 0.18V ou menor, ou seja, células que segundo o modelo utilizado teriam 4 anos ou mais.

Foram realizadas simulações nos *corners* TT (*Typical - Typical*), FF (*Fast - Fast*), SS (*Slow - Slow*), FS (*Fast - Slow*) e SF (*Slow - Fast*). O significado destas siglas leva em consideração o comportamento de cada um dos tipos de transistores utilizados na simulação (tipo pMOS e nMOS). Desta forma, a primeira letra da sigla se refere à todos os transistores tipo pMOS e a segunda letra se refere aos transistores do tipo nMOS. A denominação *typical* é utilizada para denominar todos os transistores que apresentam o comportamento que lhes é esperado; A denominação *slow* refere-se à todos os transistores que apresentam um maior valor de *threshold* e que desta forma levam mais tempo para mudar de estado no momento que são acionados; Os transistores com denominação *fast* são aqueles que por outro lado tem sua tensão de *threshold* com um valor de tensão menor do que o esperado e que desta forma apresentam maior velocidade para mudar de estado quando acionados.

Em mais detalhes, a tabela 3 mostra os resultados para cada um dos *corners* utilizados. Em mais detalhes, a tabela 3 mostra os resultados referentes às 5000 iterações realizadas para cada um dos *corners* utilizados. A tabela apresenta então os *Acertos*, todos as células envelhecidas que foram detectadas como tal. Os *Erros Totais*, que correspondem a todas as detecções realizadas de forma errada. Por ultimo estes *Erros Totais* são apresentados de forma separada em *Erros - Células Velha*, que corresponde a todas as células envelhecidas que passaram pelo sensor sem serem detectadas e *Erros - Célula Jovem*, que corresponde a todas as células jovens que foram detectadas como envelhecidas.

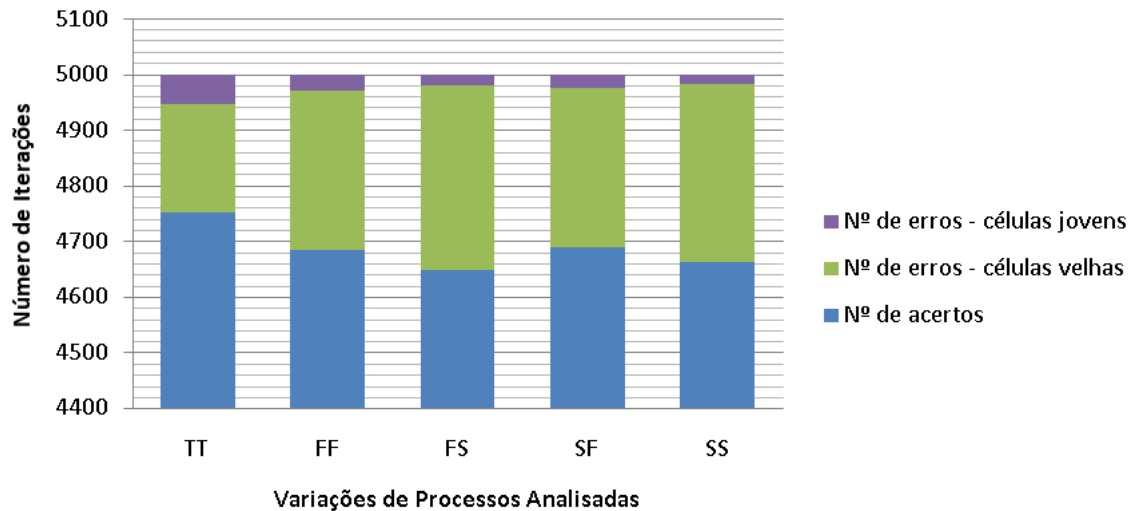
Na figura 34 são apresentados os dados para cada *corner* simulado através de uma abordagem gráfica. Nesta figura estão os mesmo dados apresentados na tabela 3 porém em barras gráficas que permitem uma melhor análise e comparação destes entre as diferentes variações de processos simuladas. Na figura 34 nota-se que a maior incidência de erros

Tabela 3 – Capacidade de detecção do sensor para todos os *corner* utilizados

<i>Corner</i>	TT	FF	FS	SF	SS
Acertos	4753	4684	4648	4690	4662
Erros Totais	247	316	352	310	338
Erros - Célula Velha	194	288	334	286	332
Erros - Célula Jovem	53	28	18	24	16

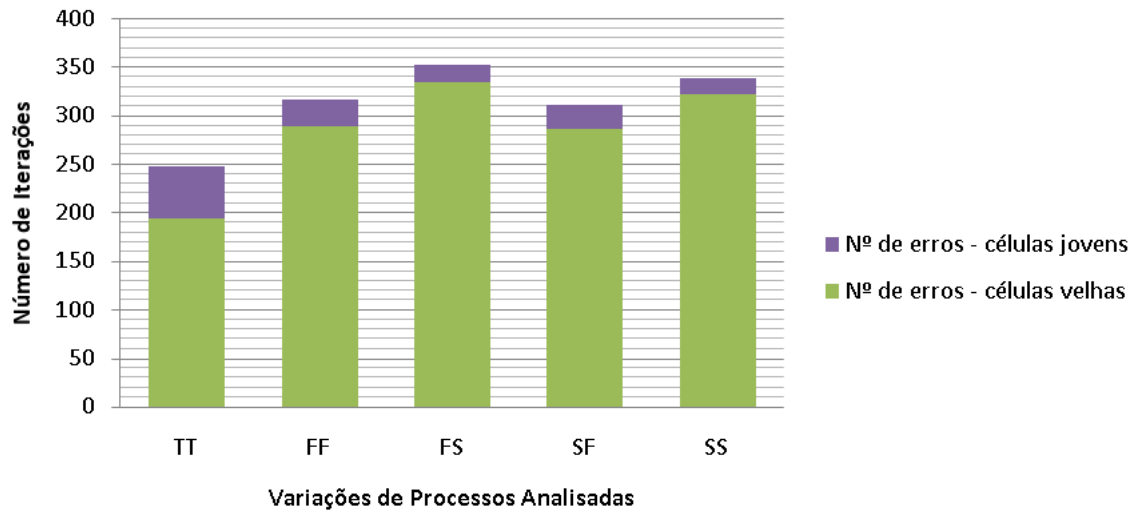
ocorreu no *corner* FS, enquanto que a menor incidência ocorreu, conforme era esperado, no *corner* TT.

Figura 34 – Gráfico dos resultados das simulações de Monte Carlo para todos os 5 *corners* simulados



A figura 35 exemplifica então os tipos de erros que foram encontrados no sensor através da técnica de Monte Carlo, decidiu-se separar estes erros em: *Nº de erros - células velhas*, são todas as células envelhecidas que deveriam ter sido detectadas pelo sensor e não foram; *Nº de erros - células jovens*, corresponde ao total de células jovens detectadas como sendo mais velhas do que o permitido pelo sensor. Nota-se pelo gráfico que a maior incidência de erros ocorreu a partir de células que mesmo envelhecidas passaram pela detecção do sensor.

Através dos dados levantados foi construída a tabela 4, que resume muito bem a necessidade de se trabalhar com uma referência (utilizada para classificar uma célula como jovem ou envelhecida) que sofra de forma igual com as variações de processo tanto

Figura 35 – Gráfico separando os tipos de falha

quanto a célula que estará sendo testada. Ainda é notado que a coluna de células é necessária, pois além da alteração dos *thresholds* dos transistores de *corner* para *corner* também existe alteração de outras variáveis que somadas alteram inclusive as capacitância das linhas de alimentação das colunas.

Tabela 4 – Funcionamento do Sensor Considerando Diferentes Variações de Processos.

Corner	TT	FF	SS	SF	FS
Vdd0 - 2 anos	1,0751V	1,0758V	1,0743V	1,0703V	1,0785V
Vdd0 - 3 anos	1,0762V	1,0767V	1,0754V	1,0718V	1,0795V
Vdd0 - 4 anos	1,0774V	1,0779V	1,0765V	1,073V	1,0804V
Vdd0 - 5 anos	1,0785V	1,0789V	1,0776V	1,0743V	1,0813V
Vdd4 REF	1,0765V	1,0772V	1,0755V	1,0718V	1,0797V

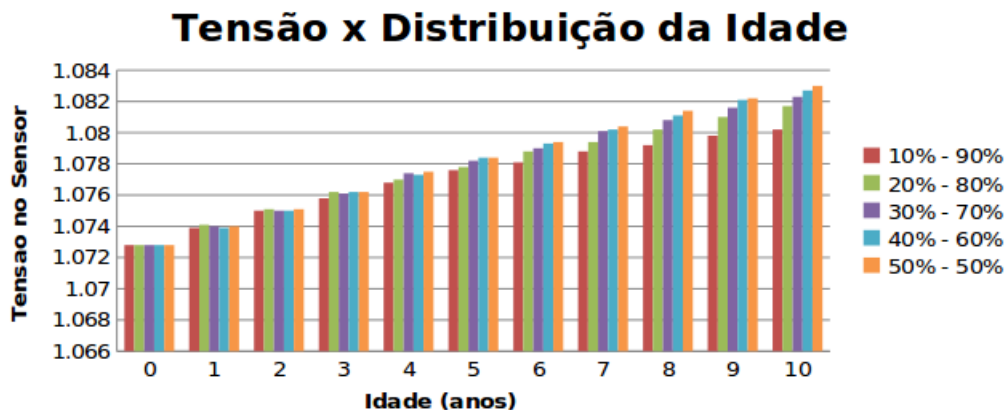
Através da tabela 4 nota-se, por exemplo, que se a referência fosse fixa uma célula com 5 anos localizada no *corner* FS poderia ser interpretada como sendo nova (com menos do que 4 anos). Quando se observa o *corner* TT, é possível observar que a tensão da célula de referência acaba por se adaptar de *corner* para *corner* e desta forma mantém a relação entre a idade e as tensões utilizadas pelo sensor para realizar a detecção.

5.2 Capacidade de Monitoramento

O trabalho então pretende detectar a idade de uma célula de memória, não levando em consideração qual foi a distribuição deste envelhecimento na célula, ou seja, não é capaz de determinar qual dos inversores da célula envelheceu mais. Desta forma técnicas que visam distribuir o envelhecimento das células de forma equilibrada podem ser associadas ao sensor para desta forma garantir um prolongamento da vida útil das memórias. Desta forma deve-se determinar qual a SNM mínima que se deseja garantir, e assim configurar o sensor para a detecção, levando em conta o comportamento da memória.

Na figura 36 percebe-se então a variação da curva de tensão que o sensor utiliza para realizar suas detecções em relação a distribuição do envelhecimento. Nesta figura tem-se uma distribuição que vai desde a pior possibilidade possível para uma célula 90% – 10% até a melhor relação de distribuição 50% – 50%.

Figura 36 – Tensão de Detecção x Distribuição do Envelhecimento.



5.3 Overhead

Nesta seção serão abordados alguns dos incrementos relacionados à implementação do sensor, os incrementos que serão abordados se referem ao aumento de área, do consumo de energia e do tempo de acesso ao bloco quando se utiliza um bloco de memória com o sensor inserido. Tais fatores foram levados em consideração durante o projeto do sensor sempre buscando a minimização destes.

5.3.1 Área

A área total ocupada pelo sensor foi estimada levando em conta o número de transistores utilizados e seus tamanhos. Note que não foi levada em consideração a área ocupada pelo roteamento. Torna-se importante salientar que em função do sensor utilizar uma coluna de células para servir como referência, o aumento do número de colunas durante o projeto do bloco de memórias que será utilizado acaba realizando uma diluição da área ocupada pelo sensor, pois para cada coluna adicionada ao bloco de memórias apenas um circuito comparador de tensão junto com algumas chaves é adicionado, em comparação, para cada linha adicionada ao bloco de memória tem-se uma célula de memória adicionada.

Tabela 5 – Custo de Área Associado ao Sensor

Tamanho do Bloco	Disposição <i>linha x coluna</i>	Custo <i>um²</i>	Acréscimo <i>em %</i>
4kbit	512 x 8	35,016	13,2
16kbit	1024 x 16	70,032	6,6
64kbit	2048 x 32	140,064	3,3

5.3.2 Performance

A velocidade do bloco SRAM não é afetada pelo sensor, pois o mesmo não diminui a tensão nas linhas de alimentação durante o funcionamento normal do bloco.

Convém mencionar que apenas durante a realização do teste é que a metodologia afeta ligeiramente a performance da SRAM, uma vez que operações de escrita são realizadas para que o sensor detecte o envelhecimento das células.

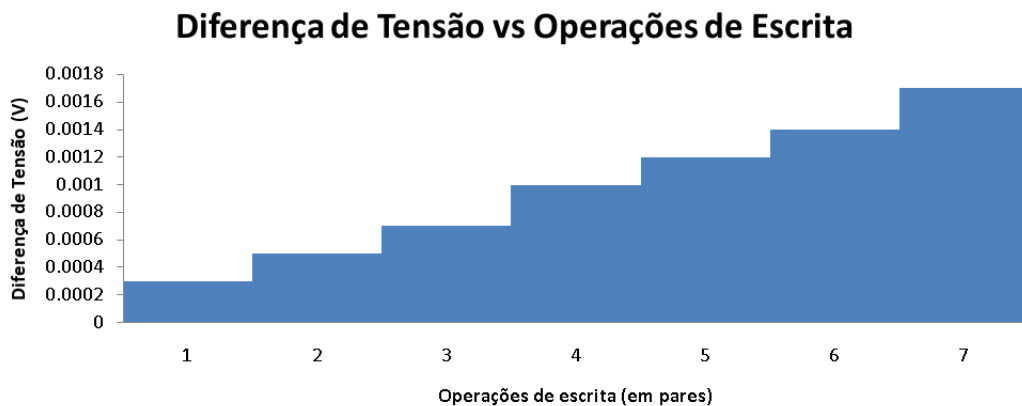
Afim de minimizar o tempo de detecção, decidiu-se por adicionar um sensor por coluna de memória, mantendo apenas uma coluna de referência para ser compartilhada em função da inviabilidade de se ter uma coluna por sensor visto seu consumo de área. Desta forma acaba sendo o número de escritas realizado por teste e seu período demandado que interfere no tempo total dispensado em cada varredura do bloco.

Além disto o tamanho do bloco acaba impactando no tempo de varredura não só pelo

maior número de células que deve ser analisado, mas também porque no momento que existem mais células por coluna, existe também mais capacitância por coluna. Como a avaliação do sensor quanto à idade das células esta ligada à tensão resultante no barramento de alimentação de cada coluna, para que se tenha uma diferença de tensão igual à obtida em bloco menor (colunas menores) deve-se aumentar o número de operações de escrita.

Verifica-se na figura 37 o aumento de tensão verificado para um bloco de 512 células por coluna quanto ao número de operações de escrita realizados, notar que conforme foi mencionado anteriormente o número de operações realizado deve sempre ser par, ou seja, ambos os inversores de cada célula devem ser testados de forma equivalente. As tensões encontradas na figura 37 se referem às tensões utilizadas pelo sensor para diferenciar uma "idade" da outra.

Figura 37 – Diferença de tensão entre idades vs. Número de operações de escrita.



Desta forma, a sensibilidade do sensor está diretamente ligada ao número de operações de escrita realizadas e também seu tempo de execução para testes.

Percebe-se pela tabela 6 que quanto maior o número de operações de escrita, maior será o tempo demandado para realizar a varredura do bloco de memória.

Através dos tempos totais gastos presentes na tabela 6, percebe-se que o tempo total gasto para cada detecção é resultado de uma detecção de 4 em 4 células em paralelo por linha. Isto ocorre porque, para diminuir o incremento de área, decidiu-se por utilizar

apenas uma coluna de células para todo o bloco, e desta forma esta é compartilhada entre os sensores. Em função disto, a corrente drenada da capacitância da coluna de referência por cada um dos sensores, quando somada, acaba alterando a leitura. Detectou-se uma divergência de leitura de 1 ano à mais (células parecem ter mais idade do que realmente tem) em função da detecção em paralelo de 8 células.

Tabela 6 – Tempo de Detecção

Frequência	Período de Escrita
500MHz	<i>2ns</i>
Bloco: 8 colunas	512 linhas
Nº de Operações	Tempo Gasto (s)
2	4,10e-6
4	8,19e-6
6	1,23e-5
8	1,64e-5
10	2,05e-5
12	2,46e-5
14	2,87e-5

5.3.3 Energia

Observa-se pela tabela 7 que a diferença de potência média entre as técnicas praticamente se manteve inalterada. Enquanto isso a máxima potência apresentou um acréscimo pequeno, de somente 3,8%, isto prova que o sensor não acrescenta significativo consumo ao bloco, apenas sua operação é que obviamente acaba por acrescentar, pois cada operação de escrita resulta em consumo do bloco.

Tabela 7 – Consumo de Energia

	Bloco com o Sensor	Bloco Sem o Sensor	Acréscimo
Potência Média (W)	1,971E-03	1,970E-03	1%
Potência Máxima (W)	6,930E-03	6,670E-03	3,79%

6 CONCLUSÃO

O presente trabalho teve como objetivo o monitoramento da idade de células de SRAM submetidas ao envelhecimento. Foi pesquisado e desta forma concluído que o principal causador deste envelhecimento é o fenômeno do NBTI, desta forma passou-se a estudar suas causas e seus efeitos sobre a célula de memória para que assim fosse escolhida a melhor forma de se buscar detectar o envelhecimento.

Observou-se então que existe uma perda da capacidade de condução dos pMOS submetidos ao NBTI, isso se deve ao aumento de 10% à 15% do seu valor de *threshold* para cada ano de utilização (estresse). Sabendo-se desta consequência associada ao NBTI, decidiu-se por monitorar o consumo dinâmico das células durante as operações de escrita. Observando-se assim que, para cada ano de envelhecimento ocorria uma diminuição do consumo dinâmico e desta forma se utilizou desta informação para determinar o quão velha estava a células testada.

A proposta visou monitorar o valor de tensão resultante na capacitância das linhas de alimentação das colunas de células. E para avaliar uma célula como envelhecida ou não, este valor de tensão foi comparado com o valor de tensão encontrado na linha de alimentação de uma coluna de células usada apenas como referência. Esta coluna, foi resguardada dos efeitos do envelhecimento durante todo o tempo que não foi utilizada para a comparação através do uso de *power-gate*.

O trabalho teve ainda o cuidado de distribuir de formas distintas este envelhecimento entre os dois transistores pMOS que compõe as células de memória para verificar como isto afetava a leitura do consumo das células. Foi concluído então que mesmo que um

dos pMOS seja estressado durante 90% do seu tempo de vida (oque faz com que o outro pMOS seja envelhecido apenas durante 10% de toda sua vida) ainda assim, a leitura do envelhecimento da célula, uma vez que é realizada mais de uma operação de escrita em cada medição, foi correta. Isto ocorreu em função de que o pMOS que envelheceu durante a maior parte do tempo consegue complementar a diferença criada por aquele que envelheceu menos.

Para mensurar o quanto o envelhecimento das células impactou na estabilidade das células, foi pesquisada uma medida de qualidade referente a tensão de um ruído capaz de inverter o valor armazenado. Esta medida, conhecida como SNM, se mostrou muito útil e capaz de determinar a estabilidade das células SRAM do tipo 6T para cada etapa do envelhecimento. Foi através desta que se pôde avaliar o real impacto do envelhecimento em uma célula.

Desta forma, o trabalho foi capaz de atingir seu objetivo, de criar um sensor capaz de detectar o envelhecimento de células de SRAM, capaz de suportar variações de processo de uma forma adaptativa (pois a referência acaba oscilando juntamente com o restante das células) e ainda de medir o impacto nas células de memória e evitar então o uso de células já definidas como envelhecidas.

7 *TRABALHOS FUTUROS*

Como trabalho futuro, pode-se sugerir a elaboração de um sensor capaz de detectar a tensão entre a coluna de referência e a coluna de células que está sendo testada. Desta forma o uso de um grupo de capacitores, no momento necessário para que se indique qual a idade que o sensor deve detectar, poderá ser deixado de lado, e assim pode-se evitar os problemas relacionados à esta "calibração" (escolha do número de capacitores que será colocado em paralelo). Assim, apenas seria indicada a diferença de tensão entre as colunas e desta forma se saberia qual o envelhecimento em função deste valor.

REFERÊNCIAS

- [1] “International technology roadmap for semiconductors (itrs),” *Semiconductor Industry Association (SIA)*, 2003.
- [2] K. Kang, H. Kufluoglu, K. Roy, and M. Ashraful Alam, “Impact of negative-bias temperature instability in nanoscale sram array: Modeling and analysis,” *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 26, oct. 2007.
- [3] R. Vattikonda, W. Wang, and Y. Cao, “Modeling and minimization of pmos nbtı effect for robust nanometer design,” in *Design Automation Conference, 2006 43rd ACM/IEEE*, pp. 1047–1052, 0-0 2006.
- [4] F. J. List, “The static noise margin of sram cells,” in *Solid-State Circuits Conference, 1986. ESSCIRC '86. Twelfth European*, pp. 16–18, sept. 1986.
- [5] Z. Qi, J. Wang, A. Cabe, S. Wooters, T. Blalock, B. Calhoun, and M. Stan, “Sram-based nbtı/pbtı sensor system design,” in *Design Automation Conference (DAC), 2010 47th ACM/IEEE*, pp. 849–852, june 2010.
- [6] F. Ahmed and L. Milor, “Reliable cache design with on-chip monitoring of nbtı degradation in sram cells using bist,” in *VLSI Test Symposium (VTS), 2010 28th*, pp. 63–68, april 2010.
- [7] K. Kang, M. Alam, and K. Roy, “Characterization of nbtı induced temporal performance degradation in nano-scale sram array using iddq,” in *Test Conference, 2007. ITC 2007. IEEE International*, pp. 1–10, oct. 2007.
- [8] C. Ferri, D. Papagiannopoulou, R. Bahar, and A. Calimera, “Nbtı-aware data allocation strategies for scratchpad memory based embedded systems,” in *Test Workshop (LATW), 2011 12th Latin American*, pp. 1–6, march 2011.
- [9] K. Kang, S. Gangwal, S. P. Park, and K. Roy, “Nbtı induced performance degradation in logic and memory circuits: How effectively can we approach a reliability solution?,” in *Proceedings of the 2008 Asia and South Pacific Design Automation Conference, ASP-DAC '08*, (Los Alamitos, CA, USA), pp. 726–731, IEEE Computer Society Press, 2008.
- [10] D. K. Schroder and J. A. Babcock, “Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing,” *Journal of Applied Physics*, vol. 94, no. 1, pp. 1–18, 2003.
- [11] W. Wang, V. Reddy, A. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao, “Compact modeling and simulation of circuit reliability for 65-nm cmos technology,” *Device and Materials Reliability, IEEE Transactions on*, vol. 7, pp. 509–517, dec. 2007.

- [12] A. Calimera, E. Macii, and M. Poncino, "Analysis of nbtI-induced snm degradation in power-gated sram cells," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pp. 785–788, 30 2010-june 2 2010.
- [13] A. Calimera, E. Macii, and M. Poncino, "NbtI-aware power gating for concurrent leakage and aging optimization," in *Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design*, ISLPED '09, (New York, NY, USA), pp. 127–132, ACM, 2009.
- [14] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation*. Wiley-IEEE Press, 3rd ed., 2010.
- [15] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of the surface-state charge (q_{ss}) of thermally oxidized silicon," *Journal of The Electrochemical Society*, vol. 114, no. 3, pp. 266–274, 1967.
- [16] M. Bagatin, S. Gerardin, A. Paccagnella, and F. Faccio, "Impact of nbtI aging on the single-event upset of sram cells," *Nuclear Science, IEEE Transactions on*, vol. 57, pp. 3245–3250, dec. 2010.
- [17] D. K. and Schroder, "Negative bias temperature instability: What do we understand?," *Microelectronics Reliability*, vol. 47, no. 6, pp. 841 – 852, 2007. <ce:title>Modelling the Negative Bias Temperature Instability</ce:title>.
- [18] J. Massey, "NbtI: What we know and what we need to know - a tutorial addressing the current understanding and challenges for the future," in *Integrated Reliability Workshop Final Report, 2004 IEEE International*, pp. 199 – 211, oct. 2004.
- [19] M. Alam, "A critical examination of the mechanics of dynamic nbtI for pmosfets," in *Electron Devices Meeting, 2003. IEDM '03 Technical Digest. IEEE International*, pp. 14.4.1 – 14.4.4, dec. 2003.
- [20] A. Calimera, E. Macii, and M. Poncino, "NbtI-aware clustered power gating," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 16, pp. 3:1–3:25, November 2010.
- [21] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Impact of nbtI on sram read stability and design for reliability," in *Proceedings of the 7th International Symposium on Quality Electronic Design*, ISQED '06, (Washington, DC, USA), pp. 210–218, IEEE Computer Society, 2006.
- [22] A. Pavlov and M. Sachdev, *CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies: Process-Aware SRAM Design and Test*. Springer Publishing Company, Incorporated, 1st ed., 2008.
- [23] B. Shilpi, S. R.K., and M. Pattnaik, "Static noise margin analysis of various sram topologies," vol. 3, pp. 304–309, june 2011.
- [24] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of mos sram cells," *Solid-State Circuits, IEEE Journal of*, vol. 22, pp. 748 – 754, oct 1987.
- [25] A. Bhavnagarwala, X. Tang, and J. Meindl, "The impact of intrinsic device fluctuations on cmos sram cell stability," *Solid-State Circuits, IEEE Journal of*, vol. 36, pp. 658 –665, apr 2001.

- [26] B. Calhoun and A. Chandrakasan, “Analyzing static noise margin for sub-threshold sram in 65nm cmos,” in *Solid-State Circuits Conference, 2005. ESSCIRC 2005. Proceedings of the 31st European*, pp. 363 – 366, sept. 2005.
- [27] J. Vazquez, V. Champac, I. Teixeira, M. Santos, and J. Teixeira, “Programmable aging sensor for automotive safety-critical applications,” in *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, pp. 618 –621, march 2010.
- [28] A. Ricketts, J. Singh, V. Narayanan, and D. Pradhan, “Investigating the impact of nbtI on different power saving cache strategies,” in *Proceedings of IEEE International Conference, DATE, 2010, Germany*, DATE 2010, March 2010.
- [29] H. Singh and H. Mahmoodi, “Analysis of sram reliability under combined effect of nbtI, process and temperature variations in nano-scale cmos,” in *Future Information Technology (FutureTech), 2010 5th International Conference on*, pp. 1 –4, may 2010.
- [30] A. Moshovos, B. Falsafi, F. Najm, and N. Azizi, “A case for asymmetric-cell cache memories,” *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 13, pp. 877 –881, july 2005.
- [31] Y.-J. Chang and F. Lai, “Dynamic zero-sensitivity scheme for low-power cache memories,” *IEEE Micro*, vol. 25, pp. 20–32, July 2005.

APÊNDICE A - EXEMPLO DE ESTIMAÇÃO DO VALOR DE SNM PARA UMA CÉLULA SRAM

Nas próximas páginas será apresentado o modelamento utilizado no HSPICE para a obtenção dos valores relativos a SNM de uma célula. Esta análise é importante pois disponibiliza o máximo ruído suportado por uma célula antes que esta tenha seu valor invertido durante uma leitura em função de um ruído DC.

```
.include './tech/corners_TT'
```

```
*-----INVERSOR A-----
```

```
.subckt INV_bitb bit1 bitb vdd gnd vth1=0
```

```
*Nome S G D B
```

```
XM1 bitb x vdd vdd PSVTLP L=0.06 W=0.22
```

```
XM2 gnd bit1 bitb gnd NSVTLP L=0.06 W=0.20
```

```
v1 x bit1 dc vth1 *-----fonte para nbt1-----
```

```
.param vth1=0
```

```
.ends INV_bitb
```

```
*-----INVERSOR B-----
```

```

.subckt INV_bit bitb1 bit vdd gnd vth2=0
*Nome S    G    D    B
XM1 bit y vdd vdd PSVTLP L=0.06 W=0.22
XM2 gnd bitb1 bit gnd NSVTLP L=0.06 W=0.20
v1 y bitb1 dc vth2 *-----fonte para nbt-----
.param vth2=0
.ends INV_bit

vfonte vdd gnd dc 1.1V
vin nvin gnd dc 1.1v
.dc vin -1.1V 1.1V 0.01V

x1 bit1 bitb1 vdd gnd INV_bitb
x2 bitb1 vdd vdd gnd NSVTLP L=0.06 W=0.12 *TRANSISTOR DE ACESSO
*-----
* +/-polinomial nós referência P0 P1
Eou1 nvou1 gnd POLY(2) bitb1 0 nvin 0 0 'Sqrt(2)' 1 0 0 0
Ei1a bit1 bitm1 POLY(1) nvou1 0 0 '1/Sqrt(2)'
Ei1b bitm1 gnd POLY(1) nvin 0 0 '1/Sqrt(2)'
*-----
x3 bit2 bitb2 vdd gnd INV_bit
x4 bitb2 vdd vdd gnd NSVTLP L=0.06 W=0.12 *TRANSISTOR DE ACESSO

* +/-polinomial nós referência P0 P1
Eou2 nvou2 gnd POLY(2) bitb2 0 nvin 0 0 'Sqrt(2)' -1 0 0 0
Ei2a bit2 bitm2 POLY(1) nvou2 0 0 '1/Sqrt(2)'
Ei2b bitm2 gnd POLY(1) nvin 0 0 '1/Sqrt(2)*(-1)'

```



```
.probe dc u=v(nvin) inv1in=v(bit1) inv2in=v(bit2) inv1out=v(bitb1) inv2out=v(bitb2)

.MEAS dc SNM1 MAX PAR('v(nvou1)-v(nvou2)') FROM=-0.660V TO=0.660V
.MEAS dc SNM2 MIN PAR('v(nvou1)-v(nvou2)') FROM=-0.660V TO=0.660V
.MEAS dc nbtivTH1 MAX PAR(vth1) FROM=-0.660V TO=0.660V
.MEAS dc nbtivTH2 MAX PAR(vth2) FROM=-0.660V TO=0.660V

.connect gnd 0

.temp 27

.option post probe
.option MEASFILE=1
.param step='0.043'
.param vth1= 0
.param vth2= 0
.alter
.param vth1='step * 1'
.param vth2='step * 1'
.alter
.param vth1='step * 2'
.param vth2='step * 2'
.alter
.param vth1='step * 3'
.param vth2='step * 3'
.alter
.param vth1='step * 4'
.param vth2='step * 4'
```

```
.alter  
.param vth1='step * 5'  
.param vth2='step * 5'  
.alter  
.param vth1='step * 6'  
.param vth2='step * 6'  
.alter  
.param vth1='step * 7'  
.param vth2='step * 7'  
.alter  
.param vth1='step * 8'  
.param vth2='step * 8'  
.alter  
.param vth1='step * 9'  
.param vth2='step * 9'  
.alter  
.param vth1='step * 10'  
.param vth2='step * 10'  
.end
```

APÊNDICE B – MODELAMENTO HSPICE DA SIMULAÇÃO DE MONTE CARLO

Neste anexo estão os passos necessários para que se possa iniciar uma sequência de simulações através da técnica de Monte Carlo.

```
.param max='0.43' * 10% de deslocamento ao ano para
*apenas um transistor envelhecendo
```

```
.PARAM val=5000 *val é o número de interações que serão realizadas
.OPTION MCBRIEF=1 * (0 coloca tudo, 1 não coloca nenhum parâmetro de monte
* carlo na saída, coloca os parâmetros num arquivo lis)
.OPTION MODMONTE=1 *SERVE PARA QUE O SIMULADOR RETORNE APENAS OS VALORES
*NECESSÁRIOS
```

```
.TRAN 0.1n 42n SWEEP MONTE=val
```

```
.PARAM age1=UNIF(0.5, 0.5) * DISTRIBUIÇÃO DOS VALORES RELACIONADOS À
* IDADE DA CÉLULA
.PARAM dis1=GAUSS(0.5, 0.5, 3) * DISTRIBUIÇÃO DA DISTRIBUIÇÃO DO ENVELHECIMENTO
* ENTRE OS INVERSORES DA CÉLULA
.PARAM idade = 'age1*max'
```

```
.param comple = '(1-dis1)'
```

```
.param real_aging1 = 'idade'
```

```
.param NBTI2_bit=0
```

```
.param NBTI2_bitb=0
```

```
.param NBTI1_bit = 'dis1*idade'
```

```
.param NBTI1_bitb= 'comple*idade'
```

```
.MEAS tran Age min PAR('real_aging1') FROM=38.7ns TO=39ns *MEDIÇÃO DA IDADE
```

```
.MEAS tran Dist MAX PAR('dis1') FROM=38.7ns TO=39ns *MEDIÇÃO DA DISTRIBUIÇÃO
```

```
.MEAS tran Out max v(xsensor0.outinv1,gnd0) FROM=40ns TO=41.7ns *MEDIÇÃO DE SAÍDA
```

APÊNDICE C – PUBLICAÇÕES REALIZADAS

Temos abaixo as publicações que foram realizadas através do trabalho realizado nesta dissertação.

LATW 2012: Investigating the Use of an On-Chip Sensor to Monitor NBTI Effect in SRAM

DDECS 2012: On-Chip Aging Sensor to Monitor NBTI Effect in Nano-Scale SRAM