

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROELETRÔNICA

MAURÍCIO BANASZESKI DA SILVA

**Circuito On-Chip para a Caracterização em Alta Escala do Efeito de
Bias Temperature Instability**

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em
Microeletrônica

Prof. Dr. Gilson Inácio Wirth
Orientador

Porto Alegre, Março de 2012.

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Banaszeski da Silva, Maurício

Circuito On-Chip para a Caracterização em Alta Escala do Efeito de Bias Temperature Instability [manuscrito] / Maurício Banaszkeski da Silva. – 2012.

15 f.:il.

Orientador: Gilson Inácio Wirth.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Microeletrônica. Porto Alegre, BR – RS, 2012.

1.BTI 2.Caracterização em alta-escala 3.Transistores nano métricos. . I. Wirth, Gilson Inácio. III. Circuito On-Chip para a Caracterização em Alta Escala do Efeito de Bias Temperature Instability.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PMICRO: Prof. Ricardo Reis

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais e irmã pelo apoio incondicional dado, sem qual seria extremamente difícil a jornada ao longo desses dois anos.

Gostaria de expressar minha gratidão para com todos os funcionários e professores da Pós Graduação em Microeletrônica desta Universidade pelo conhecimento passado e pelo apoio. Gostaria de agradecer em especial ao Professor Doutor Gilson I. Wirth pela confiança depositada como orientador desse projeto e principalmente por todo incentivo dado ao longo do Mestrado.

Gostaria de agradecer, também, a todos os funcionários e pesquisadores do IMEC, Bélgica, com os quais convivi por um ano, agradecendo em especial a Ben Kaczer e a Geert Van der Plas pelo incentivo e conhecimento passado.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	5
LISTA DE FIGURAS	6
LISTA DE TABELAS	8
RESUMO	9
ABSTRACT	10
1 INTRODUÇÃO	11
2 BIAS TEMPERATURE INSTABILITY (BTI)	13
2.1 O que é?	13
2.2 Origens	15
2.2.1 Modelo Reaction-Difusion	15
2.2.2 Modelo Trapping/Detrapping	16
2.3 PMOS vs NMOS	17
2.4 Impacto de BTI nos circuitos digitais	18
2.5 Variabilidade	20
3 CARACTERIZAÇÃO ELÉTRICA DE BTI	22
3.1 Extraíndo V_{th}	22
3.1.1 Método da corrente constante	22
3.1.2 Extrapolação da região linear	23
3.1.3 Segunda derivada.....	24
3.2 Extraíndo ΔV_t devido ao BTI	25
3.2.1 Rápida medida de I_d	26
3.2.2 Rápida medida de V_g	27
3.2.3 Método On the Fly.....	28
3.2.4 Extrapolação da Recuperação.....	30
3.3 Automatizando caracterização de BTI	31
4 CIRCUITO PROPOSTO	35
4.1 Motivação	35
4.2 Idéia Central	35
4.2.1 Detalhamento das Chaves.....	37
4.2.2 Detalhamento do multiplexador	38
4.2.3 Método <i>Force and Sense</i> (Kelvin)	38
4.3 Controlando o estresse e a medida	39
4.4 Circuito Final	43
4.5 Uso do circuito para caracterização automatizada	46
4.5.1 Estado 1 – Preparação para medida de V_{th}	46
4.5.2 Estado 2 – Medida de V_{th}	47
4.5.3 Estado 3 – Preparação para o estresse	48
4.5.4 Estado 4 – Início do estresse	49
4.5.5 Estado 5 - Estresse	50
4.5.6 Estado 6 – Medida de ΔV_{th}	50
4.5.7 Tempos de <i>Hold</i> e <i>Setup</i>	51
5 CONSIDERAÇÕES DO LEIAUTE	53
5.1 Célula Básica	53
5.2 Célula BTI	55
5.3 Circuito Completo	57

5.4	Seis Módulos BTI	58
5.5	Benefícios do circuito	60
6	SIMULAÇÕES	61
6.1	Caracterização do Circuito	61
6.2	Medindo V_{th}	63
6.3	Caracterizando BTI	63
7	CONCLUSÃO	66
	REFERÊNCIAS	67

LISTA DE ABREVIATURAS E SIGLAS

BTI	Bias Temperature Instability
DIBL	Drain Induced Barrier Lowering
DF	Duty Factor
DUT	Device Under Test
HCI	Hot Carrier Injection
HK	High-k
MG	Metal Gate
MOSFET	Metal Oxide Silicon Field Effect Transistor
NMOS	N-channel Metal Oxide Silicon
NBTI	Negative Bias Temperature Instability
PBTI	Positive Bias Temperature Instability
PMOS	P-channel Metal Oxide Silicon
PTM	Predictive Technology Model

LISTA DE FIGURAS

Figura 2.1 Degradação de V_{th} com o tempo, mostrando claramente um comportamento $\log(t)$	13
Figura 2.2 Dependências do efeito BTI com a) Campo elétrico, b) temperatura.	14
Figura 2.3 Ilustração do fenômeno de NBTI dinâmico, o período de relaxação recupera o V_{th} logo após a retirada do sinal de estresse.	15
Figura 2.4 Impacto da a) frequência e b) do <i>duty factor</i> no efeito de BTI	15
Figura 2.5 Modelo <i>Reaction-Difusion</i>	16
Figura 2.6 Diferença entre a teoria Reaction-Diffusion e as medidas práticas.	16
Figura 2.7 Ilustração da origem de BTI dada pela captura de cargas.	17
Figura 2.8 Diferenças entre NBTI e PBTI em ambos transistores PMOS e NMOS.	18
Figura 2.9 Exemplo do cálculo da probabilidade de estresse dos transistores de um circuito.	19
Figura 2.10 Impacto do dimensionamento na degradação do atraso máximo de um inversor, melhor relação W_p/W_n variando após degradação.	19
Figura 2.11 a) <i>Percolation Path</i> criado pelos dopantes do canal e b) obstrução do caminho por uma armadilha ocupada. Figura adaptada de Kaczer, 2010.	21
Figura 2.12 Diminuição do número de defeitos dada pela miniaturização dos transistores a) Desvio de V_{th} em transistores com 800 armadilhas. b) Desvio de V_{th} em transistores com 12 armadilhas.	21
Figura 3.1 Método da Corrente Constante	23
Figura 3.2 Método da extrapolação linear	24
Figura 3.3 Método da Segunda Derivada	25
Figura 3.4 O transistor é polarizado como mostrado enquanto a corrente I_d é monitorada.	26
Figura 3.5 I_d medido é traduzido em ΔV_{th} deslocando horizontalmente a curva $I_d \times V_g$ inicial.	27
Figura 3.6: Circuito utilizado no método da rápida medida de V_g	28
Figura 3.7: Curvas I-V antes e depois do estresse, linhas sólidas representam medidas e linhas pontilhadas representam a equação (2.2).	29
Figura 3.8 Exemplo das medidas realizadas durante o período de recuperação.	30
Figura 3.9 Extrapolação realizadas utilizando as medidas realizadas na recuperação. (KACZER, 2008).....	31
Figura 3.10 Esquemático do circuito apresentado em	32
Figura 3.11 Esquemático do circuito de medida.	32
Figura 3.12 Degradação do oscilador dada em função do valor de saída N do circuito.	33
Figura 3.13 Célula baseado em uma combinação NAND-NOR.	33
Figura 4.1 Idéia básica para a célula da matriz com DUT do tipo PMOS	36
Figura 4.2 Esquemático da chave utilizada.	37
Figura 4.3 Esquemático do multiplexador utilizado.	38
Figura 4.4 Demonstração do uso do método Force and Sense.	39
Figura 4.5 Dois comportamentos diferentes para estresse e medida	39
Figura 4.6 Erro médio de medida para 1000 transistores simulados	40
Figura 4.7 Novo comportamento proposto para o estresse e a medida.	41
Figura 4.8 Comportamento do circuito e seus 6 estados	42
Figura 4.9 Mapa de Karnaugh	42

Figura 4.10 Célula BTI para um DUT do tipo PMOS	43
Figura 4.11 Bloco dos <i>shift registers</i>	44
Figura 4.12 Exemplo da troca de linha quando em estresse.	44
Figura 4.13 Funcionamento do <i>flip-flop</i> do tipo FF1	45
Figura 4.14 Sinais externos e internos para uma matriz 3x3	46
Figura 4.15 Comportamento dos sinais para definir nível alto em todas linhas e colunas.	47
Figura 4.16 Comportamento dos sinais para medir V_{th} em todos DUTs sequencialmente	48
Figura 4.17 Comportamento dos sinais para definir nível baixo em todas linhas e colunas.	49
Figura 4.18 Comportamento dos sinais para colocar todos DUTs sob estresse.	50
Figura 4.19 Comportamento dos sinais para medir ΔV_{th} sequencialmente em todos os DUTs.	51
Figura 4.20 Ilustração do tempo necessário a se esperar para mudar Row1 após a mudança de CK2.	52
Figura 5.1 a) Célula Básica completa e b) sem algumas camadas. c) Menu para alteração de parâmetros manualmente.	54
Figura 5.2 Comportamento projetado ao se colocar células básicas lado-a-lado.	54
Figura 5.3 Exemplo de uma NAND2 criada a partir de uma Célula Básica.	55
Figura 5.4 Vista completa de todas camadas da Célula BTI.	56
Figura 5.5 Vista das camadas de conexão intermediária da Célula BTI	56
Figura 5.6 Vistas das conexões da Célula BTI.	56
Figura 5.7 Disposição dos blocos no circuito completo. Exemplo com uma matriz 3x3.	57
Figura 5.8 Visão geral do circuito completo. Exemplo com uma matriz 3x3.	58
Figura 5.9 a) Circuito completo 30x3 b) disposição de 2 circuitos 30x3 no módulo	58
Figura 5.10 Seis módulos utilizando totalizando 1024 DUTs.	59
Figura 6.1 ΔV_{th} encontrado para os primeiros 15 transistores da matriz.	64
Figura 6.2 Corrente I_d medida em V_{force}	65

LISTA DE TABELAS

Tabela 2.1: diferentes dimensionamentos e seu impacto no atraso após 8 anos de degradação	20
Tabela 3.1: Funcionamento da célula proposta	33
Tabela 5.1: Geometrias utilizadas e suas localizações.	59
Tabela 6.1: Tempos de <i>setup</i> e <i>hold</i> para os <i>flip-flops</i> das colunas.	61
Tabela 6.2: tempos de <i>hold</i> para os <i>flip-flops</i> das linhas.	62
Tabela 6.3: Caracterização do atraso e do <i>slope</i> nos sinais internos do circuito.	62
Tabela 6.4: Comparação entre os métodos de medida de V_{th} para DUTs na Matriz e Isolados.	63

RESUMO

O trabalho propõe um circuito para caracterização estatística do fenômeno *Bias Temperature Instability* (BTI). O circuito tem como base uma matriz de transistores para caracterização eficiente em larga escala de BTI. O design proposto visa o estudo da variabilidade de BTI dependente do tempo em dispositivos altamente miniaturizados. Para tanto se necessita medir centenas de dispositivos, a fim de se obter uma amostra estatisticamente significativa. Uma vez que variações nos tempos de estresse e medida dos dispositivos podem gerar erros no processo de caracterização, o circuito implementa em chip (*on-chip*) o controle dos tempos de estresse e de medida, para que ocorra uma caracterização estatística precisa. O circuito de controle implementado faz com que todos os dispositivos testados tenham os mesmos tempos de estresse e os mesmos tempos de recuperação (relaxamento). Desta forma, o circuito proposto melhora significativamente tanto a área utilizada quanto o tempo de medida, quando comparado a alternativas anteriormente implementadas. O leiaute do circuito foi realizado no novo nó tecnológico de 28 nanômetros do IMEC.

Palavras-Chave: *Bias Temperature Instability* (BTI), Caracterização em Alta-escala, Transistores Nano Métricos, Confiabilidade, Degradação de Performance.

On-Chip Circuit for Massively Parallel BTI Characterization

ABSTRACT

This work proposes an array-based evaluation circuit for efficient and massively parallel characterization of Bias Temperature Instability (BTI). This design is highly efficient when studying the BTI time-dependent variability in deeply-scaled devices, where hundreds of devices should be electrically characterized in order to obtain a statistically significant sample size. The circuit controls stress and measurement times for accurate statistical characterization, making sure all the devices characterized have the same stress and recovery times. It significantly improves both area and measurement time. The circuit layout is laid out in the new 28nm node IMEC technology.

Keywords: Bias Temperature Instability (BTI), Large-scale Characterization, Deeply Scaled Transistors, Reliability, Performance Degradation (Aging).

1 INTRODUÇÃO

Novos avanços no desempenho dos dispositivos integrados são principalmente devidos a melhoria da tecnologia, proporcionando o dimensionamento cada vez menor dos transistores. Tais avanços como a redução da potência consumida, cálculos mais rápidos e a integração de mais transistores permitem a existência de memórias maiores e a integração de novas funcionalidades ao chip.

Com o dimensionamento atingindo escalas nano-métricas, as variações *on-chip* devido a flutuações estocásticas no processo de fabricação começaram a tornar as respostas do circuito (ex. o atraso de uma porta lógica) aleatoriamente distribuídas (ASHRAF, 2011). Variabilidades causadas no processo e suas incertezas estão cada vez mais presentes em todos os níveis do fluxo de projeto e também impactam a confiabilidade dos sistemas. Logo, a confiabilidade dos dispositivos se tornou um aspecto crítico da tecnologia CMOS.

Nas tecnologias mais recentes, o tempo de vida e a confiabilidade dos circuitos integrados estão sendo afetados pela degradação das características elétricas dos transistores com o tempo. As características físicas dos transistores também podem sofrer degradações significativas, causando mudanças estocásticas e determinísticas nos parâmetros elétricos do transistor, especialmente na tensão de limiar (V_{th}).

Um efeito de degradação que tem recebido atenção é o efeito chamado de *Bias Temperature Instability* ou BTI. Com a tensão de alimentação sendo dimensionada menos agressivamente em comparação com as dimensões do dispositivo, dispositivos diminutos estão sujeitos a altas densidades de corrente e temperaturas. Adicionalmente, o dimensionamento cada vez menor da espessura do óxido dos transistores resulta em campos elétricos altíssimos. As altas temperaturas e altos campos elétricos originam o efeito BTI fazendo com que a tensão limiar do transistor se desvie gradualmente para maiores valores absolutos (JEPPSON, 1977). Assim, aumentando a incerteza e degradando o desempenho do circuito.

É sabido que BTI é causado pelo aumento no número de armadilhas ou ligações quebradas no dielétrico do transistor; o desvio da tensão de limiar ocorre quando uma carga é capturada por um desses defeitos (GRASSER, 2011). Uma distribuição estatística do número de defeitos, posição e propriedades (magnitude do desvio do V_{th} , constante de tempo de captura e de emissão) e o fato de com a miniaturização do transistor o número absoluto de defeitos diminuem, são indicadores que uma pequena amostra de transistores não é mais adequada para prever o envelhecimento de bilhões de pequenos dispositivos usados nas aplicações digitais, que utilizam o estado da arte na tecnologia de miniaturização dos transistores. Adicionalmente, o custo e tempo de caracterização de grandes amostras de transistores para o teste de circuitos complexos com baixa taxa de falhas impõem um sério desafio ao uso de futuras tecnologias.

Este trabalho objetiva contribuir para a uma eficiente caracterização e análise estatística de BTI. Para tanto, primeiramente foi realizado um estudo sobre as causas do fenômeno e seus efeitos na degradação de dispositivos e circuitos integrados. Em seguida realizou-se o estudo sobre os métodos e circuitos existentes para a caracterização de BTI.

A partir disto, iniciou-se o desenvolvimento de um circuito que atendesse aos requisitos de caracterização especificados. O circuito desenvolvido permite que centenas de dispositivos sejam caracterizados de forma automática e paralela, sendo altamente eficiente na caracterização da componente estatística de BTI.

Nesse trabalho também são abordadas as técnicas de leiaute utilizadas no desenvolvimento do circuito no nó tecnológico de 28 nanômetros do IMEC, bem como as simulações realizadas com o *netlist* pós-leiaute.

2 BIAS TEMPERATURE INSTABILITY (BTI)

2.1 O que é?

Bias Temperature Instability, ou apenas BTI, é um fenômeno de degradação ou envelhecimento que afeta a confiabilidade dos transistores MOSFETs. O efeito degrada a tensão de limiar (V_{th}) dos transistores fazendo seu valor aumentar em módulo de forma logarítmica com o tempo (WIRTH, 2011), como mostrado na Figura 2.1.

Na tecnologia CMOS, a qual é largamente empregada no projeto de circuitos digitais, as tensões necessárias para que os transistores NMOS e PMOS conduzam corrente são respectivamente positivas e negativas. Por essa razão esse efeito nos transistores NMOS é chamado de *Positive Bias Temperature Instability* (PBTI) e nos transistores PMOS a degradação é chamada de *Negative Bias Temperature Instability* (NBTI). Porém seu impacto é usualmente de maior intensidade nos transistores PMOS (SCHRODER, 2007).

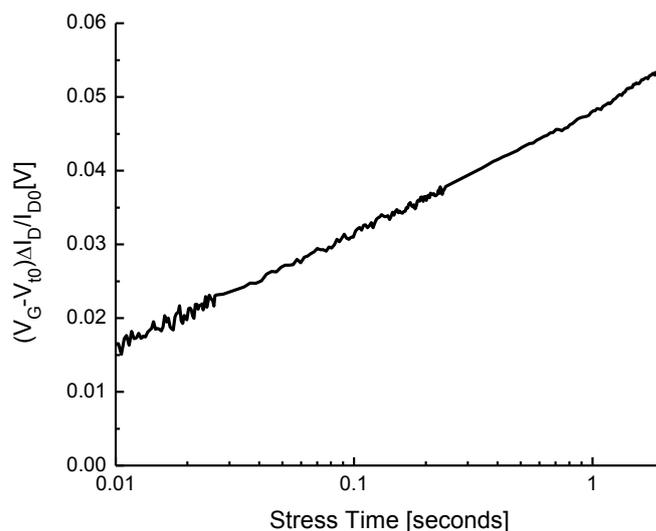


Figura 2.1 Degradação de V_{th} com o tempo, mostrando claramente um comportamento $\log(t)$ (WIRTH, 2011)

O efeito de NBTI foi primeiramente observado há mais de 40 anos atrás (MIURA, 1966), mas apenas começou a ganhar atenção nos últimos anos devido ao aumento de sua relevância à medida que a miniaturização da tecnologia CMOS avança. As características relacionadas à miniaturização, ou *scaling*, aumentaram a importância do efeito, tornando BTI uma das principais preocupações que afetam a confiabilidade os circuitos integrados. Tais características são:

- Presença de alto campo elétrico no óxido, devido ao dimensionamento do óxido ocorrer em proporções maiores que o dimensionamento das tensões aplicadas. (JEPPSON, 1977)
- Presença de altas temperaturas devido ao aumento de potência dissipada causada pela alta densidade de transistores.

- Uso de dielétricos formados por óxidos nitretados com alta constante dielétrica k , bem como novos materiais dielétricos conhecidos como *high-k* (DEGRAVE, 2008).

BTI é um efeito que depende da temperatura e do campo elétrico no óxido do transistor. Esse efeito é agravado ao se aumentar a temperatura ou o campo elétrico, através da diminuição da espessura do óxido ou do aumento da tensão aplicada à porta do transistor. Essas dependências podem ser vistas na Figura 2.2a e Figura 2.2b apresentadas em (VATTIKONDA, 2006).

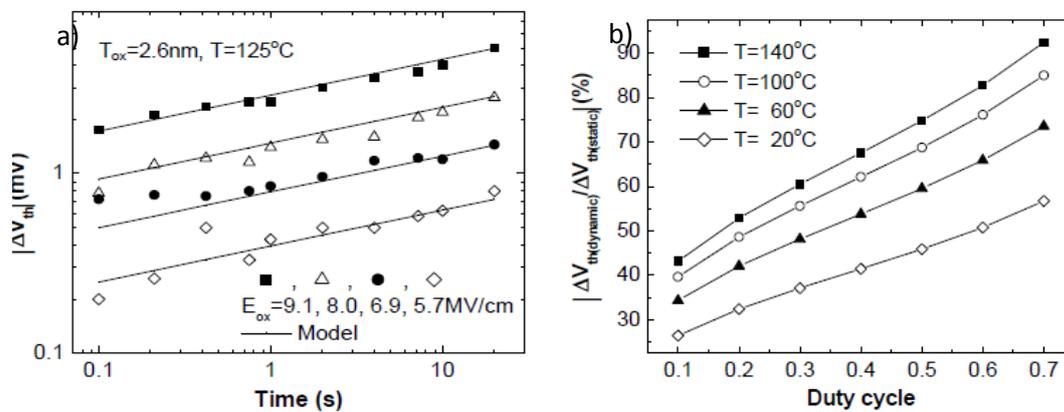


Figura 2.2 Dependências do efeito BTI com a) Campo elétrico, b) temperatura. (VATTIKONDA, 2006)

Além das já mencionadas dependências com a temperatura e tensão BTI, também é dependente do ciclo de trabalho, ou *duty factor*. Essa dependência se deve ao fenômeno conhecido como recuperação ou relaxação.

O fenômeno de recuperação ocorre quando a tensão de estresse é retirada ou até mesmo diminuída, fazendo com que a tensão limiar do transistor seja parcialmente recuperada. Esse tipo de estresse muitas vezes é chamado de BTI dinâmico para se contrapor ao efeito dado por uma tensão estática, chamado de BTI estático. A Figura 2.3, retirada de CHEN 2003, demonstra claramente esse efeito, onde após a retirada do estresse uma parcela do V_{th} é recuperada e quando a tensão do estresse é recolocada a degradação reinicia.

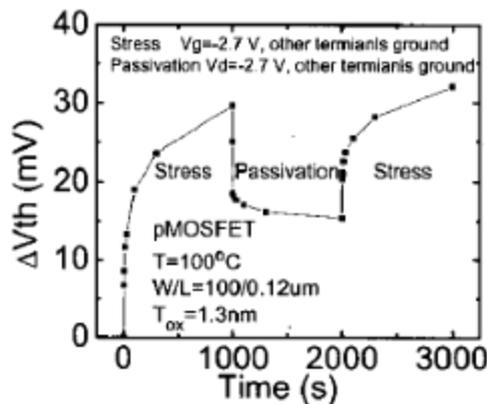


Figura 2.3 Ilustração do fenômeno de NBTI dinâmico, o período de relaxação recupera o V_{th} logo após a retirada do sinal de estresse. (CHEN, 2003)

Uma vez que BTI tem se demonstrado independente da frequência do sinal (CHEN, 2003) (como mostrado na Figura 2.4a), para um sinal AC, onde ocorrem sucessivos estresses e recuperações, o estresse total após certo tempo será dependente da relação entre o período que o sinal ficou em estresse e o período em que ficou sem estresse, dado pelo *duty-factor* $DF = (T_{estresse} - T_{relax}) / T_{estresse}$. Sendo assim, quanto maior o tempo de recuperação em relação ao tempo de estresse, ou seja, menor o *duty factor* do sinal, menor será o impacto do fenômeno de BTI no transistor, como mostrado na Figura 2.4b.

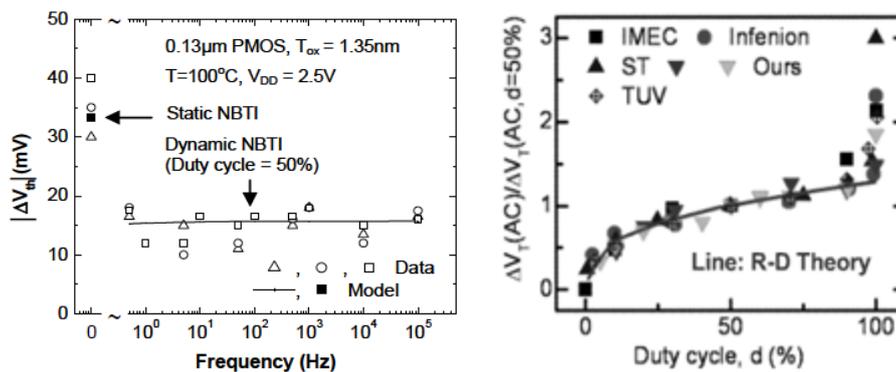


Figura 2.4 Impacto da a) frequência e b) do *duty factor* no efeito de BTI (CHEN, 2003).

2.2 Origens

As origens do efeito de BTI ainda são muito estudadas e discutidas (GRASSER, 2011). Atualmente, as duas teorias com maior presença na literatura são a teoria da criação e restauração de armadilhas na interface (ou *interface traps*) ao longo do tempo, chamada de modelo *Reaction-Difusion* e a teoria do aprisionamento (*trapping*) e liberação (*detrapping*) de portadores em armadilhas presentes no interior do óxido, chamada de modelo *trapping/detrapping*.

O modelo *Reaction-Difusion* foi um dos primeiros modelos a tentar modelar BTI e foi amplamente aceito na literatura por um longo período. Porém, atualmente o modelo *trapping/detrapping*, baseado em armadilhas no interior do óxido ou na interface, vem ganhando amplo espaço na literatura. Enquanto o modelo de *Reaction-Difusion* prediz uma lei de potência para a dependência temporal (t^n), o modelo de *trapping/detrapping* prediz uma dependência logarítmica (WIRTH, 2011). Os dois modelos serão discutidos brevemente a seguir.

2.2.1 Modelo Reaction-Difusion

O modelo *Reaction-Difusion*, ou modelo de Reação e Difusão, explica BTI como sendo efeito da criação de armadilhas na interface ao longo do tempo, dada pela quebra da ligação entre o hidrogênio, usado na passivação, e o silício. De acordo com o modelo essa quebra da ligação Si-H ocorre devido ao alto campo elétrico e às altas temperaturas presentes nos transistores nanométricos. Uma vez quebrada a ligação Si-H o átomo de hidrogênio migra em direção à porta do transistor, reagindo com outro átomo de

hidrogênio e formando H_2 , que por sua vez também se difunde em direção a porta, deixando assim, uma ligação incompleta de silício na interface.

Quando o campo elétrico é retirado, temos a recuperação do BTI. De acordo com o modelo, na falta de campo elétrico, o hidrogênio se difunde de volta à interface, se ligando novamente ao átomo de silício, assim recriando a ligação Si-H e passivando os defeitos da interface. A Figura 2.5, a seguir ilustra o modelo *Reaction-Difusion*.

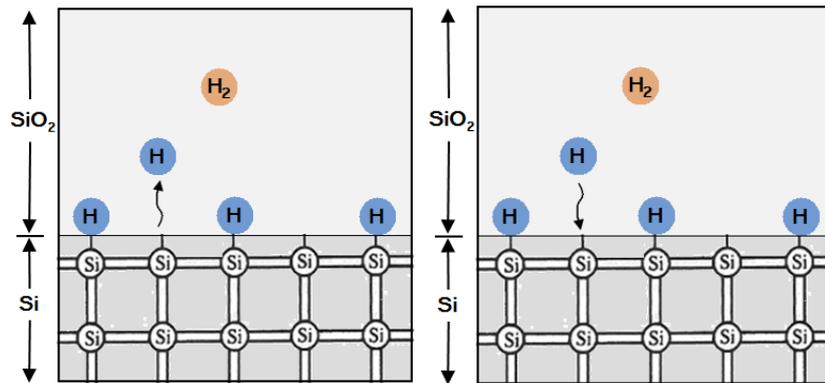


Figura 2.5 Modelo *Reaction-Difusion*

Porém, ao tentar descrever o fenômeno da relaxação, o modelo *Reaction-Difusion* acaba não sendo adequado para explicar a rápida recuperação que ocorre logo após a retirada do estresse. A diferença entre a teoria e a prática, apresentado por (REISINGER, 06) na Figura 2.6, se deve ao longo tempo para ocorrer a difusão do hidrogênio e a recriação das ligações Si-H descritos pela teoria.

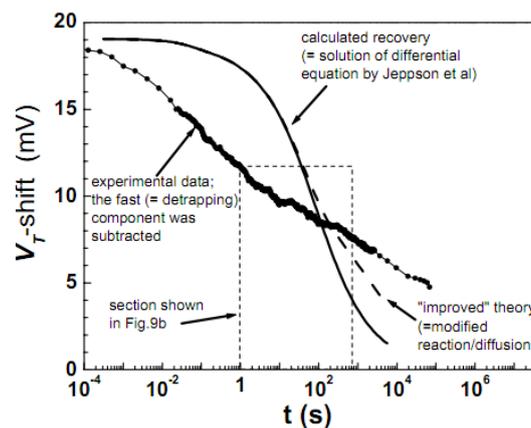


Figura 2.6 Diferença entre a teoria *Reaction-Difusion* e as medidas práticas. (REISINGER, 2006)

2.2.2 Modelo Trapping/Detrapping

O modelo *Trapping/Detrapping* foi inicialmente desenvolvido para explicar a rápida recuperação do efeito de BTI. Assim surgiu um modelo misto, onde dois modelos coexistem: criação de defeitos na interface, dado pelo modelo *Reaction-Difusion*, e a captura e liberação de cargas por armadilhas pré-existentes no interior do óxido do

transistor (HUARD, 2007). Neste modelo, a criação de defeitos na interface é o responsável pela parte permanente, não recuperável, da degradação enquanto as armadilhas, pré-existentes no interior do óxido, seriam responsáveis pela parte recuperável de BTI.

Modelos mais recentes, (GRASSER, 2009) (KACZER, 2009), propõem que BTI pode ser explicado unicamente pelo efeito de captura e liberação de portadores no interior do dielétrico.

No modelo *Trapping/Detrapping*, é explicada a variação de V_{th} pela pré-existência de armadilhas no interior do óxido, na qual cada armadilha tem como propriedades as probabilidades de captura e emissão, assim como, o desvio que causará no V_{th} . As probabilidades de captura e emissão são dadas pelo de tempo de captura e pelo tempo de emissão, que são os tempos médios transcorridos para que a armadilha capture um portador e emita esse portador, respectivamente. Esses tempos são log-uniformemente distribuídos, ou seja, armadilhas com diversas ordens de grandezas diferentes podem ser encontradas com a mesma probabilidade no óxido do transistor (GRASSER, 2010).

De acordo com o modelo, quando uma armadilha é ocupada por um portador de carga (elétron ou lacuna) ela ficará eletricamente carregada. Assim, a tensão de limiar do transistor é alterada, de acordo com a localização da armadilha nas três dimensões do óxido (espessura, largura e comprimento). A localização na espessura do óxido determina o efeito eletrostático que a armadilha terá na porta e no canal de inversão do transistor, já a posição da armadilha na largura e no comprimento do óxido poderá determinar quão impactante será seu efeito na mobilidade dos portadores no transistor, já que o seu efeito eletrostático pode obstruir um caminho de percolação, ou *percolation-path*, no canal (KACZER, 2010) (Ver Seção 2.5), diminuindo drasticamente a sua condutividade. A localização da armadilha no comprimento do canal também determina o impacto da armadilha na tensão superficial ao longo da fonte e do dreno. Combinando todos os efeitos, dados pela localização da armadilha no canal, tem-se o desvio no V_{th} dado por uma determinada armadilha.

Durante o período de estresse, armadilhas com diferentes tempos de captura e diferentes desvios de V_{th} são povoadas, alterando assim, a tensão limiar ao longo do tempo, conforme ilustrado na Figura 2.7, a seguir.

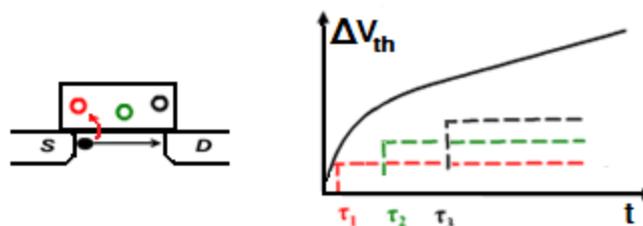


Figura 2.7 Ilustração da origem de BTI dada pela captura de cargas.

2.3 PMOS vs NMOS

Como mencionado anteriormente, o BTI é observável em ambos os transistores PMOS e NMOS, porém, o efeito PBTI sempre foi desconsiderado em tecnologias

baseadas em SiO_2 e $SiON$. Já o efeito NBTI em transistores PMOS foi considerado o principal responsável na degradação de circuitos. A Figura 2.8 mostra a diferença do impacto de NBTI e PBTI em ambos os transistores PMOS e NMOS.

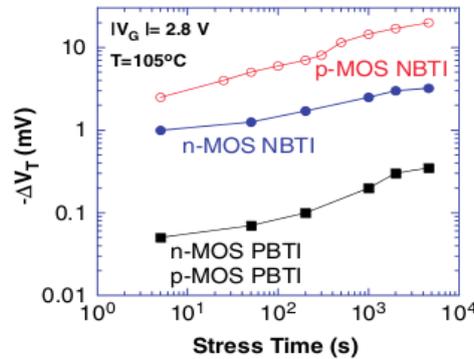


Figura 2.8 Diferenças entre NBTI e PBTI em ambos transistores PMOS e NMOS. (SCHRODER, 2007)

Entre as possíveis explicações para a diferença no impacto do BTI entre transistores do tipo PMOS e do tipo NMOS estão: difusão do hidrogênio do substrato e diferença na polarização do potencial de superfície. (SCHRODER, 2007)

A teoria da difusão do hidrogênio do substrato diz que a quebra das ligações do fósforo (usado como dopante nos transistores NMOS) com o hidrogênio, ocorrem mais facilmente que a quebra das ligações do boro, usado nos transistores PMOS. Como resultado da quebra da ligação do hidrogênio com os átomos dopantes do silício, o hidrogênio livre se difunde em direção à interface, assim, passivando as ligações incompletas do silício.

A teoria da diferença na polarização da superfície diz que os defeitos na interface causam uma polarização positiva na interface dos transistores PMOS enquanto causam uma polarização negativa na interface dos transistores NMOS. Porém, como as armadilhas existentes no interior do óxido são armadilhas para lacunas, ou seja, aprisionam cargas positivas, independente de ser PMOS ou NMOS, o efeito de BTI no NMOS é parcialmente anulado conforme:

$$\text{Canal tipo } n: \Delta Q_{\text{tot}} = \Delta Q_{\text{ox}} - \Delta Q_{\text{it}}$$

$$\text{Canal tipo } p: \Delta Q_{\text{tot}} = \Delta Q_{\text{ox}} + \Delta Q_{\text{it}}$$

Porém, ao contrário dos dielétricos clássicos formados por SiO_2 ou $SiON$ onde o PBTI era insignificante, com o advento do uso de dielétricos de alta permissividade baseados em camadas de SiO_2/HfO_2 se observou um grande aumento no efeito de PBTI, (ZAFAR, 2006) (DEGRAVE, 2008). Também, se constatou que em alguns casos, o PBTI poderia ter efeito ainda maior que NBTI, como observado por ZAFAR 2006 em dispositivos HfO_2 com portas formadas por $NiSi$.

2.4 Impacto de BTI nos circuitos digitais

O efeito de BTI impacta diretamente no desempenho dos circuitos digitais, uma vez que o efeito de envelhecimento aumenta V_{th} em módulo e ao aumentar V_{th} , diminui a

corrente de dreno (I_{ds}) do transistor, que por sua vez influencia o tempo de propagação do sinal elétrico em uma porta lógica.

Um dos desafios para se obter a degradação do atraso de um circuito, após algum certo tempo de estresse, está na dependência de BTI com o *duty factor*. O *duty factor* deve ser determinado para cada transistor do circuito, o qual depende da topologia deste, da probabilidade do transistor ser estressado e dos sinais de entrada do circuito. A Figura 2.9, abaixo, exemplifica o cálculo do *duty factor*, para os transistores PMOS de um circuito digital formado por portas lógicas NAND.

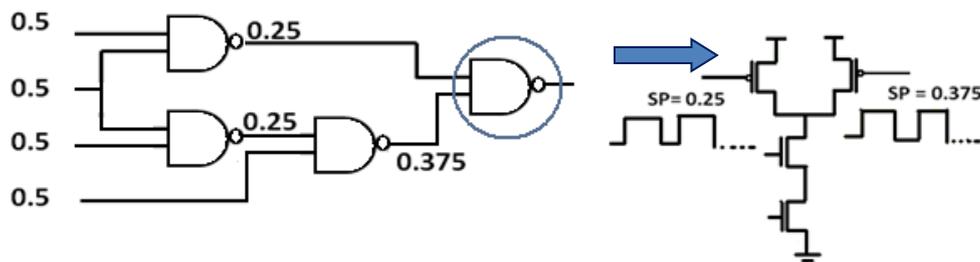


Figura 2.9 Exemplo do cálculo da probabilidade de estresse dos transistores de um circuito.

Outro fator que influencia a degradação de circuitos é o dimensionamento dos transistores utilizado nas portas lógicas, onde transistores com maior possibilidade de degradação devem ser corretamente dimensionados. O trabalho desenvolvido em SILVA 2009 mostra a importância de um correto dimensionamento dos transistores, a fim de minimizar o impacto do NBTI no atraso do circuito, no qual a melhor relação da largura do PMOS com a largura do NMOS deve ser achada levando-se em conta a degradação e a expectativa de uso do circuito. A Figura 2.10 mostra a melhor relação W_p/W_n de um inversor variando com o tempo de degradação e a Tabela 2.1 demonstra diferentes metodologias de dimensionamentos e seu impacto no atraso após 8 anos de degradação.

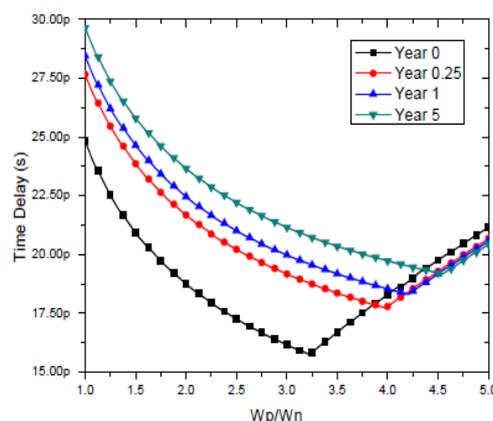


Figura 2.10 Impacto do dimensionamento na degradação do atraso máximo de um inversor, melhor relação W_p/W_n variando após degradação. (SILVA, 2009)

Tabela 2.1 diferentes dimensionamentos e seu impacto no atraso após 8 anos de degradação

		Slow	Fast	NBTI-aware
INV	Atraso máx.(ps)	22,6	20,21	19,04
	Wp(nm)	307	408	440
	Wn(nm)	96	128	96
NOR	Atraso máx.(ps)	29,92	21,42	26,57
	Wp(nm)	615	758	780
	Wn(nm)	96	118	96
NAND	Atraso máx.(ps)	30,28	28,46	25,61
	Wp(nm)	307	392	440
	Wn(nm)	192	245	192

Fonte: SILVA, 2009

2.5 Variabilidade

Nas recentes tecnologias que usam transistores nanométricos o aspecto estocástico de BTI ganha grande relevância, observando-se uma grande variabilidade na degradação dos transistores. Isto faz com que a previsão da degradação acabe se tornando complexa. Entre as causas para o aumento da variabilidade com a diminuição do tamanho dos transistores está: o alto efeito que uma única armadilha tem na degradação do V_{th} e o efeito de flutuações aleatórias na dopagem, ou *Random Dopant Fluctuation*.

O impacto de uma armadilha na tensão limiar pode ser estimado pela equação 2.1. E o alto efeito que uma única armadilha tem em V_{th} é explicado pelas diminutas dimensões, W_{eff} e L_{eff} .

$$\Delta V_{th} = \frac{q_t}{W_{eff} \cdot L_{eff} \cdot C_t} \quad (2.1)$$

Random Dopant Fluctuation é um efeito de variabilidade que afeta transistores diminutos. A variabilidade é dada pela pequena quantidade de dopantes necessária para dopar esses transistores e pela localização aleatória dos dopantes no canal do transistor. É sabido que a região onde se encontra o átomo dopante é uma região de maior resistência para a passagem do portador, funcionando como barreiras onde a energia de condução é mais elevada. Assim, em transistores nanométricos, onde não há presença de dopantes formam-se caminhos definidos pelo qual ocorre a condução da corrente, chamados de caminho de percolação, ou *percolation-path* (ASHRAF, 2011). Quando uma armadilha posicionada logo acima de um *percolation-path* é ocupada, a força eletrostática do portador preso acaba por afastar os outros portadores, obstruindo a passagem de corrente e assim diminuindo drasticamente a corrente do transistor (KACZER, 2010).

A obstrução completa ou parcial de um *percolation-path* é um grande responsável pelo alto desvio e variabilidade na tensão limiar que ocorre devido ao BTI. A figura a seguir ilustra bem esse conceito demonstrando as “ilhas” criadas pelos átomos dopantes e o caminho de percolação dos portadores.

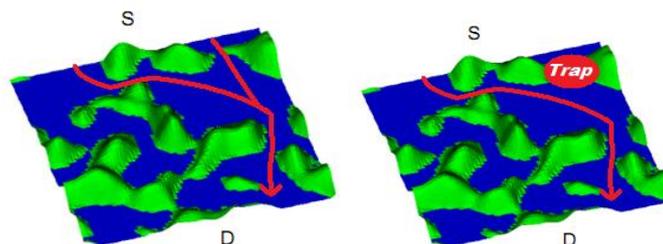


Figura 2.11 a) *Percolation Path* criado pelos dopantes do canal e b) obstrução do caminho por uma armadilha ocupada. Figura adaptada de Kaczer, 2010.

Ao diminuir gradualmente o tamanho do transistor a quantidade de defeitos contida em cada transistor é diminuída proporcionalmente. Dada a natureza estocástica dos defeitos no óxido, devido a sua localização associada ao fenômeno do *percolation-path*, BTI tem alta variabilidade em transistores diminutos. Assim, mesmo com transistores com mesmas dimensões, parâmetros e até mesmas quantidades de armadilhas, o efeito de BTI pode ser completamente diferente entre eles. A Figura 2.12 ilustra o aumento de variabilidade devido a miniaturização dos transistores e a diminuição da quantidade de armadilhas.

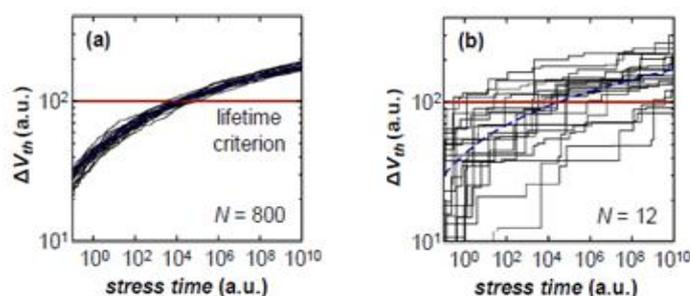


Figura 2.12 Variabilidade dada pela miniaturização dos transistores a) Desvio de V_{th} em transistores com 800 armadilhas. b) Desvio de V_{th} em transistores com 12 armadilhas. (KACZER, 2011)

3 CARACTERIZAÇÃO ELÉTRICA DE BTI

3.1 Extrairindo V_{th}

A técnica tradicional para a medida da tensão limiar do transistor (V_{th}) consiste em extrair a curva $I_d \times V_g$ na região linear do transistor, para assim se extrair seu V_{th} utilizando um dos muitos métodos disponíveis. Entre os métodos existentes estão: corrente-constante, extrapolação da região linear, extrapolação da transcondutância e método da segunda derivada.

A seguir esses métodos serão discutidos de forma sucinta, com base no trabalho realizado por (Ortiz-Conde, 2002). O objetivo não é uma discussão completa e exaustiva, mas exemplificar como a escolha do método pode impactar o resultado e a complexidade do trabalho de extração de V_{th} . Estas técnicas também serão utilizadas para avaliar a precisão obtida pelo método utilizado neste trabalho. Será medido o V_{th} de um transistor PMOS com as dimensões de comprimento do canal igual a 30 nm e largura igual a 150 nm utilizando o simulador SPICE com a tecnologia PTM (*Predictive Technology Model*) (ZHAO, 2006).

3.1.1 Método da corrente constante

Esse método avalia a tensão de limiar como o valor da tensão na porta do transistor V_g presente em uma dada corrente de dreno constante, I_d , para tensões V_{ds} pequenas, $V_{ds} < 100\text{mV}$. Um valor típico para a corrente de dreno é utilizar a relação $I_{ds} = W/L \times 10^{-7}$. Este método é bastante utilizado na indústria pela simplicidade e rapidez com que pode ser determinada a tensão de limiar, podendo ser determinada com apenas uma medida de tensão na porta do transistor para uma dada corrente I_d aplicada. Apesar da simplicidade, esse método pode apresentar pouca precisão por depender da escolha correta do valor para a corrente de dreno.

A Figura 3.1 mostra o exemplo do uso do método da corrente constante, sendo I_{ds} calculada por $I_{ds} = (150\text{nm}/30\text{nm}) \times 10^{-7} = 5 \times 10^{-7}$ A, assim, tendo como resultado $|V_{th}| = 0.242$ V.

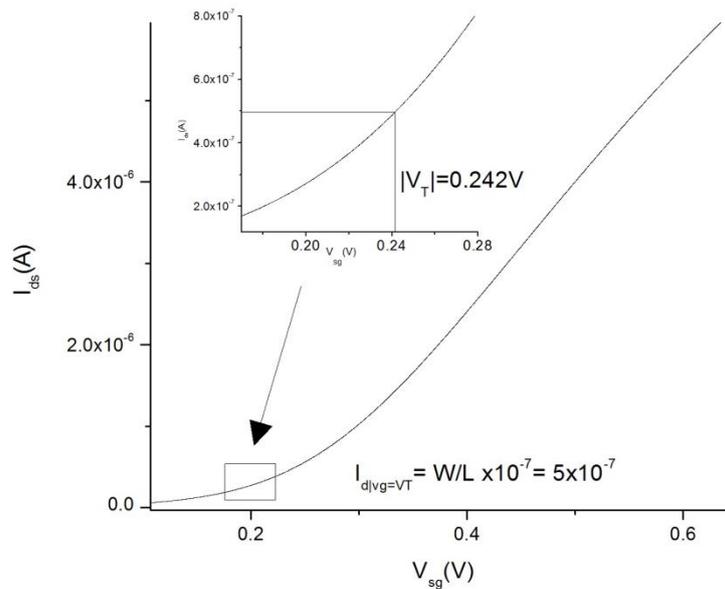


Figura 3.1 Método da Corrente Constante

3.1.2 Extrapolação da região linear

Conhecido como ELR (*Extrapolation in Linear Region*) é provavelmente o método mais conhecido de extração de V_{th} . O método consiste em achar o ponto de interceptação da extrapolação linear da curva $I_d \times V_g$ em sua máxima derivada no eixo das tensões de porta ($I_d = 0$). V_{th} , então, é achado adicionando-se $V_d/2$ no valor resultante da interceptação no eixo de V_g . Porém, esse método pode ser impreciso devido à incerteza de se achar a máxima derivada da curva $I_d \times V_g$, devido à degradação da mobilidade e a existência de grande resistência parasita entre a fonte e o dreno, que acabam desviando a curva $I_d \times V_g$ do comportamento linear, ideal, mesmo em tensões levemente acima de V_{th} .

Na Figura 3.2 é demonstrada a utilização do método ELR onde o valor da máxima derivada encontrada é 1.6×10^{-5} quando $V_g = 0.445$ V, logo o $|V_t|$ encontrado é igual a 0.25 V no transistor simulado.

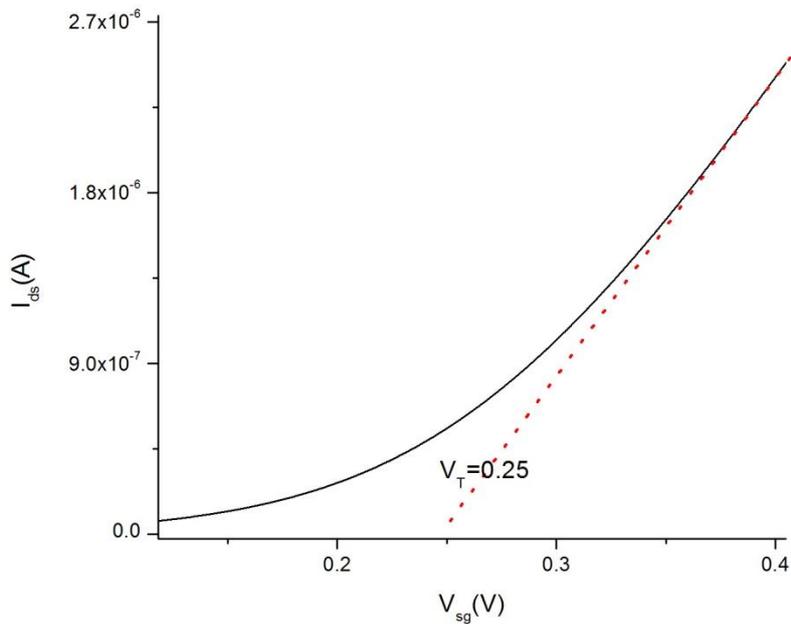


Figura 3.2 Método da extrapolação linear

3.1.3 Segunda derivada

O método SD (*Second Derivative*) foi desenvolvido para evitar a dependência da medida de V_{th} com as resistências em série com a fonte e o dreno do transistor a ser testado. Nesse método, V_{th} é definido como a tensão na porta na qual a derivada da transcondutância ($\partial g_m / \partial V_g = \partial^2 I_d / \partial^2 V_g$) é máxima. A ideia desse método vem da idealização da região linear, onde I_d tende a ser igual a zero para $V_g < V_{th}$ e I_d é proporcional a V_g para $V_g > V_{th}$. Assim, $\partial I_d / \partial V_g$ tende a ser uma função degrau e $\partial^2 I_d / \partial^2 V_g$ tende a ser infinita (na prática tende a ter o valor máximo) quando $V_g = V_{th}$. Como mostrado na Figura 3.3 esse método é muito susceptível ao ruído, nota-se que antes de o ruído do sinal ser suavizado, o máximo era difícil ser encontrado, já após a utilização de um filtro passa baixa, o V_{th} achado foi igual a 0.265 V.

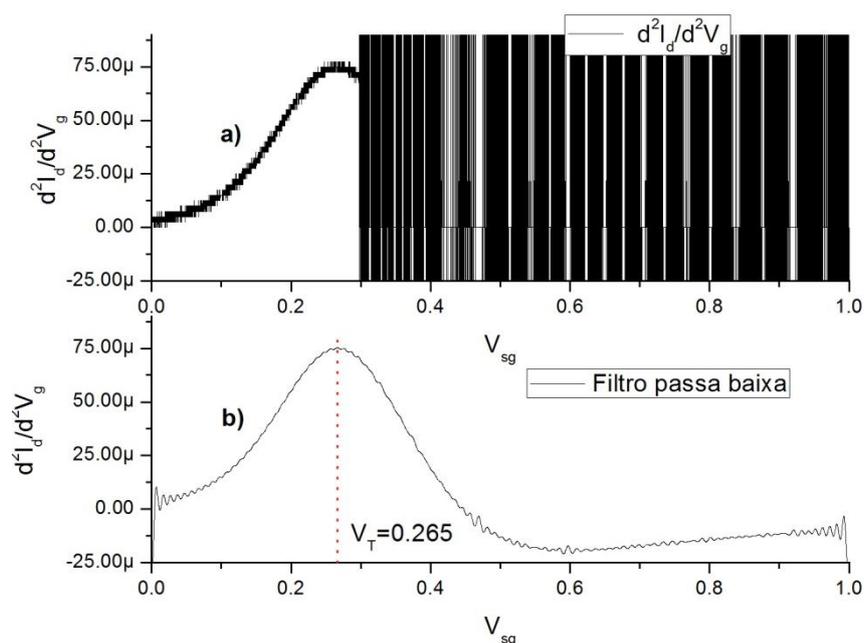


Figura 3.3 Método da Segunda Derivada, a)Alto ruído na medida da segunda derivada b)Após passar por um filtro passa baixa V_t pode ser facilmente determinado.

3.2 Extrairindo ΔV_t devido ao BTI

Como discutido anteriormente, uma interrupção na tensão de estresse resulta em uma parcial recuperação da degradação. Uma vez que a remoção temporária do estresse é necessária a fim de determinar V_{th} , e como ΔV_{th} diminui abruptamente após pequenos períodos de relaxação, a recuperação da tensão limiar que ocorre nesse período de medida causa erros na caracterização do efeito de BTI.

Os métodos que necessitam a extração da curva $I_d \times V_g$, mencionados anteriormente, requerem medidas de vários valores de I_d para diferentes valores de V_g , ou seja, é um método demorado para medir V_{th} requerendo um longo período sem a tensão de estresse no dispositivo para que a curva $I_d \times V_g$ seja feita. Por essa razão, esses métodos se tornam imprecisos ao tentar se extrair o desvio de V_{th} . Logo, o uso de outros métodos mais rápidos são necessário.

Assim, foram introduzidos alguns métodos que servem para caracterizar especificamente a degradação de BTI, utilizando abordagens onde o período de estresse é retirado em um intervalo rápido, como o método da rápida medida de I_d (Secção 3.2.1) e o método da rápida medida de V_g (Secção 3.2.2), ou utilizando a abordagem onde a caracterização é feita sem a retirada do estresse, como no método *on-the-fly* (Secção 3.2.3).

A caracterização de BTI também é difícil por requerer grandes tempos de estresse, sendo necessário que haja a caracterização de um mesmo dispositivo por horas ou dias, para modelar e prever BTI precisamente. Dado a característica logarítmica da degradação de V_{th} , as medidas podem ser feitas em intervalos de tempo também em escala logarítmica, a fim de diminuir a quantidade de medidas necessária e de diminuir ainda

mais o impacto da medida. As medidas de I_d na caracterização de BTI ocorrem como mostrado na Figura 3.4.

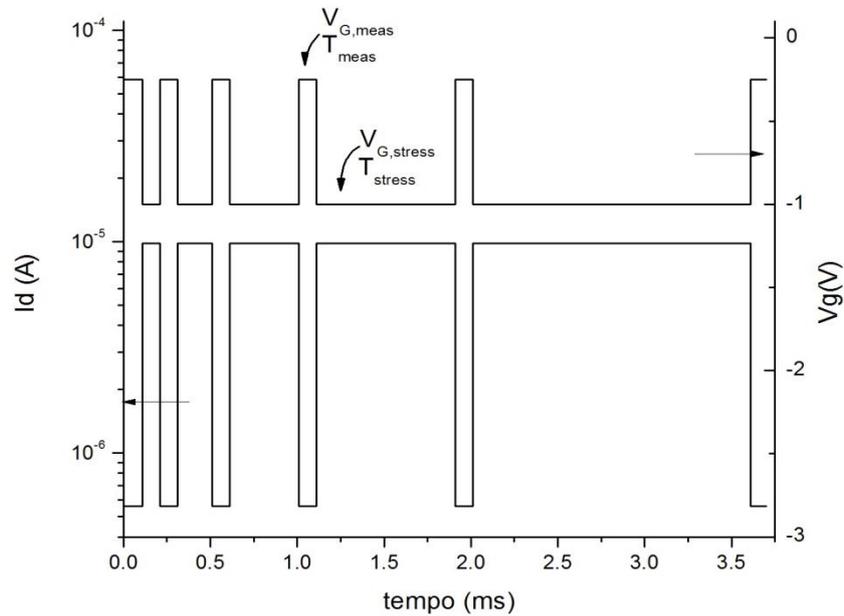


Figura 3.4 O transistor é polarizado como mostrado enquanto a corrente I_d é monitorada.

3.2.1 Rápida medida de I_d

Um método rápido para extração do desvio da tensão limiar foi apresentado no trabalho (KACZER, 2005). Nesse método faz-se o uso da curva $I_d \times V_g$ extraída anteriormente ao período de estresse. Uma vez começado o estresse, a medida é feita ao mudar-se a tensão, na porta do transistor, de uma tensão de estresse para a tensão de medida ($V_{estresse}$ para V_{medida}). Mantendo V_{medida} por um período curto o suficiente para a medição de apenas um ponto de I_d . Para essa medida se utiliza $V_g = V_{medida} \approx V_{th}$. Uma vez medido o novo valor da corrente de dreno, o desvio na tensão limiar é determinado deslocando-se a curva $I_d \times V_{th}$ conforme a Figura 3.5, a seguir.

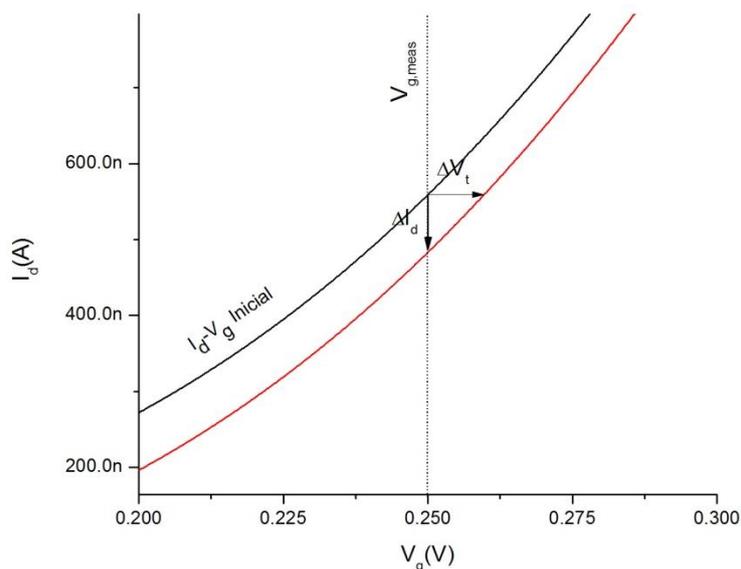


Figura 3.5 I_d medido é traduzido em ΔV_{th} deslocando horizontalmente a curva $I_d \times V_g$ inicial.

Como pode ser visto na Figura 3.5, uma mudança em de 1 mV em V_{th} significa uma mudança de quase 10% em I_d , assim uma precisão de 0.1% na medida de I_d é suficiente e pode ser realizada em um período de 1 milissegundo, utilizando um analisador de parâmetros padrão (REISINGER, 2007).

3.2.2 Rápida medida de V_g

O método da rápida medida de V_g foi introduzido por Reisinger, 2006 e permitiu uma rápida medida da degradação causada por BTI, após apenas alguns microssegundos da retirada do estresse. Dados colhidos com esse método foram essenciais para se colocar em dúvida a validade da teoria de *reaction/difusion*, como foi mostrado na Figura 2.6. Ao contrário do método da rápida medida de I_d , o método da rápida medida de V_g utilizada uma corrente constante, perto da encontrada quando $V_g = V_{th}$, e então, a tensão na porta do transistor é monitorada.

Para o método ser utilizado é necessário o uso de um circuito especial, mostrado na Figura 3.6. O circuito mostrado garante, enquanto o transistor é medido, uma tensão V_{ds} constante e uma corrente I_d também constante, através do uso de um amplificador operacional com um laço de realimentação que modifica V_g quando há degradação na tensão limiar do transistor. Assim, o método considera que o ΔV_g medido equivale ao ΔV_{th} proveniente da degradação por BTI.

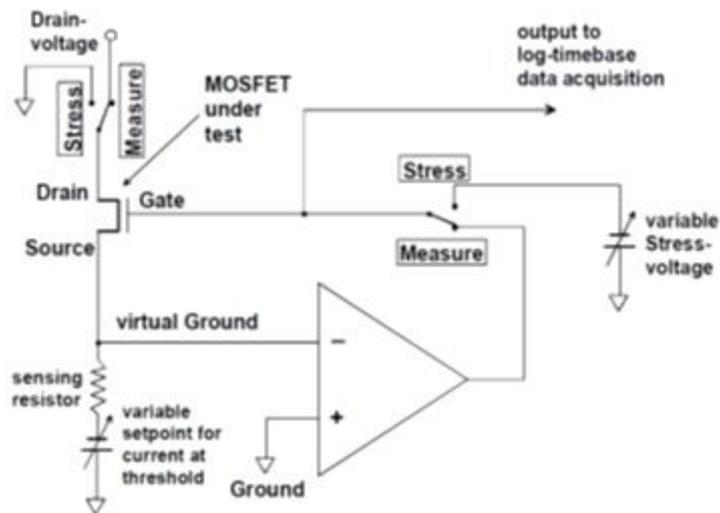


Figura 3.6: Circuito utilizado no método da rápida medida de V_g . (Reisinger, 2006)

Esse método de medida é mais rápido que o método anterior pois a medida de ΔV_{th} não depende da medida precisa de I_d . Permitindo que V_g seja medido com o uso de apenas um osciloscópio em um período de alguns microsegundos, apenas dependendo das capacitâncias do circuito e do tempo de acomodação do laço de realimentação.

3.2.3 Método On the Fly

O método On-the-Fly ou OTF foi primeiramente apresentado em (DENAIS, 2004), diferentemente dos dois métodos anteriores, esse método permite a medida de V_{th} sem a retirada da tensão de estresse, ou seja, sem que ocorra recuperação do BTI.

Como explicado em Reisinger, 2007, para isso ser possível, um método deve ser achado a fim de determinar V_{th} a partir de parâmetros medidos com tensões próximas às condições de estresse. O maior problema que o método OTF deve solucionar é que a degradação da mobilidade tem praticamente o mesmo efeito que o desvio de V_{th} : Uma diminuição na mobilidade inclina a curva I-V (Figura 3.7); uma degradação de V_{th} desloca a curva para a direita. Ambos efeitos são praticamente indistinguíveis no regime próximo à tensão de estresse. A situação piora pelo fato que a curva I-V se achata para altas tensões de V_g , assim, I_d se torna pouco sensível a variações de V_{th} . Logo, o método OTF deve conseguir separar os efeitos da mobilidade e do V_{th} .

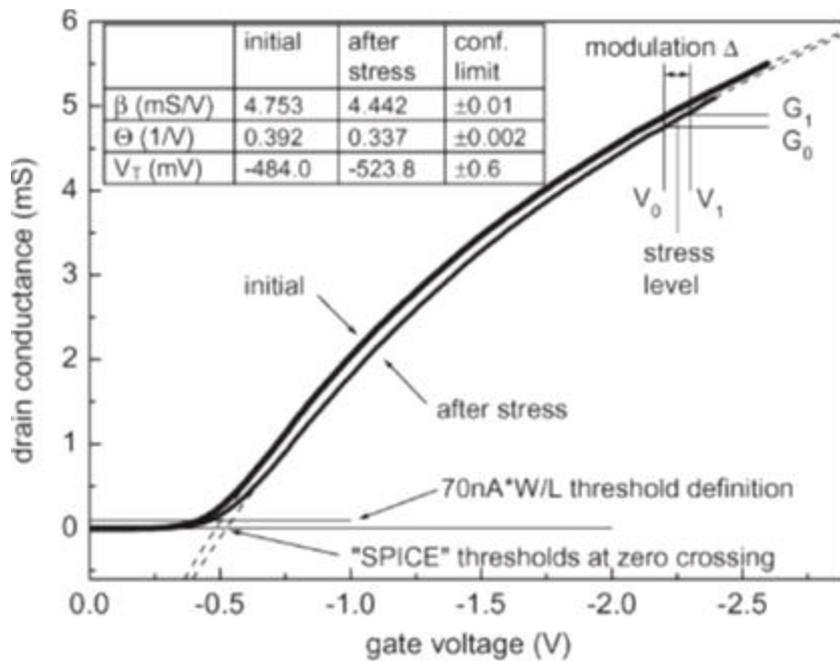


Figura 3.7: Curvas I-V antes e depois do estresse, linhas sólidas representam medidas e linhas pontilhadas representam a equação (2.2). (REISINGER, 2007)

O método OTF funciona com uma equação de três parâmetros de uma simulação SPICE para o regime linear de V_d

$$G_{SD} = \beta \times \frac{V_g - V_{th}}{1 + \Theta \times (V_g - V_{th})} \quad \text{para } V_g > V_{th} \quad (2.2)$$

onde G_{SD} é a condutância entre dreno e fonte, β age como a mobilidade independente de V_g , e Θ leva para uma aproximação assintótica da condutância onde o valor máximo $G_{SD} = \beta/\Theta$ descreve a diminuição da mobilidade em altos campos elétricos devida a rugosidade da superfície. β controla a mobilidade perto do máximo enquanto Θ faz a curvatura da curva I-V. A Equação (2.2) não tem significado físico. Ela é apenas um meio para aproximar a curva I-V real o melhor possível, usando apenas três parâmetros. Essa equação não descreve o regime perto de V_{th} , onde há um pico na mobilidade. Entretanto, como pode ser visto na Figura 3.7, a Equação (2.2) aproxima muito bem o resultado da simulação elétrica para correntes de dreno acima do pico na mobilidade. Assim, esta equação pode ser considerada apropriada para a extração dos parâmetros.

No método OTF V_{th} é encontrado determinando dois pontos na curva G_{SD} , como visto na Figura 3.7. Esses dois valores de G_{sd} (G_0 e G_1) e seus correspondentes valores da tensão na porta (V_0 e V_1) são inseridos na Equação (2.2). Assim, V_{th} é obtido de acordo com

$$V_{th} = V_0 - \frac{-(1 + \Theta(V_1 - V_0)) + \sqrt{(1 + \Theta(V_1 - V_0))^2 + 4\Theta G_0/g}}{2\Theta} \quad (2.3)$$

onde Θ é sabido de uma prévia caracterização do transistor e $g = \frac{G_1 - G_0}{V_1 - V_0}$.

Apesar do método OTF não necessitar a retirada do estresse, esse método é susceptível a muitos erros dados pelas aproximações feitas nas equações (2.2) e (2.3), pelo erro das medidas para se determinar g e por se considerar Θ constante, onde uma pequena mudança em Θ tem um grande impacto em V_{th} (Reisinger, 2007).

3.2.4 Extrapolação da Recuperação

Além do método *on-the-fly*, os métodos da rápida medida de I_d e V_g , apesar de serem realizados com muito mais rapidez que os métodos tradicionais para se encontrar V_{th} (medida lenta da curva $I_d \times V_g$), ainda necessitam da retirada do estresse para ocorrer a medida. Assim, a recuperação do BTI ocorrerá durante a medida e o efeito de degradação pode muitas vezes ser subestimado.

Para que BTI não seja subestimado foi introduzido em Kaczer, 2008 um método onde mais de um ponto de I_d ou V_g é medido quando o dispositivo está em recuperação (Figura 3.8). Permitindo, assim, que a degradação do transistor seja extrapolada com precisão para nano segundos ou mesmo pico segundos depois que o estresse foi retirado.

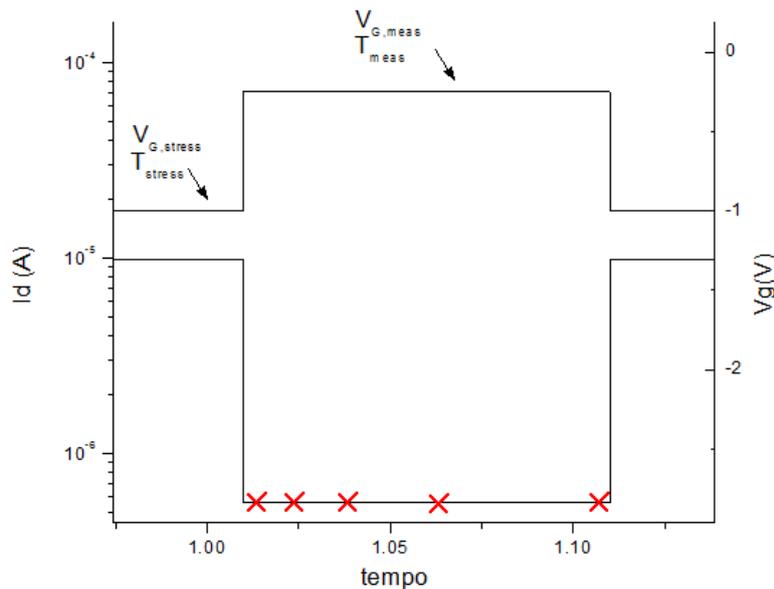


Figura 3.8 Exemplo das medidas realizadas durante o período de recuperação.

A Figura 3.9 demonstra a extrapolação feita nas medidas de BTI. Na figura os círculos sólidos são as medidas realizadas de acordo com o método da Figura 3.8, os símbolos X representam a extrapolação para $t_{relax}=0$, e os quadrados representam a extrapolação para t_{relax} tendendo a infinito.

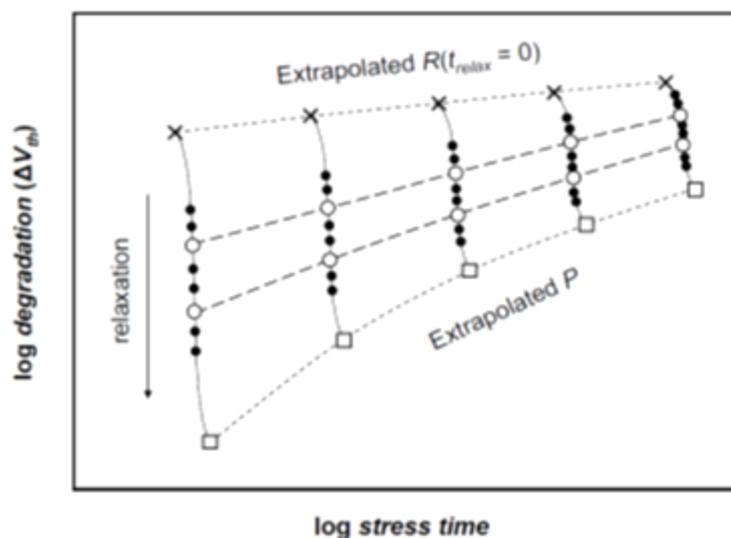


Figura 3.9 Extrapolação realizadas utilizando as medidas realizadas na recuperação. (KACZER, 2008)

3.3 Automatizando caracterização de BTI

Dada a quantidade de tempo em que é necessário um transistor ser submetido ao estresse e a necessidade de se caracterizar grandes quantidades de transistores, devido à variabilidade do efeito de BTI, foram desenvolvidas nos últimos anos, técnicas que utilizam circuitos *on-chip* com a finalidade de automatizar a caracterização do efeito.

Muitos trabalhos anteriores que tentam simplificar ou automatizar a medida do envelhecimento de transistores utilizam a degradação da frequência de oscilação de um oscilador em anel para prever a degradação por NBTI. Tais trabalhos podem ser visto em (REDDY, 2002) e (KETCHEN, 2007) onde a diferença entre a frequência antes e depois do estresse é medida através de um osciloscópio. O esquemático simplificado do circuito pode ser visto na Figura 3.10, o qual utiliza um anel de inversores, uma NAND - para alternar entre oscilação (medida) e o estresse - e um divisor de frequência para que se possa observar a frequência de oscilação no osciloscópio.

Essa é uma técnica que facilita a monitoração da degradação ao mesmo tempo em que ela ocorre. Entretanto, essa técnica tem baixa resolução devida às dificuldades associadas com: *i*) colocar sinais de alta frequência na saída de um circuito, *ii*) o grande tempo de medida requerido para uma medida precisa da frequência, *iii*) temperaturas e mudanças de tensões locais, e *iv*) o uso de divisores de frequência.

Esses trabalhos propostos também são incapazes de separar os efeitos NBTI, PBTI e *hot-carrier-injection* (HCI). Quando uma cadeia de um oscilador em anel é estaticamente estressada, o estresse se alterna entre o PMOS de um inversor e o NMOS do próximo inversor da cadeia. Quando sujeito a estresse dinâmico todos transistores PMOS e NMOS dos inversores presentes no oscilador em anel são estressados, as saídas dos inversores se alternarão entre GND e VDD e assim, o efeito HCI também ocorrerá.

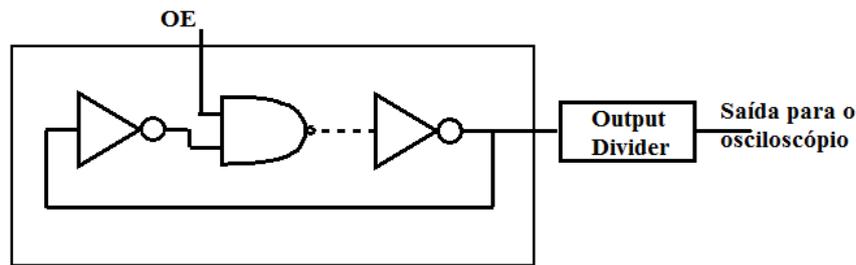


Figura 3.10 Esquemático do circuito apresentado em (REDDY, 2002)

O trabalho presente em (KIM, 2007) resolve o problema com a resolução utilizando um par de osciladores em anel. O circuito proposto utiliza um oscilador em anel como dispositivo em teste, o qual será estressado, e outro oscilador como referência, habilitado apenas na hora das medidas. O funcionamento do circuito se dá através do batimento da frequência dos dois osciladores utilizando um circuito de comparação de fase, como mostrado na Figura 3.11. A saída do circuito é dada pela contagem do número de oscilações (N) do oscilador de referência necessária para que o oscilador em teste e o de referência tenham suas fases re-alinhadas, como mostrado na Figura 3.12. Com essa abordagem, é possível conseguir medidas com precisões de pico-segundos. Como exemplo, para uma diferença anteriormente ao estresse entre a referência e o oscilador em teste de 1% e com a saída $N = 100$, após um certo tempo de estresse a degradação do oscilador fará, por exemplo, a diferença ser igual a 2% e o novo valor de N será igual a 50. Assim, neste exemplo se terá a saída variando de 50 contagens para apenas 1% de degradação. Logo, utilizando o monitor proposto com frequências de oscilação na ordem de Ghz se consegue precisões altíssimas, na ordem de pico-segundos. Além da alta precisão, esse método diferencial de medida anula efeitos de variações ambientais uma vez que ambos osciladores de referência e de teste estão sujeitos as mesmas temperaturas e tensões.

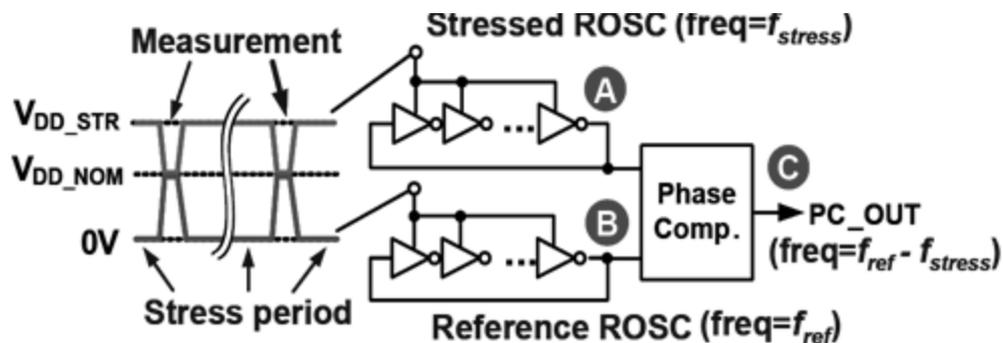


Figura 3.11 Esquemático do circuito de medida. (KIM, 2007)

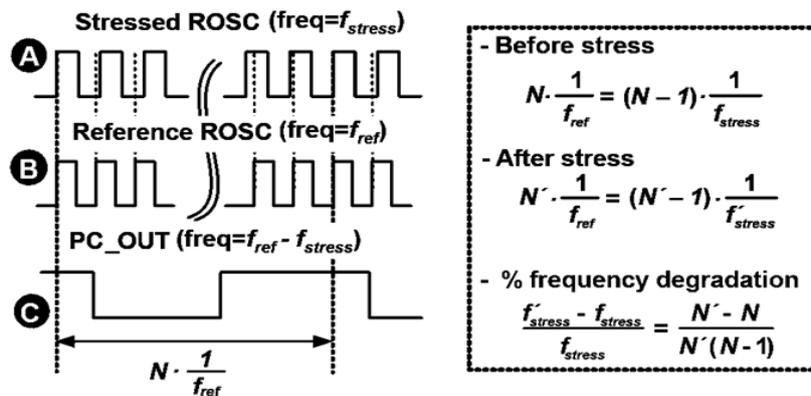


Figura 3.12 Degradação do oscilador dada em função do valor de saída N do circuito.

Trabalhos apresentados em (KIM, 2008) e (KEANE, 2010) apresentam circuitos de testes também baseados em osciladores em anel, o qual podem separar a análise de NBTI, PBTI e HCI. O trabalho proposto em (KEANE, 2010) separa os efeitos BTI e HCI através do desligamento dos drenos dos transistores da fonte de alimentação durante o estresse. O trabalho apresentado em (KIM, 2008) propõem uma nova célula para o oscilador em anel baseada na combinação NAND-NOR para selecionar entre os estresses e medidas de NBTI ou PBTI. A Figura 3.13 demonstra a célula proposta e a Tabela 3.1 descreve seu funcionamento.

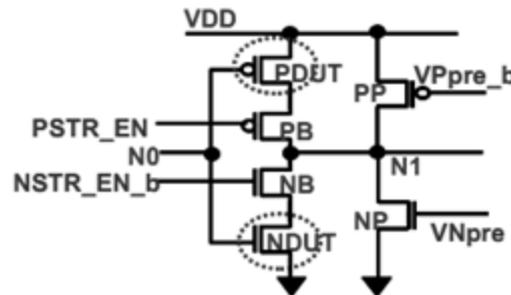


Figura 3.13 Célula baseado em uma combinação NAND-NOR. (KIM, 2008)

Tabela 3.1 Funcionamento da célula proposta

Sinal	Tensão utilizada em Modo de Estresse		Tensão utilizada em Modo de Medida
	NBTI	PBTI	
VDD	VDD_stress	VDD_stress	VDD
PSTR_EN	VDD_stress	VDD_stress	GND
NSTR_EN_b	GND	GND	GND
VPpre_b	VDD_stress	GND	VDD
VNpre	VDD	GND	GND

Fonte: KIM, 2008

Finalmente, o trabalho apresentado em (KEANE e ZHANG, 2010) aperfeiçoa o trabalho presente em (KIM, 2007) utilizando uma matriz de osciladores em anel para a caracterização estatística de NBTI.

Porém, esses trabalhos apresentados não são adequados para a caracterização de um transistor individual, e logo, não são adequados para entender os efeitos de BTI no

nível do dispositivo. Recentemente, o trabalho apresentado (SCHLUNDER, 2011) descreveu um circuito baseado numa matriz de transistores capaz de fazer medida de um único transistor. Entretanto, o método utilizado tem a desvantagem de aplicar diferentes tempos de estresse em cada dispositivo em teste. Essa desvantagem será tratada em detalhes mais adiante.

A proposta apresentada, nesse trabalho, propõe e descreve um circuito baseado em uma matriz de transistores para que se possa caracterizar *on-chip* o efeito de *Bias Temperature Instability* de forma automática, eficientemente e massivamente. O design é altamente eficiente quando estudado a componente estatística de BTI, onde centenas de dispositivos devem ser testados para que se possa obter uma amostra estatisticamente significativa. Como será visto o circuito controla os tempos de estresse e medida para uma caracterização precisa, fazendo com que todos dispositivos em teste tenham os mesmos tempos submetidos a períodos de estresse e tenham os mesmos tempos submetidos a períodos de medida.

4 CIRCUITO PROPOSTO

4.1 Motivação

Como já discutido anteriormente, os defeitos nos transistores são responsáveis pelos efeitos de degradação, como o BTI. Ao diminuir gradualmente o tamanho do transistor a quantidade de defeitos contida em cada transistor é diminuída proporcionalmente, assim, a contribuição de cada defeito se torna relevante para o efeito de envelhecimento. Dada a natureza estocástica dos defeitos no óxido ou na interface, BTI apresenta grande variabilidade em transistores diminutos.

Devido à alta variabilidade do BTI, o efeito deve ser caracterizado estatisticamente. Os atuais circuitos digitais contêm cerca de bilhões de transistores. Para se ter uma boa prevenção ou predição do efeito se necessita de uma boa caracterização estatística. Logo, deve se ter precisos valores de média, desvio padrão e informações sobre a forma da distribuição (como *skewness* e *kurtosis*), que para serem determinados com precisão necessitam da caracterização de milhares de dispositivos.

Para suprir as condições necessárias à boa caracterização do efeito de envelhecimento, esse trabalho propõe a utilização de um circuito *on-chip* que possa diminuir o tempo total de caracterização e ao mesmo tempo prover uma boa caracterização estatística através das seguintes características:

Reduzir Tempo de Avaliação:

- Automatizar o teste utilizando interfaces simples
- Avaliar um grande número de dispositivos similares para caracterização estatística.
- Avaliar um grande número de dispositivos diferentes
- Testar dispositivos paralelamente

Boa Caracterização Estatística:

- Medir variação de um único dispositivo
- Medir variação de parâmetros
- Boa resolução, baixo erro
- Circuito simples

4.2 Idéia Central

Com o objetivo de caracterizar rapidamente centenas de transistores, foi desenvolvido, nesse trabalho, o circuito que paraleliza a caracterização de BTI. Assim, todos os transistores podem ser estressados ao mesmo tempo, diminuindo drasticamente o tempo de medida em comparação a todos os transistores sendo testados individualmente.

Para a paralelização dos testes, o circuito proposto é baseado em uma matriz de $m \times n$ transistores, chamados aqui de *Devices Under Test* ou DUTs, no qual todos os transistores são estressados em paralelo e medidos sequencialmente.

O circuito foi projetado para funcionar com o método da rápida medida de I_d , de acordo com o explicado na secção 3.2.1. Primeiramente é traçada uma curva $I_{ds} \times V_{gs}$ previamente ao estresse, onde o V_{th} possa ser determinado. Após o estresse apenas um ponto de I_{ds} (para $V_{gs} \approx V_{th}$) é necessário para determinar ΔV_{th} , minimizando assim a recuperação que pode ocorrer durante a medida.

A matriz é formada por células, na qual, cada célula é formada pelo DUT, por um multiplexador e por duas chaves. Sendo a porta do DUT conectada no multiplexador, que por sua vez, é o responsável por selecionar entre tensões de estresse e tensões de medida. O dreno do DUT é conectado as duas chaves nas quais são conectadas as trilhas de *force* e *sense*. Já a fonte do DUT é conectada em VDD para o caso dos DUTs serem transistores PMOS ou em GND caso transistores do tipo NMOS. A Figura 4.1 demonstra a idéia básica da célula da matriz. Os dispositivos de teste PMOS e NMOS são utilizados em matrizes diferentes para a caracterização de NBTI e PBTI, respectivamente.

As trilhas de *force* e *sense* são utilizadas pela técnica *force-and-sense* (também chamada Kelvin) que permite medidas precisas compensando a queda de tensão, nas resistências das interconexões e nas chaves de pequenas dimensões, pelo uso de um laço de realimentação. Essa técnica consiste de duas interconexões usadas independentes, uma para estímulo (*force*) e a outra para a medida da tensão (*sense*). A trilha de *force* modifica sua tensão de acordo com a tensão medida na trilha de *sense*, e assim, alcançado a tensão desejada no dreno do DUT.

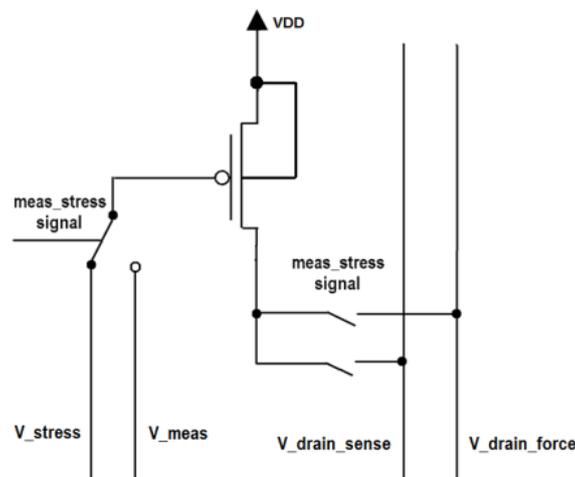


Figura 4.1 Idéia básica para a célula da matriz com DUT do tipo PMOS

Para permitir que apenas um DUT seja medido independentemente dos demais, existem as duas chaves, as quais permitem conectar um DUT às trilhas de medida (*force* e *sense*), enquanto os DUTs restantes permanecem desconectados das trilhas. As chaves são feitas utilizando dois *transmission gates* em série para que se diminua a corrente de fuga, ou *leakage*, o que também aumenta a confiabilidade do circuito. Caso contrário, um único *transmission gate* com curto circuito levaria toda a matriz a falha, uma vez que um DUT estaria sempre conectado as trilhas de medida.

As chaves também permitem que os DUTs sejam desconectados das trilhas de medida quando em estado de estresse. Assim, não há passagem de corrente pelos DUTs e evita-se a degradação pelo fenômeno de HCI, assegurando que apenas o efeito BTI seja caracterizado.

O multiplexador que possibilita a seleção entre as tensões de estresse e medida na porta do DUT é também feito com *transmission gates* permitindo o uso de uma variedade de tensões de *GND* até *VDD* para o estresse e a medida. *Transmission gates* também permitem o estresse utilizando sinais com diferentes frequências e *duty factors*, para que assim se possa realizar o estudo do impacto desses fatores na degradação BTI. As tensões de estresse (V_{stress}) e a tensão de media (V_{meas}) são providas externamente, assim como o *sense* e *force* do dreno do DUT.

O sinal *stress_meas* mostrado na Figura 4.1 é gerado internamente por um circuito de controle e será explicado adiante. A tensão *VDD* para PMOS e a tensão *GND* para os DUTs do tipo NMOS são compartilhadas com todos os transistores no circuito, assim, permitindo o comportamento adequado dos *transmission gates*.

4.2.1 Detalhamento das Chaves

As chaves foram feitas utilizando dois *transmission gates* em série com o objetivo de diminuir a corrente de fuga I_{off} , como mostrado na Figura 4.2. Entre os efeitos responsáveis para a corrente de *leakage* estão: inversão fraca, DIBL e o efeito de corpo, podendo ser modelado de acordo com a fórmula a seguir:

$$I_D = A \frac{W}{L} \times e^{\frac{1}{m\gamma T}(V_g - V_s - V_{th0} - \gamma V_s + \eta V_{ds})} \times (1 - e^{-\frac{V_{ds}}{V_T}}) \quad (2.4)$$

Pode-se ver que a corrente de fuga é exponencialmente dependente de V_d e V_s enquanto é inversamente dependente de L . Então para buscar a minimização de I_{off} ao invés de apenas utilizar maiores valores para o comprimento do canal L pode-se inserir mais um transistor em série, assim, os transistores funcionam como um divisor de tensão e V_{ds} será menor em cada transistor assim como o V_{gs} .

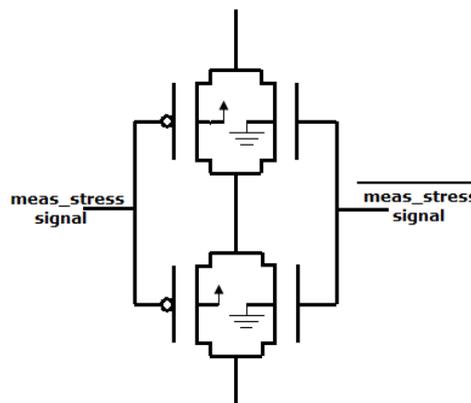


Figura 4.2 Esquemático da chave utilizada.

4.2.2 Detalhamento do multiplexador

O multiplexador também foi projetado com *transmission gates* e segue o esquemático de acordo com a Figura 4.3, abaixo:

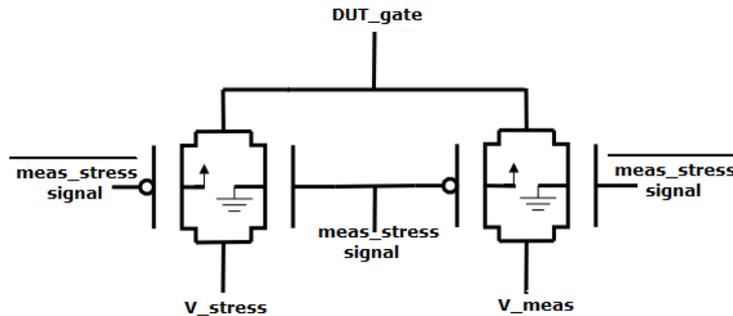


Figura 4.3 Esquemático do multiplexador utilizado.

4.2.3 Método *Force and Sense* (Kelvin)

O método *force and sense*, também conhecido como Kelvin ou medida de 4 pontas, serve para a medida precisa de pequenas resistências, na qual a resistência encontradas nos fios utilizados na medida afetam a medição. A configuração bastante usada na medida de 4 pontas é encontrada na Figura 4.4a e consiste de fios separados para a fonte de corrente e para a medida da tensão. A corrente é fornecida pela conexão *force*, gerando uma queda de tensão na impedância a ser medida de acordo com a lei de Ohm $V = RI$. Porém, essa corrente também irá gerar queda de tensão nas resistências parasitas da conexão. A fim de evitar a introdução desse erro na medida, a conexão *sense* é adicionada adjacente à impedância a ser medida. A eficácia do método vem do fato que quase nenhuma corrente passa através da conexão de medida (*sense*) e a queda de tensão nas resistências parasitas da conexão *sense* é desprezada e a impedância pode ser achada por $R = \frac{V_{sense}}{I_{force}}$.

Para o circuito proposto, utiliza-se as conexões *force* e *sense* do aparelho de medição disponível, que é representado pelo amplificador operacional da Figura 4.4b, nessa abordagem a entrada (-) do amplificador serve como a conexão *sense*, tendo corrente desprezível. Nessa configuração as conexões *force* e *sense* estão no circuito de realimentação do amplificador, assim a resistências parasitas são compensadas e a tensão na fonte do transistor é dada pela fonte presente no terminal (+).

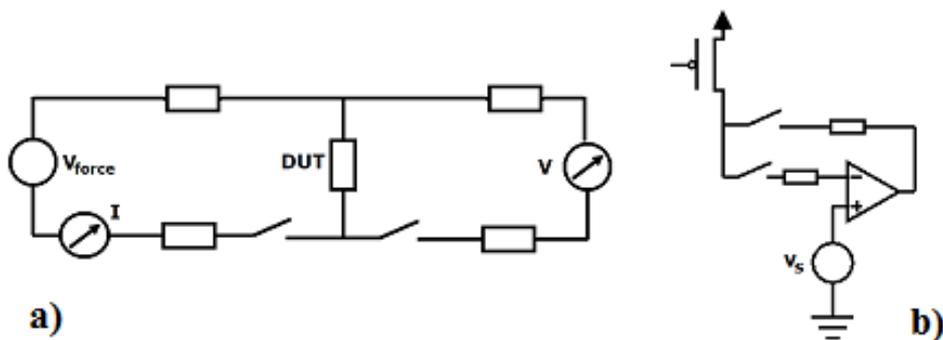


Figura 4.4 Demonstração do uso do método Force and Sense. a) Configuração básica para achar impedância b) Configuração utilizada no circuito proposto.

4.3 Controlando o estresse e a medida

Devido ao mecanismo de recuperação que ocorre durante a medida de BTI o principal problema associado com a paralelização da caracterização se encontra nas diferenças entre os tempos de estresse e entre os tempos de medida (no qual ocorre à recuperação ou relaxamento). Para uma caracterização estatística é obrigatório que todos os dispositivos testados tenham os mesmos tempos de estresse e os mesmos tempos de recuperação; caso contrário os dados estatísticos seriam impactados por essas diferenças.

Utilizando a matriz com simples células formadas por DUTs, como visto na Figura 4.1, existem duas possibilidades de medidas, apresentadas na Figura 4.5. Na Abordagem I o período de estresse para todos os transistores começa ao mesmo tempo e para, para a realização das medidas, também ao mesmo tempo, levando a diferentes períodos de recuperação durante a medida sequencial dos DUTs. Na Abordagem II o tempo de estresse para apenas no transistor a ser medido enquanto os outros DUTs continuam sobre estresse, levando a diferentes tempos de estresse para cada DUT.

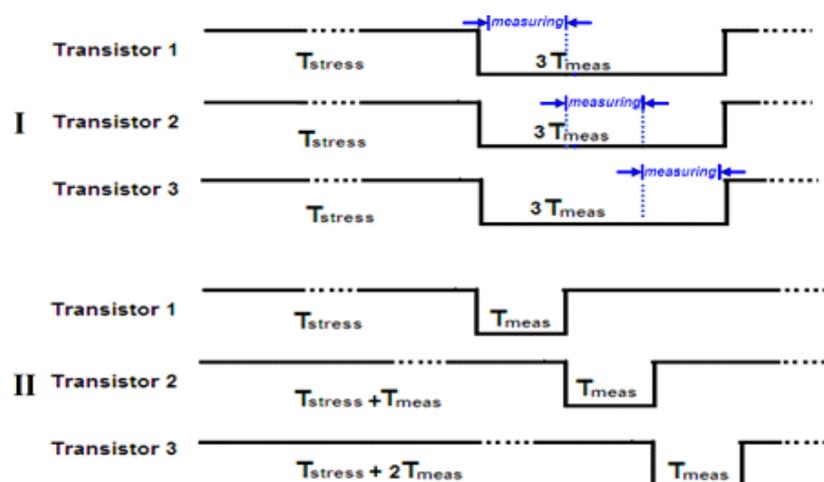


Figura 4.5 Dois comportamentos diferentes para estresse e medida

Escolher a segunda abordagem parece ser a melhor opção uma vez que a recuperação ocorre rapidamente nos primeiros milissegundos enquanto a degradação, durante o estresse, ocorre mais lentamente a medida que o tempo de estresse aumenta (devido ao comportamento logarítmico em função do tempo). Então, a escolha por tempos diferentes de estresse aparenta ser mais precisa para medidas estatísticas. Porém, essa abordagem também começa a ser imprecisa com o aumento do tamanho da matriz, uma vez que a diferença no tempo de estresse entre primeiro e o último DUT a ser medido será maior, e também com o aumento do tempo de medida (i.e. para que se observe e estude o comportamento da recuperação de BTI). Essa abordagem também leva a grandes erros quando caracterizando BTI nos primeiros segundos de estresse. Como exemplo, 1 ms de período de medida em uma matriz de 1000 dispositivos irá levar em uma diferença de 1 s entre o primeiro dispositivo e o último dispositivo. Se o tempo de estresse a ser testado

é 1 s esse método adiciona um erro determinístico de 100% no estresse aplicado entre o último e o primeiro dispositivo.

Com o auxílio dos gráficos presentes na Figura 4.6, obtidos através de simulações utilizando o simulador em (Kaczer 2011), pode-se ver o erro introduzido utilizando-se a Abordagem I e II, Figura 4.6a e Figura 4.6b respectivamente. Os gráficos mostram o erro médio de medida (para 1000 transistores simulados) dado pela diferença nos tempos de medida, Figura 4.6a, e nos tempos de estresse, Figura 4.6b, após 100 milissegundos, 1 segundo e 100 segundos de estresse.

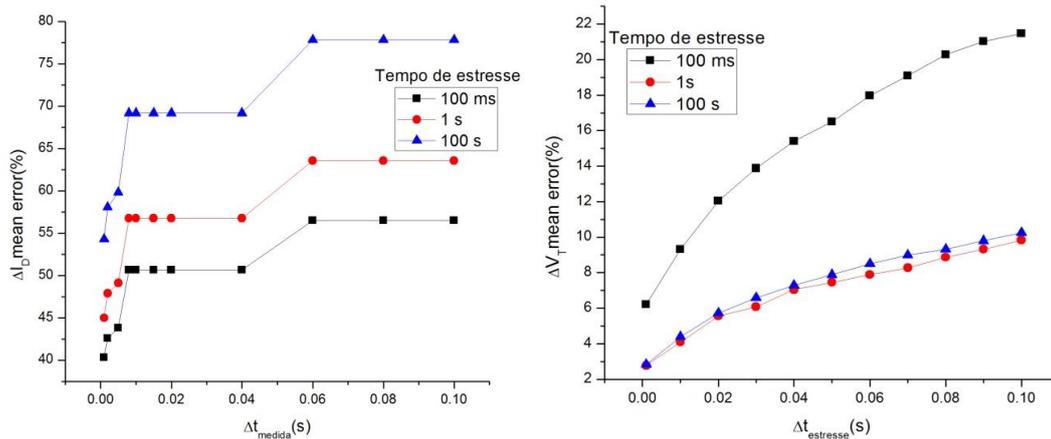


Figura 4.6 Erro médio de medida, para 1000 transistores simulados, para diferentes períodos de estresse, devida às diferenças a) nos tempos de medida b) nos tempos de estresse.

Analisando os resultados da Abordagem I, se percebe que o erro é maior quanto maior é o período de estresse que o transistor ficou submetido, pois após um certo tempo de estresse haverá mais armadilhas a serem recuperadas. Também se nota que nos primeiros milissegundos de diferença no tempo de medida, o erro já é elevado, devido à rápida recuperação que ocorre no efeito BTI.

Analisando os gráficos da Abordagem II, nota-se que o tempo de erro médio devido à diferença no tempo de estresse é maior para valores menores de estresse, isso se deve, como já mencionado, às diferenças no tempo de estresse estar na mesma ordem de grandeza do período de estresse. E ao aumentar o tempo de estresse diminui-se o erro devido ao comportamento logarítmico do efeito BTI.

Portanto, este trabalho propõe uma nova abordagem que torna ambos períodos de estresses e de medidas o mesmo para todos os DUTs. Esse comportamento é alcançado deslocando o início do estresse e da medida entre DUTs. O comportamento do sinal de controle é apresentado na Figura 4.7. O começo do estresse acontece sequencialmente em cada DUT com o atraso de um período de medida em relação ao dispositivo anterior, então o estresse é mantido paralelo em todos os transistores até que um DUT seja medido.



Figura 4.7 Novo comportamento proposto para o estresse e a medida.

O comportamento do circuito proposto pode ser dividido em 6 estados (Figura 4.8):

1) Estado de espera: nenhum DUT está sendo medido e nenhum DUT está sendo estressado. O circuito está esperando, se preparando, para a realização das primeiras medidas.

2) Medida de V_{th} : um DUT está sendo medido sem que os outros dispositivos sejam estressados. Medida começa seqüencialmente. V_{stress} deve ser igual a uma tensão de não-estresse, i.e. $V_{stress} = GND$ para NMOS e VDD para transistores PMOS.

3) Estado de espera: nenhum DUT está sendo medido e nenhum DUT está sendo estressado. Tensão na porta do DUT é V_{meas} enquanto V_{stress} se torna uma real tensão de estresse. Circuito está esperando, se preparando, para a realização do estresse. Assim $V_{stress} < VDD - V_{th}$ para PMOS e $V_{stress} > GND + V_{th}$ para NMOS.

- 4) Estresse começa seqüencialmente com um atraso igual ao tempo de medição.
- 5) Todos os DUTs estão sendo estressados paralelamente
- 6) Medida de ΔV_{th} começa seqüencialmente.

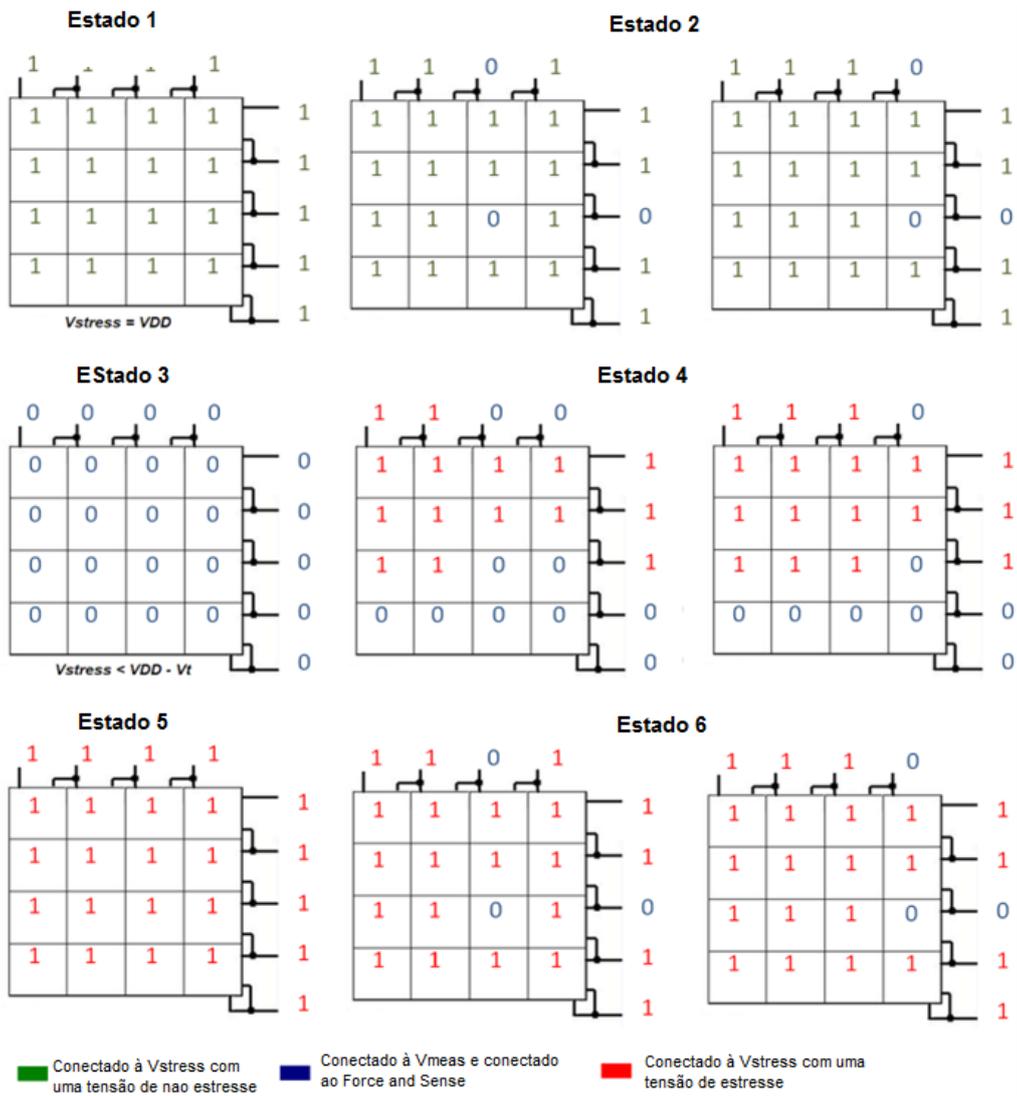


Figura 4.8 Comportamento do circuito e seus 6 estados

Através do comportamento desejado do circuito e da Figura 4.8 chega-se a ao mapa de Karnough, Figura 4.9, e o circuito responsável pelo sinal de controle dos tempos de estresse e medida pode ser deduzido.

$C_n C_{n+1}$ \ $R_m R_{m+1}$	00	01	11	10
00	0	0	0	0
01	0	0	1	1
11	1	1	1	1
10	0	1	1	1

Figura 4.9 Mapa de Karnough

$$R_m \cdot R_{m+1} + (R_m + R_{m+1})C_m + R_m C_{m+1} = \overline{\overline{R_m \cdot R_{m+1} \cdot (R_m + R_{m+1})C_m \cdot R_m C_{m+1}}} \quad (2.5)$$

4.4 Circuito Final

Para ser possível obter o comportamento acima descrito, foi acrescentado à célula da Figura 4.1 o circuito lógico que controla os tempos de medida e estresse, representado pela equação (2.5). A célula final, chamada de Célula BTI, pode ser vista na Figura 4.10.

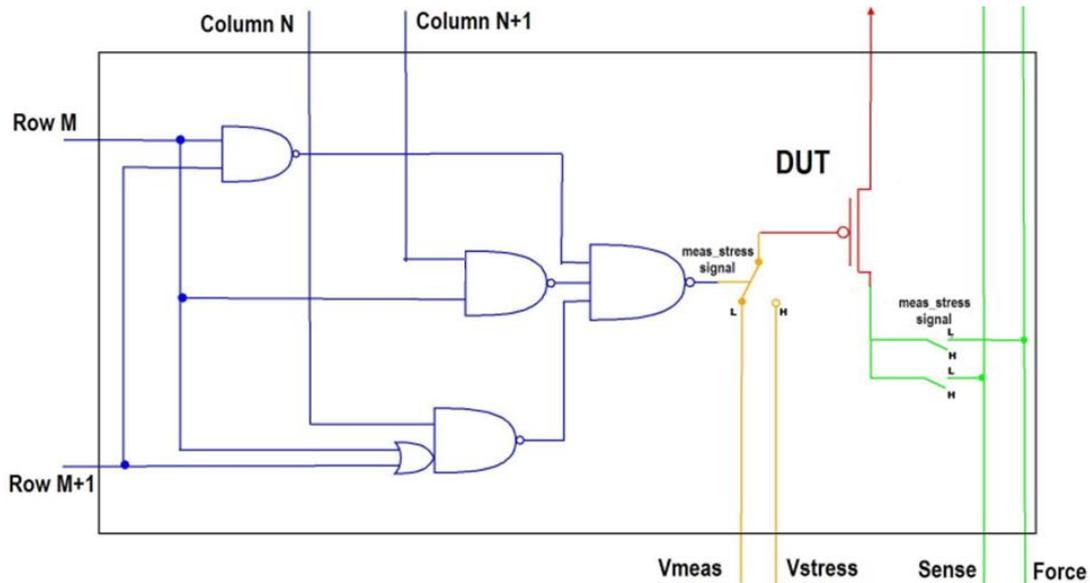


Figura 4.10 Célula BTI para um DUT do tipo PMOS

Para o endereçamento dos sinais de controle das linhas e colunas foi escolhido o uso de *shift registers* (mostrado na Figura 4.11), o que permite obter o comportamento desejado, diminuindo o número necessário de contatos externos e diminuindo a complexidade do circuito. O endereçamento é dividido em dois blocos de *shift register* — um para o endereçamento de colunas e o outro para o endereçamento de linhas. O endereçamento foi projetado tal que a última saída do bloco das colunas, $m+1$, é o sinal de *clock* do bloco das linhas. O *shift register* das colunas utiliza dois *clocks* não sobrepostos, para evitar problemas com as transições lentas do sinal de *clock* vindas de sinais externos ao chip. Além dos dois sinais de *clock* os sinais *Coll*, *Row1* e *start_stress* são sinais externos ao chip.

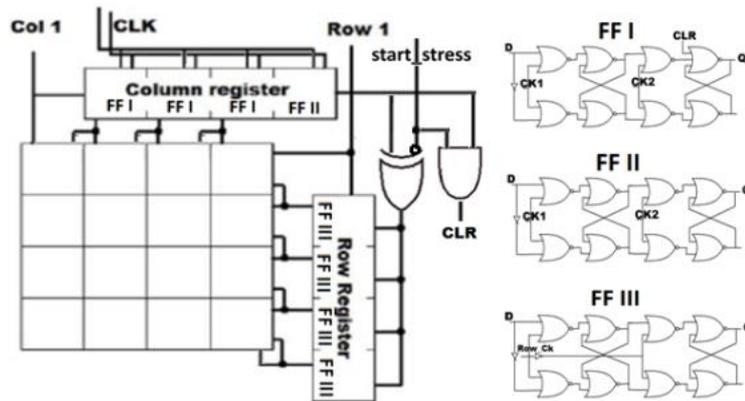


Figura 4.11 Bloco dos *shift registers*

No circuito proposto, para que uma nova linha comece a ser estressada, os *flip-flops* presentes no registrador das colunas, devem ser limpos imediatamente após o sinal de relógio, como mostrado na Figura 4.12. Assim, o circuito utiliza um sinal, chamado aqui de *start_stress*, junto com uma NAND ligada ao sinal $ColM+1$, controlando quando os valores das colunas devem ser limpos na chegada do *clock*. O sinal *start_stress* também controla quando o registrador das linhas é sensível à borda de descida (para os estados 2 e 6) ou sensível à borda de subida (para o estado 4), utilizando o sinal invertido e o sinal $ColM+1$ ligados a uma porta XOR. O controle da sensibilidade à borda de descida e subida é importante, pois quando em estado de medida o circuito deve trocar de linha quando o último flip-flop - que não está conectado a matriz - muda seu valor para zero. Quando em estado de estresse a mudança de linha deve ocorrer quando o último flip-flop muda seu valor para 1. Com essa abordagem se evita a espera de um período de medida na troca de linha e assim, se evita a diferença nos tempos de medidas ou de estresses.

A Figura 4.11 também mostra que os sinais *Row1* e *Col1* endereçam diretamente a primeira linha e a primeira coluna, respectivamente, e são também as entradas dos *shift registers*. Essa abordagem foi utilizada, pois assim a primeira célula de uma nova linha estará sobre estresse imediatamente quando o registrador das colunas for limpo, fazendo com que não haja atraso nessa célula quando o estresse mudar para a próxima linha, como mostrado na Figura 4.12.

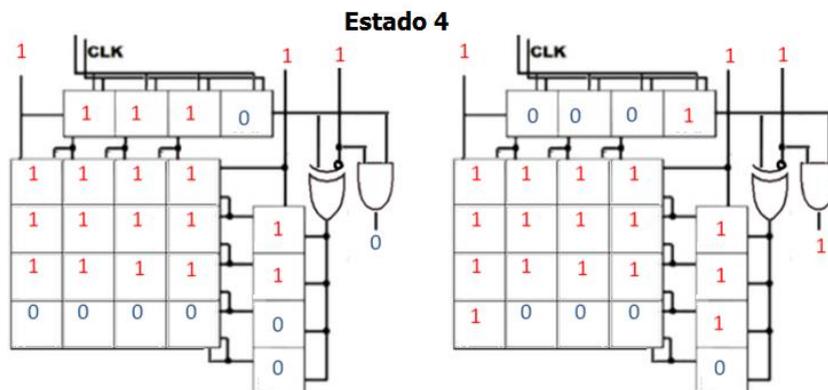


Figura 4.12 Exemplo da troca de linha quando em estresse.

O design apropriado dos *flip-flops* da coluna evita corridas quando o *shift register* limpa. Apenas os Flip-flops do tipo FF1 da Figura 4.11, são limpos pelo sinal, sendo que o design escolhido faz com que apenas a saída seja limpa mantendo o flip-flop utilizável mesmo após o sinal *CLR* mantido ativo. Assim, as saídas dos FF1s só trocarão de valor quando *CLR* se desativar, evitando corridas, conforme mostrado na Figura 4.13.

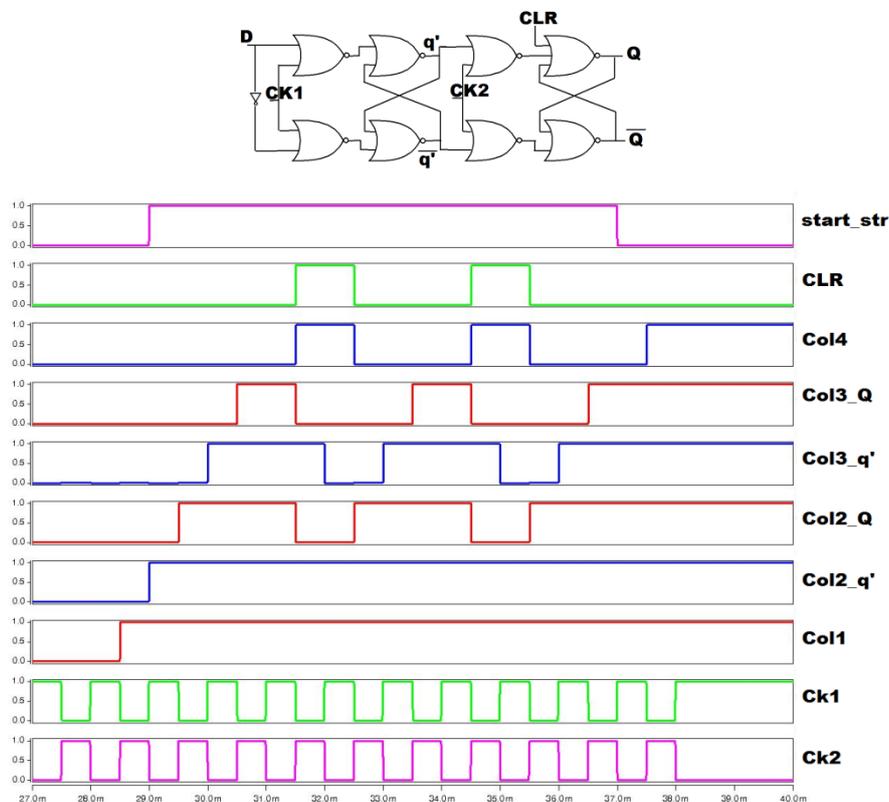


Figura 4.13 Funcionamento do *flip-flop* do tipo FF1

O comportamento proposto é alcançado completamente com a associação das células dos DUTs com os blocos de *shift registers* e com a sincronização externa dos sinais *CLK1*, *CLK2*, *Col1*, *Row1* e *Start_stress*. Os sinais externos de controle e os sinais internos dos registradores são mostrados na Figura 4.14 como exemplo para uma matriz 3x3.

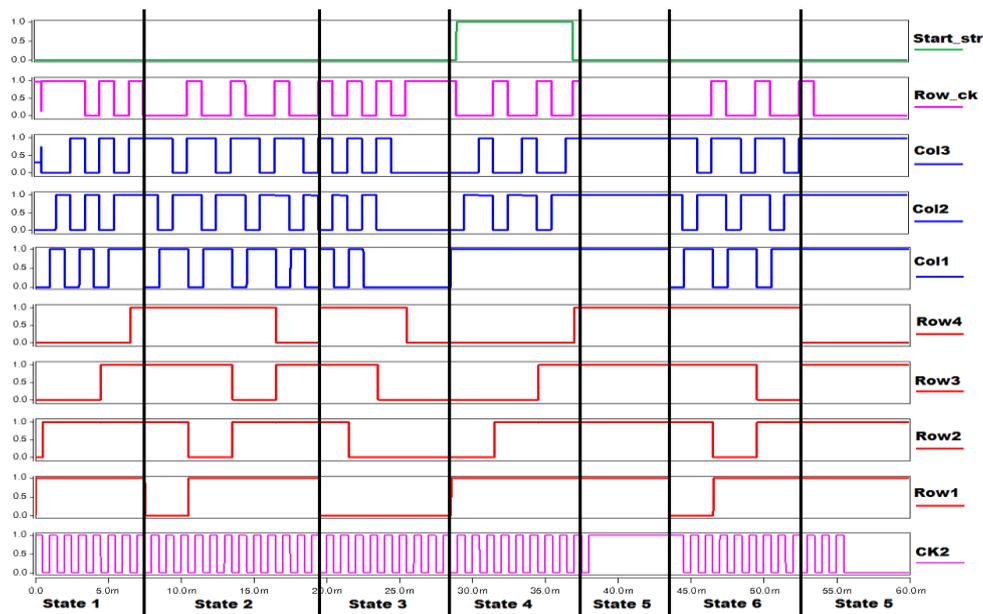


Figura 4.14 Sinais externos e internos para uma matriz 3x3

4.5 Uso do circuito para caracterização automatizada

Para utilizar o circuito corretamente os sinais de entrada *CLK1*, *CLK2*, *Col1*, *Row1* e *Start_stress* devem ser acionados de forma adequada, a fim de colocar o circuito nos seis estados descritos anteriormente. Logo, foi criado um script na linguagem Perl, que gera todos sinais externos de acordo com o descrito abaixo:

4.5.1 Estado 1 – Preparação para medida de V_{th}

Nesse estado o circuito deve ficar pronto para o início da medida dos DUTs. Assim, ao final do estado 1 os sinais *meas_stress* devem estar em nível alto em todas as células da matriz, ou seja, sinais de endereçamento das colunas e das linhas devem estar em nível alto ao final do estado 1. Com nível alto no sinal *meas_stress* a porta dos DUTs estarão ligadas à tensão V_{stress} . e então, para evitar o estresse antecipado dos transistores a tensão V_{stress} , no estado 1, deve ser uma tensão que não cause estresse, ou seja, $V_{stress} = VDD$ para PMOS e $V_{stress} = GND$ para transistores do tipo NMOS.

Para que todos sinais de endereçamento fiquem em nível alto no final do estado, a entrada *Row1* deve permanecer em nível alto, enquanto *Col1* deve alternar entre nível alto e baixo gerando o sinal de *clock* para o registrador das linhas. Após todas as linhas estarem em nível lógico 1 então *Col1* permanece em 1 até que todas as colunas fiquem em nível alto.

Ao final, após $(1 + N_{linhas} + 2 \cdot N_{colunas})$ sinais de *clock* todos *flip-flops* estarão em nível alto e o circuito estará pronto para a realização das primeiras medidas, conforme Figura 4.15.

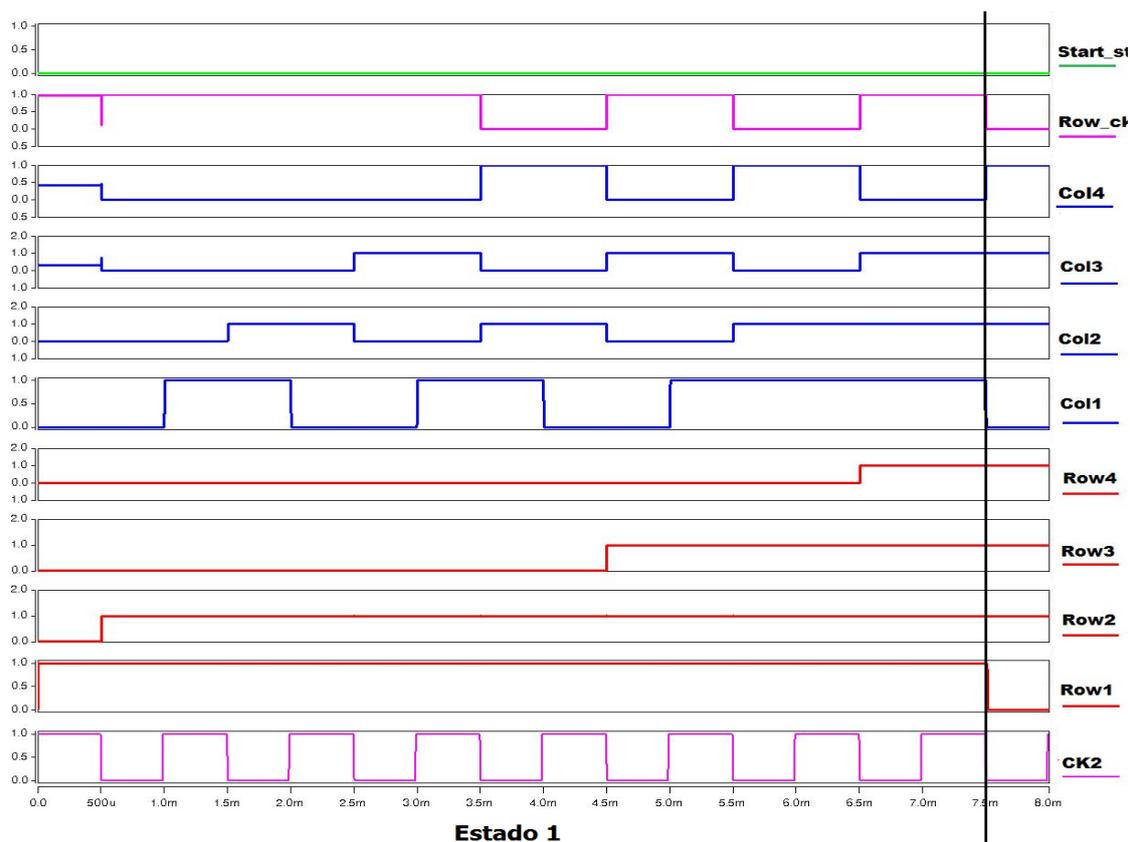


Figura 4.15 Comportamento dos sinais para definir nível alto em todas linhas e colunas.

4.5.2 Estado 2 – Medida de V_{th}

No estado 2 são feitas as medidas iniciais de V_{th} sequencialmente em cada DUT. Um DUT é medido quando o sinal $stress_meas$ está em nível baixo, conectando a porta do DUT na tensão V_{meas} e ligando seu dreno às conexões de $force$ e $sense$. Para que o nível baixo no sinal $stress_meas$ percorra sequencialmente todos os DUTs o sinal $Row1$ deve permanecer em nível baixo até que as medidas da primeira linha terminem e a segunda linha fique em nível baixo, ou seja, $Row1$ deve estar em 0 lógico por $N_{colunas}+1$ períodos de $clock$ e depois mudar seu valor para 1. $Col1$ deve ser 0 lógico por 1 período de $clock$ depois deve ser 1 lógico por $N_{colunas}$ períodos de $clock$, repetindo esse comportamento por N_{linhas} vezes. A correta medida de V_{th} é alcançada tendo o período do sinal de $clock$ maior que o tempo de medida do V_{th} . Após $N_{coluna} \cdot N_{linha}$ períodos de $clock$ todas as medidas são feitas, conforme Figura 4.16:

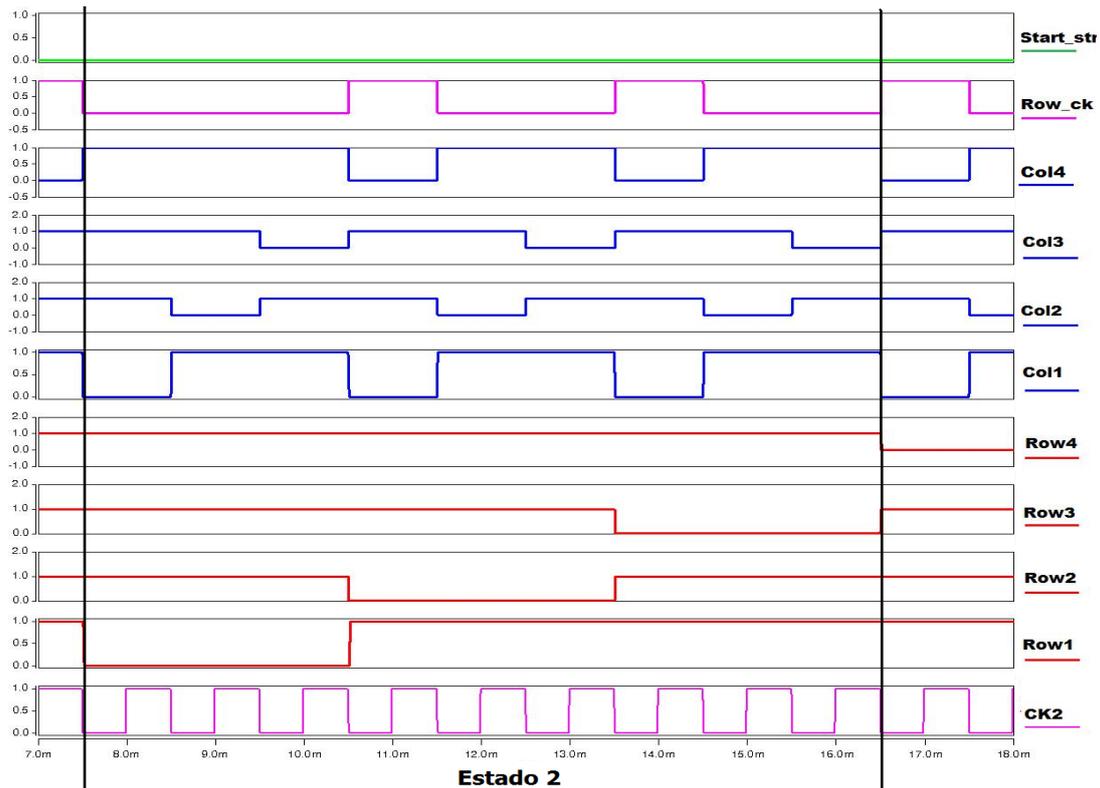


Figura 4.16 Comportamento dos sinais para medir V_{th} em todos DUTs sequencialmente

4.5.3 Estado 3 – Preparação para o estresse

No estado 3 o circuito deve ficar pronto para o início do estresse, assim todos os DUTs devem estar conectados a V_{meas} , ou seja o sinal $meas_stress$ deve ser 0 para todas as Células BTI. Para que isso ocorra o comportamento das entradas deve ser semelhante ao estado 1. *Row1* deve permanecer em nível baixo (0), enquanto *Col1* deve alternar entre nível alto e baixo gerando o sinal de *clock* para o registrador das linhas. Após todas as linhas estarem em nível lógico 0, *Col1* permanece em 0 até que todas as colunas fiquem em nível alto.

Ao final, após $(1 + N_{linhas} + 2 \cdot N_{colunas})$ sinais de *clocks* todos *flip-flops* estarão em nível baixo. V_{stress} , então, deverá ter a tensão escolhida para o estresse, $V_{stress} < V_{dd} - |V_{th}|$ para PMOS e $V_{stress} > GND + V_{th}$ para transistores NMOS e assim o circuito estará pronto para a o início do estresse, conforme Figura 4.17.

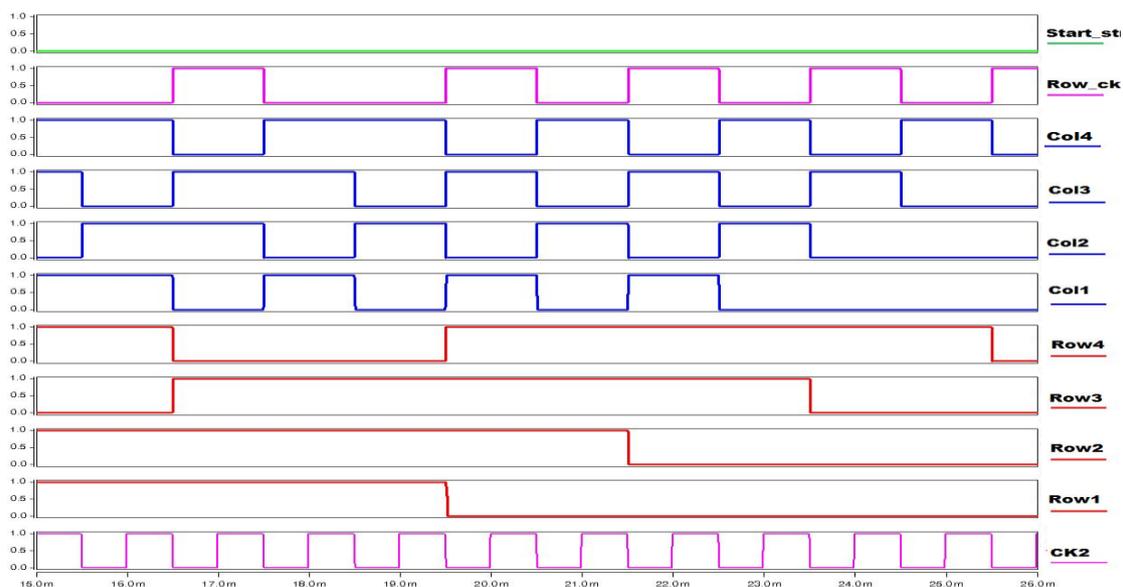


Figura 4.17 Comportamento dos sinais para definir nível baixo em todas linhas e colunas.

4.5.4 Estado 4 – Início do estresse

No estado 4 o estresse deve começar serialmente em cada DUT, com um atraso de um período de medida entre cada DUT, subsequente. Para que isso ocorra como desejado os sinais *Col1* e *Row1* devem permanecer em nível alto. O sinal *Start_stress* deve ser colocado em nível alto no início do estado 4 e deve ter seu valor colocado em nível baixo antes do último *clock* evitando que ambos *start_stress* e *ColM+1* fiquem em nível lógico 1 e acabe limpando todos os FFs ao final do estado. Assim, foi escolhido colocar *Start_stress* em 0 durante meio período de *clock*, antes do estado 4 acabar, ou seja, antes do ciclo de *clock* número $N_{colunas} * N_{linhas}$ do estado, conforme Figura 4.18.

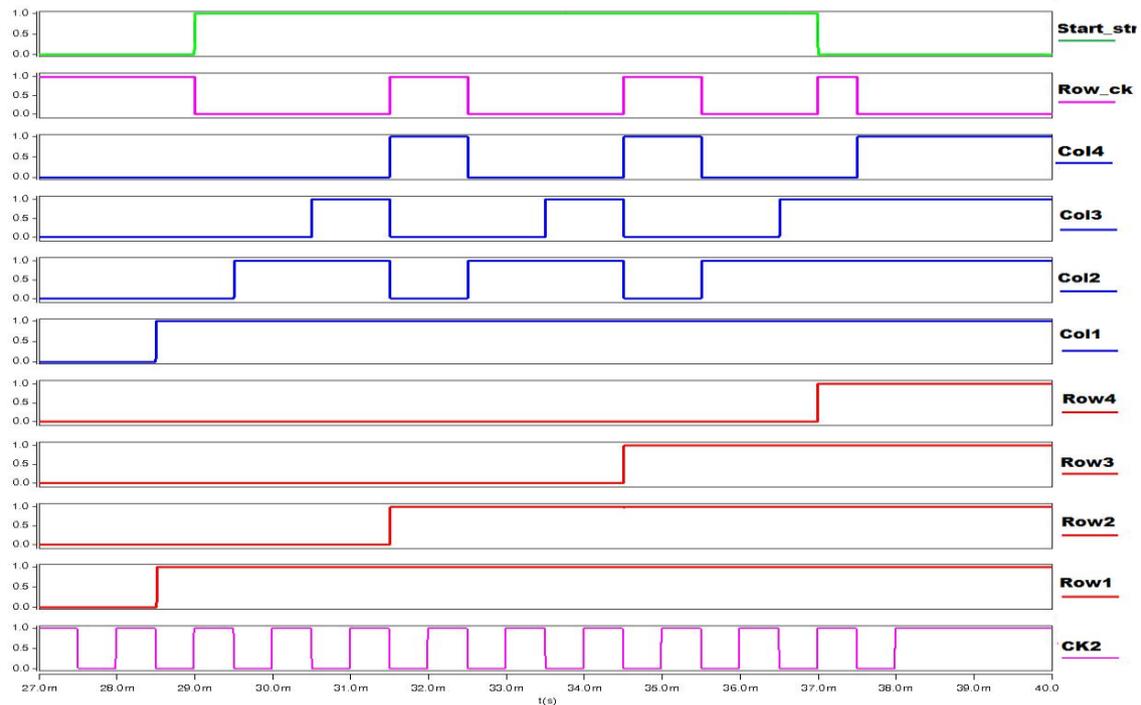


Figura 4.18 Comportamento dos sinais para colocar todos DUTs sob estresse.

4.5.5 Estado 5 - Estresse

No estado 5, todos os transistores estão sob estresse paralelamente, esperando o fim do período de estresse escolhido.

4.5.6 Estado 6 – Medida de ΔV_{th}

O estado 6 é semelhante ao estado 2 onde os transistores começam a ser medidos serialmente. Porém, nesse estado o período de medida é menor que o necessário no estado 2, pois apenas a medida de um ponto na corrente é necessária.

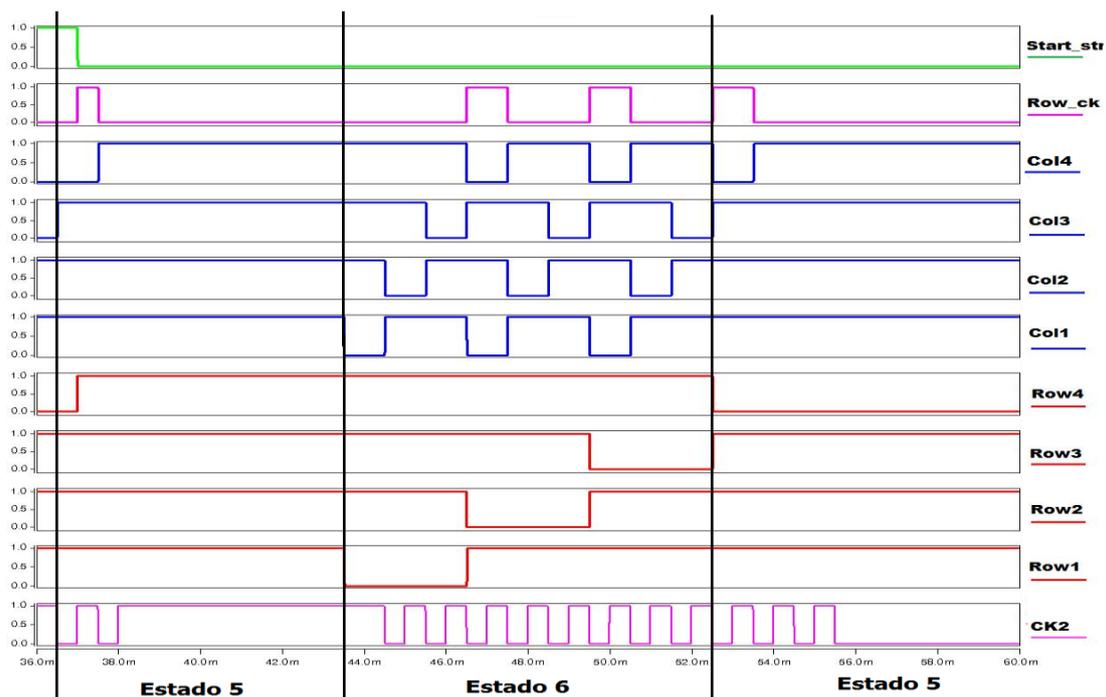


Figura 4.19 Comportamento dos sinais para medir ΔV_{th} sequencialmente em todos os DUTs.

4.5.7 Tempos de *Hold* e *Setup*

Como mostrado na Figura 4.1, *Col1* e *Row1* endereçam a primeira coluna e a primeira linha diretamente e também servem de entrada para os respectivos registradores das colunas e linhas. Uma vez que *Col2* e *Row2* mudam imediatamente após a descida do sinal *CK2*, é importante que os sinais *Col1* e *Row1* (que são sinais externos ao chip) sejam mudados de nível lógico, quando necessário, o mais próximo da mudança do sinal de *CK2*. Assim garantindo, por exemplo, que os tempos em que *Col1* e os demais sinais de coluna passam em nível lógico baixo (no caso de medida mostrado na Figura 4.19) sejam tão similares quanto possível, e que tenham pouca sobreposição de seus sinais.

No caso do registrador das colunas, que funciona com dois sinais de *clock*, se pode assegurar que *Col1* pode mudar de valor juntamente com *CK2* (Secção 6.1). Porém, o registrador das linhas depende do sinal de *clock* gerado pelo registrador das colunas. Assim, esse *clock* (*ck_row*) é gerado com atraso após a mudança de *CK2*, e se deve introduzir um tempo de *hold* para poder mudar o sinal de *Row1*. Isto causa uma leve sobreposição entre os sinais *Row1* e *Row2*, dado pelo tempo de *hold* (T_{hold}), mostrado na Figura 4.20.

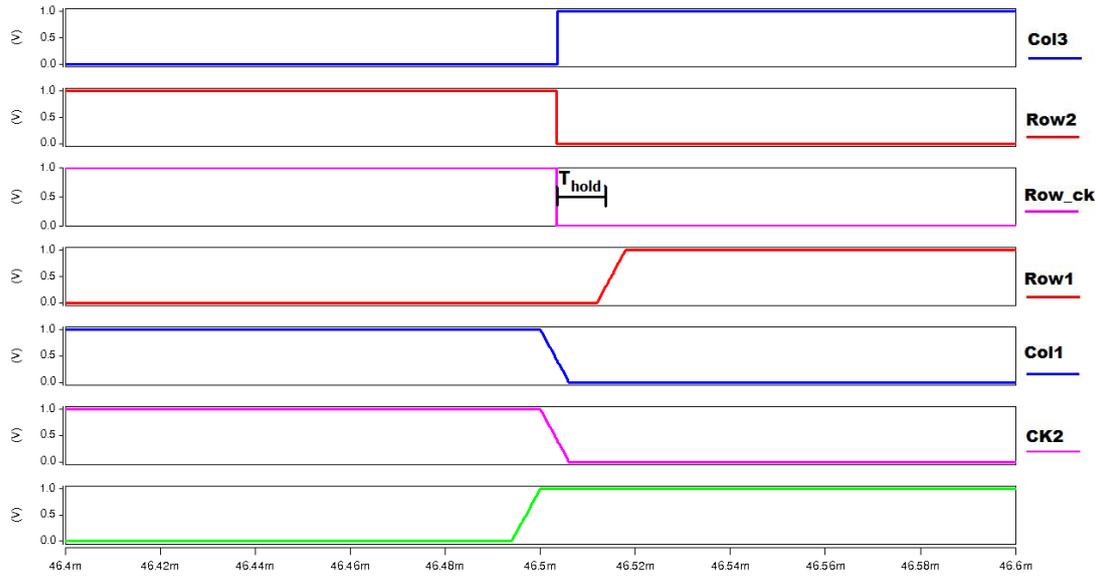


Figura 4.20 Ilustração do tempo necessário a se esperar para mudar Row1 após a mudança de CK2.

5 CONSIDERAÇÕES DO LEIAUTE

O leiaute foi realizado considerando métodos para tecnologias do estado-da-arte que são empregados em nós tecnológicos abaixo de 45 nanômetros. Tais métodos consistem na utilização de grade fixa para a camada de poly e metall e no emprego de estruturas *dummy* que asseguram a correta litografia das camadas. O layout final implementado foi realizado na tecnologia de 28nm MG-HK do imec, que utiliza *gate* de metal, dielétrico *high-K*, grade fixa de *poly* e de metall e utiliza dois níveis de conexões locais intermediárias entre região de difusão e metall.

O projeto do circuito proposto é realizado no estágio de caracterização desta nova tecnologia. Neste estágio ainda não se dispõe de bibliotecas de *standard cells*. Além disto, as características dos transistores, conexões, vias, e capacitâncias parasitas, entre outros, são meras previsões, por ainda não se dispor de modelos elétricos apropriados. Assim, o leiaute do circuito deve ser feito utilizando a metodologia *full-custom*. Nesta metodologia cada transistor e cada conexão entre os transistores devem ser feitos individualmente.

Devido à escolha da metodologia *full-custom* e o fato desta requerer que um circuito tenha seu leiaute completamente refeito quando a tecnologia ou as regras de desenho forem alteradas, o circuito proposto foi feito utilizando SKILL script, utilizado pelos softwares da Cadence®. O script SKILL é utilizado para a criação de células parametrizadas, chamadas de PCells.

PCell é um conceito utilizado largamente para o design automático de circuitos analógicos, bem como em alguns casos digitais. PCells formam blocos do circuito que são dependentes de parâmetros variáveis, sendo possível ser instanciada muitas vezes, tendo cada instância diferentes parâmetros ou não. Por exemplo, ao invés de ter diferentes células para cada tamanho de transistor utilizado no projeto, uma única PCell pode ter como parâmetros as dimensões, largura e comprimento, do transistor. Então, devido a escolha de se fazer um leiaute baseado em PCells, pode-se gerar um vasto número de diferentes células, que se diferenciam por alguns parâmetros e assim aumentando a produtividade e a capacidade de se fazer pequenas alterações no projeto sem precisar refazer grande parte de seu leiaute.

5.1 Célula Básica

Todo o leiaute do circuito proposto foi derivado de uma única PCell, chamada aqui de Célula Básica, da qual todas outras células, NANDs, NORs, Flip-Flops e etc, são derivadas. A idéia principal por trás dessa abordagem é conseguir um leiaute bem estruturado e regular, de fácil modificação, de fácil tradução para novas tecnologias e para facilitar a criação de novas células.

A Célula Básica, Figura 5.1, contém os elementos principais para a criação de outras células, tais como: a grade de *poly*, grade de metall, regiões de difusão do *pull-up* e *pull-down*, trilha de GND e VDD, contato do body-bias, região de dopagem, de poço e etc. Sendo assim ela foi projetada para ser a responsável por guardar os parâmetros das regras de desenho da tecnologia. Cada posição, distância e dimensão de cada elemento da célula, que são dependentes das regras de desenho da tecnologia (posição da região de difusão do pull-up e pull-down, distância entre polys, largura do poly, distância entre

metal e etc), podem ser facilmente trocados e o layout automaticamente se ajusta aos novos parâmetros. Além das regras de desenho, também podem ser mudados valores como: largura da região de difusão, quantidade de trilhas de metal na grade, quantidade de polys na grade, a retirada de alguma trilha de metal, a inclusão de *gaps* na trilha de metal e etc.

Essa célula foi projetada para facilitar o posicionamento e o roteamento utilizando com conexões em apenas uma dimensão. Assim, ela foi projetada para que ao ser colocada lado a lado ela se comporte como mostrado na Figura 5.2. Para isso, o roteamento das conexões é feito introduzindo *gaps* nas trilhas de metal através das propriedades da Célula Básica.

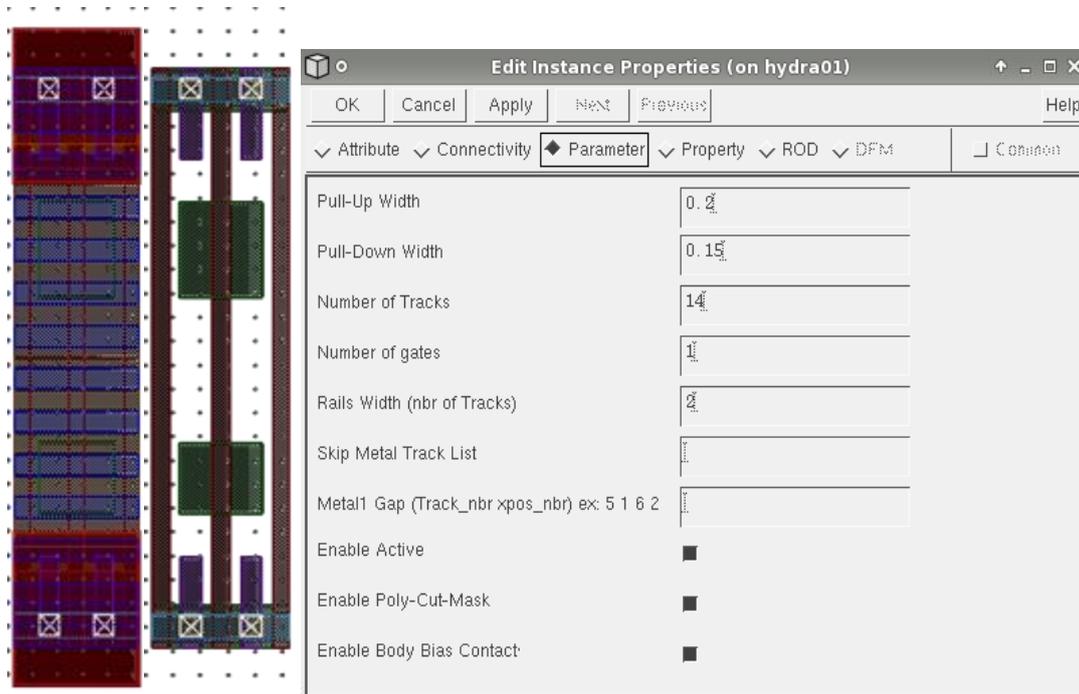


Figura 5.1 a) Célula Básica completa e b) sem algumas camadas. c) Menu para alteração de parâmetros manualmente.

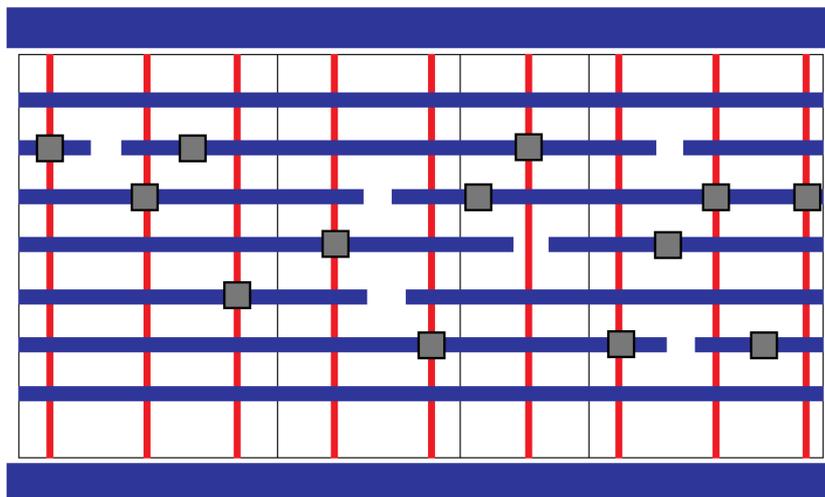


Figura 5.2 Comportamento projetado ao se colocar células básicas lado-a-lado.

Na Figura 5.3 é mostrado o exemplo de como derivar uma célula NAND2 a partir da Célula Básica. A nova célula é feita configurando os parâmetros *number of gates* para dois e removendo a trilhas de metal perto do VDD e GND, por se tratar de metais mais largos. Na vista mais a esquerda, pode-se ver que os elementos que devem ser adicionados à célula são apenas as conexões intermediárias do *pull-up* e do *pull-down* e suas vias. Pode-se notar também que as conexões intermediárias são desenhadas para que tenham o maior comprimento possível, assim, se tornando estruturas que ajudam a reforçar as demais conexões intermediárias ao seu lado. Como vistos nas outras duas vistas (do meio e da direita) o resultando final é um leiaute bem estruturado e regular.

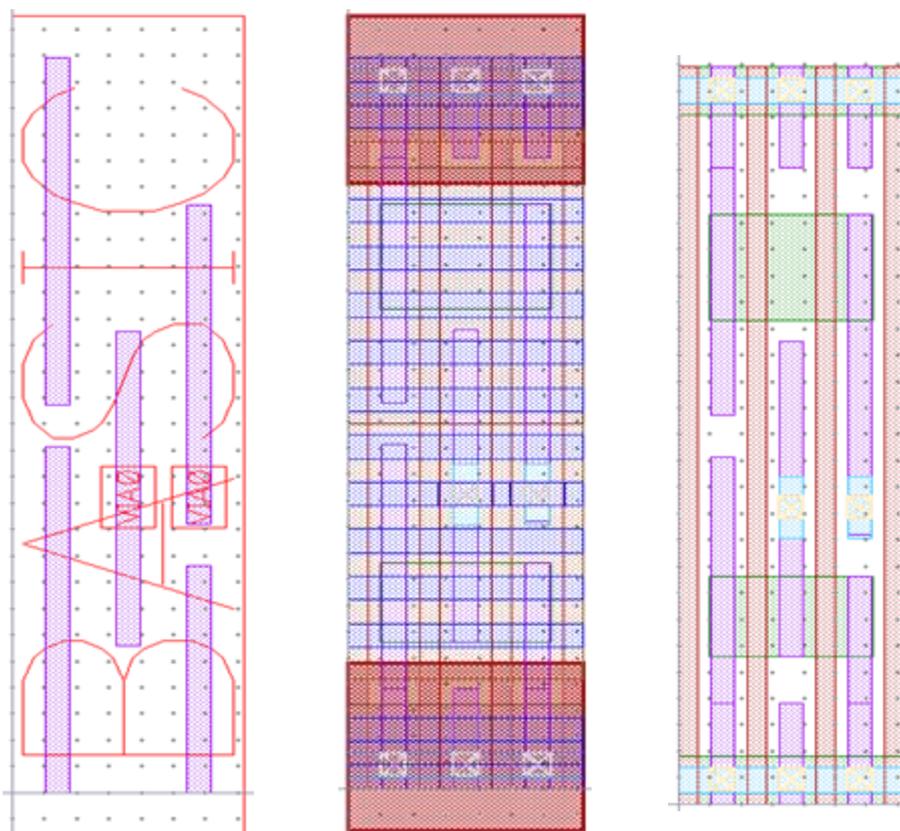


Figura 5.3 Exemplo de uma NAND2 criada a partir de uma Célula Básica.

5.2 Célula BTI

Seguindo o esquemático da Figura 5.2 todas as células necessárias (NAND2, AOI21, NAND3, MUX21, TGs e o DUT) foram colocadas lado a lado tendo seu roteamento feito através de gaps na trilha de metal e através de células de roteamento, para troca do sinal verticalmente entre as trilhas de metal. A célula BTI da matriz também contém todas as conexões de ColM, RowN, Vstress, Vmeas, Force, Sense, sendo apenas ColM uma conexão vertical. O leiaute completo de uma célula da matriz ocupa uma área final igual a $5.5\mu\text{m} \times 2.1\mu\text{m}$, podendo ser dividida em duas partes, a parte que contém apenas as conexões horizontais (parte de cima da célula) e a parte que contém as células (parte de baixo da célula), como visto na Figura 5.4, Figura 5.5 e Figura 5.6.

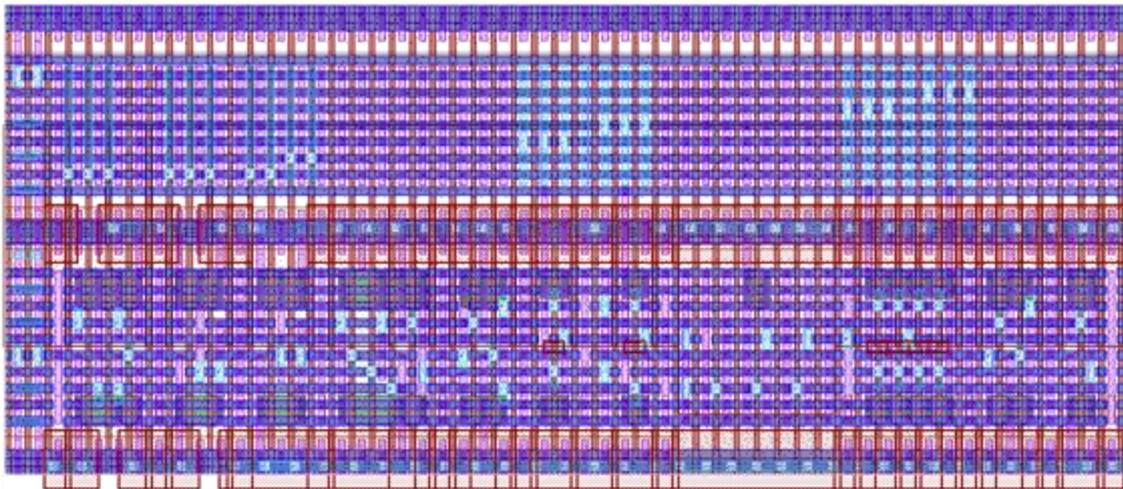


Figura 5.4 Vista completa de todas camadas da Célula BTI.

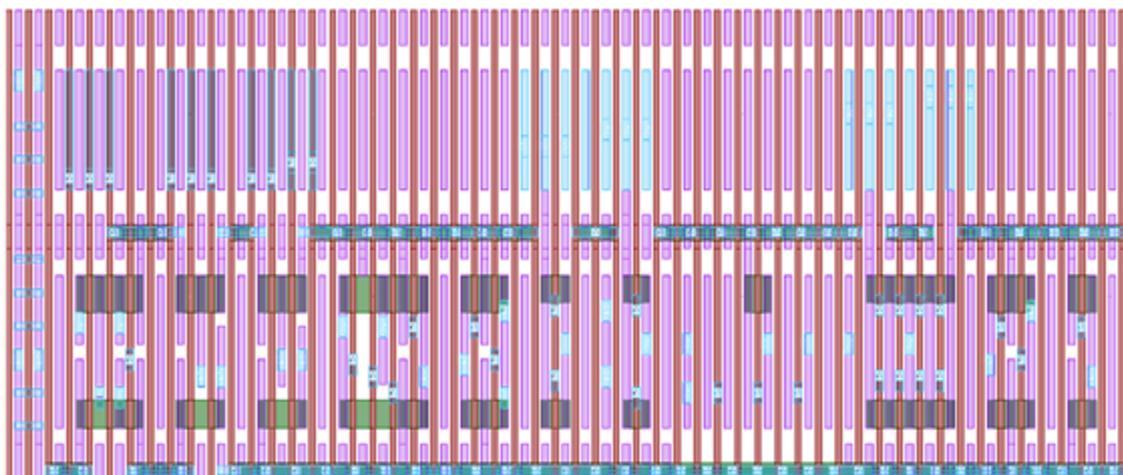


Figura 5.5 Vista das camadas de conexão intermediária da Célula BTI

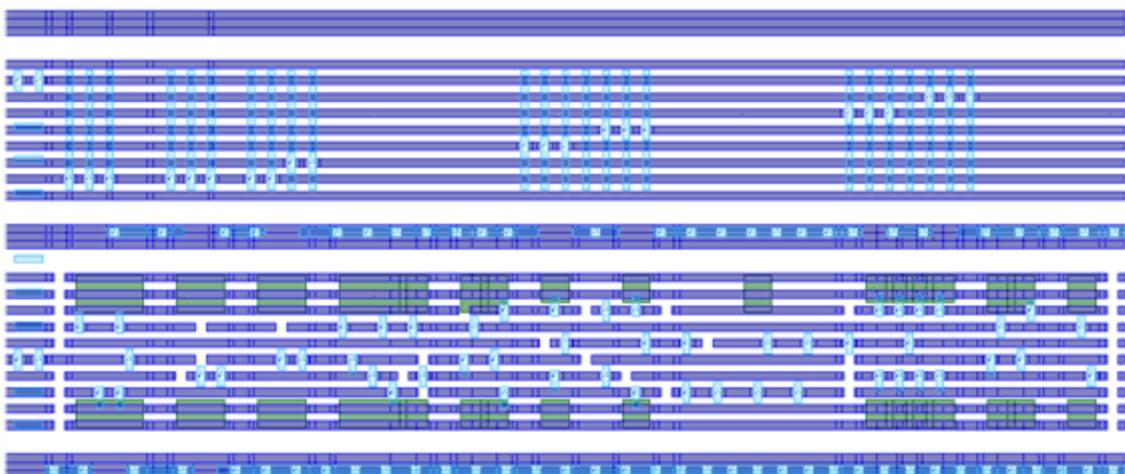


Figura 5.6 Vistas das conexões da Célula BTI.

5.3 Circuito Completo

Para formar o circuito completo as Células BTI foram posicionadas para formar uma matriz, enquanto os *shift-registers* foram colocados em torno da matriz, como pode ser visto no exemplo de uma matriz 3×3 mostrado na Figura 5.7 e na Figura 5.8.

No leiaute final se utilizou uma matriz 30×3 para otimização da área ocupada pelo circuito. Assim, todo o circuito cabe nos módulos de contatos padrão do IMEC. O circuito completo, mostrado na Figura 5.9a, com 90 DUTs, shift-register das colunas, shift-register das linhas, diodos de proteção, a lógica remanescente e as conexões, ocupa uma área de 30um × 76um. O circuito completo pode ser ainda menor, caso se utilize mais camadas de metal. Com estas dimensões o circuito é pequeno o suficiente para caber nas *drop-in/dice lanes* e, assim, pode ser usado também para monitorar o processo de fabricação.

Cada circuito ocupa 11 *pads* de contato (*CLK1*, *CLK2*, *start_stress*, *v_meas*, *v_stress*, *force*, *sense*, *VDD* e *GND*) o que significa que dois circuitos (180 DUTs) podem ser colocados em cada módulo de 24 *pads*, Figura 5.9b.

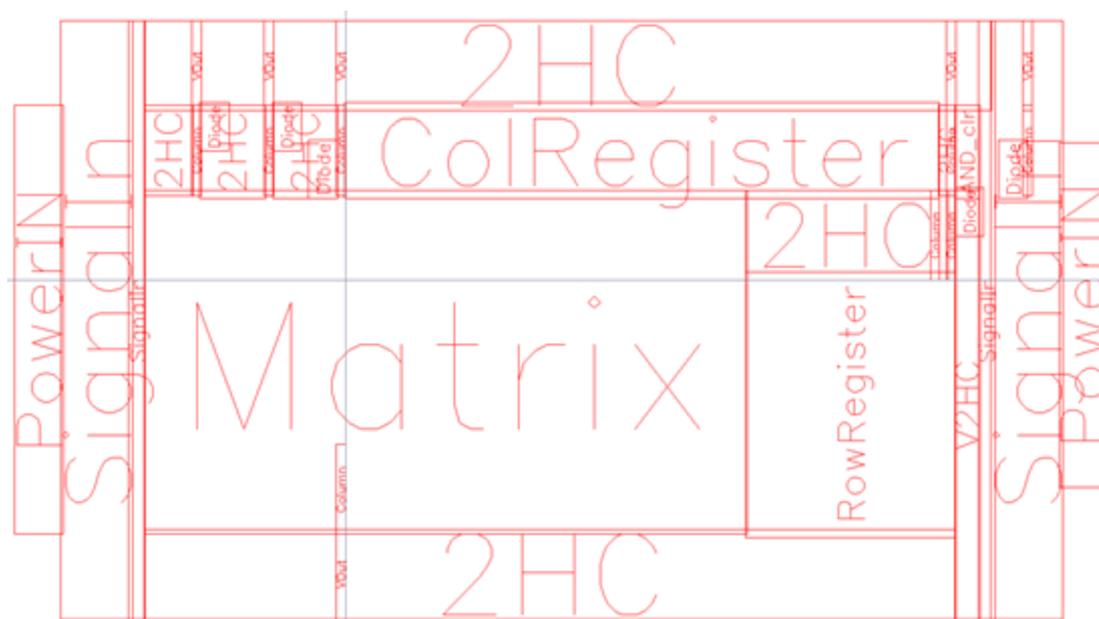


Figura 5.7 Disposição dos blocos no circuito completo. Exemplo com uma matriz 3×3.

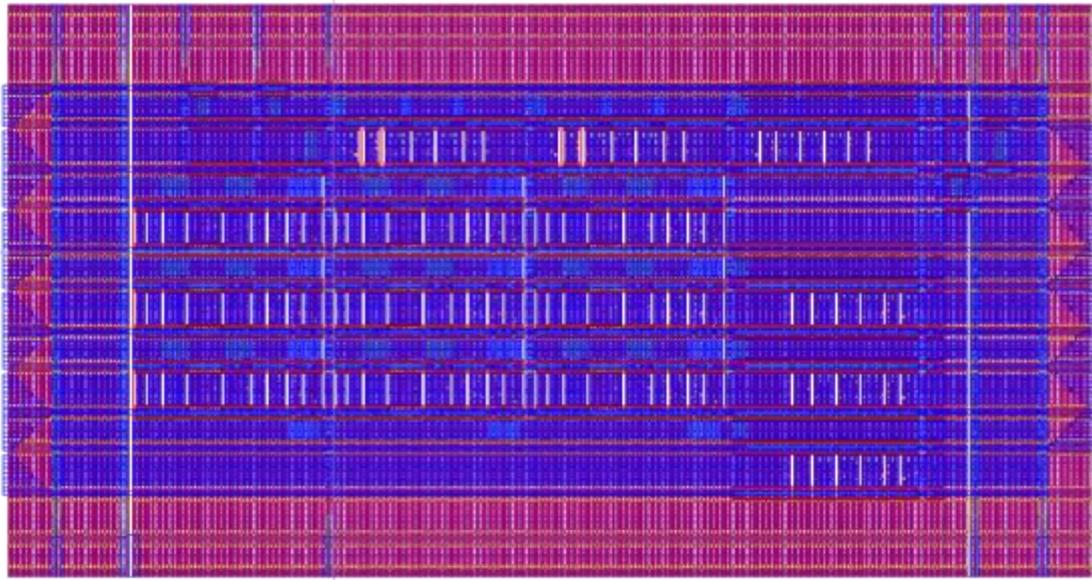


Figura 5.8 Visão geral do circuito completo. Exemplo com uma matriz 3×3.

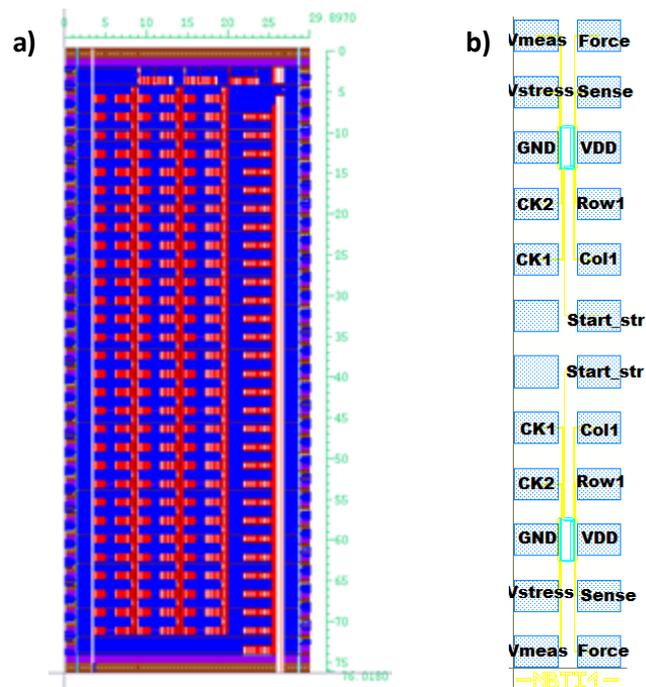


Figura 5.9 a) Circuito completo 30×3 b) disposição de 2 circuitos 30x3 no módulo de 24 pads.

5.4 Seis Módulos BTI

No total foram utilizados 6 módulos, sendo 3 destes módulos para medidas de NBTI em transistores PMOS e outros 3 módulos para a medida de PBTI em transistores NMOS, como mostrado na Figura 5.10. Cada módulo, como já mencionado, contém duas matrizes para medida de BTI. Para cada matriz foram escolhidos dois tipos diferentes de tamanhos de transistores, resultando, assim, em 24 tipos diferentes de medida, 90 transistores por matriz, com 45 transistores com a mesma geometria. Resultando na caracterização de 1080 transistores no total.

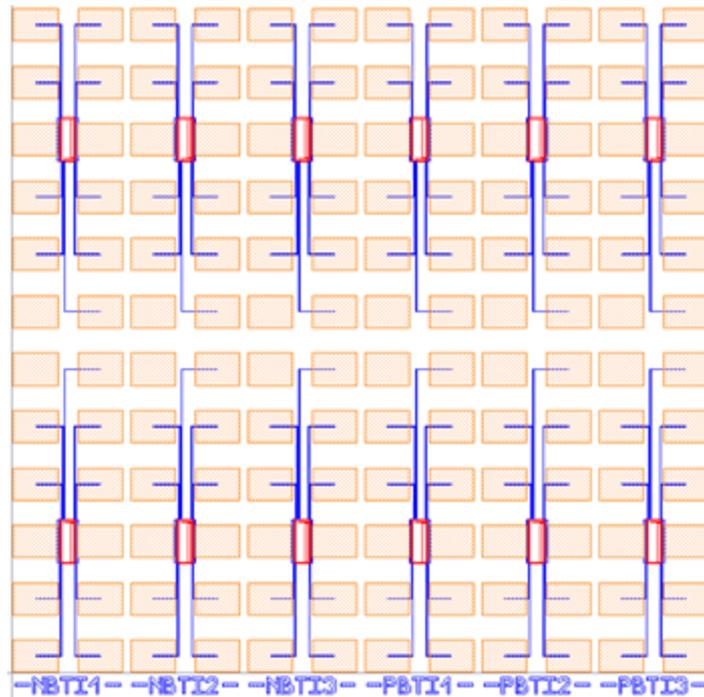


Figura 5.10 Seis módulos utilizados totalizando 1024 DUTs.

Geometrias semelhantes foram escolhidas para compartilhar a mesma matriz, assim, o impacto do *leakage* na medida de ΔV_{th} e o impacto da diferenças entre o chaveamento de estresse para medida e de medida para estresse é minimizado. Isto permite tempos de estresses e medidas similares em todos os dispositivos da matriz. As geometrias escolhidas são presentes na Tabela 5.1.

Tabela 5.1 Geometrias utilizadas e suas localizações.

Nome do Módulo	Posição1	Posição2	W	L
NBTI1	Abaixo	1 - 44	60n	30n
		45 - 90	90n	30n
	Acima	1 - 44	150n	30n
		45 - 90	250n	30n
NBTI2	Abaixo	1 - 44	60n	40n
		45 - 90	150n	40n
	Acima	1 - 44	60n	70n
		45 - 90	90n	70n
NBTI3	Abaixo	1 - 44	90n	40n
		45 - 90	400n	30n
	Acima	1 - 44	60n	120n
		45 - 90	60n	210n
PBTI1	Abaixo	1 - 44	60n	30n
		45 - 90	90n	30n
	Acima	1 - 44	150n	30n
		45 - 90	250n	30n

PBTI2	Abaixo	1 - 44	60n	40n
		45 - 90	150n	40n
	Acima	1 - 44	60n	70n
		45 - 90	90n	70n
PBTI3	Abaixo	1 - 44	90n	40n
		45 - 90	400n	30n
	Acima	1 - 44	60n	120n
		45 - 90	60n	210n

5.5 Benefícios do circuito

O leiaute final do circuito, como já mencionado, contou com seis módulos resultando num total de 1080 transistores. Atualmente, para a medida individual de um transistor se ocupa 4 pads em um módulo, assim, apenas possibilitando a medida de 6 transistores por módulo. Logo, para que 1080 transistores sejam medidos, atualmente, necessitaria de um total de 270 módulos. Utilizando o método proposto se utilizou apenas 6 módulos, o que leva a uma eficiência em área de 40 vezes.

Com o leiaute final do circuito também se consegue grandes ganhos quanto ao tempo gasto na caracterização dos dispositivos. Atualmente, para se caracterizar 1080 transistores se levaria aproximadamente $1080 \times (\text{tempo total de estresse})$, desconsiderando o tempo de medida (para T_{meas} pequeno). Já com o leiaute realizado os 1080 transistores são distribuídos em 12 circuitos individuais assim o tempo de caracterização é de aproximadamente $12 \times (\text{tempo total de estresse})$, resultando numa eficiência no tempo de caracterização de 90 vezes.

Vale lembrar que ambos benefícios na área e no tempo de caracterização crescem ao se escolher matrizes com mais transistores, ao invés dos 90 transistores escolhidos por matriz poderiam ser feitos circuitos com 1000 transistores, por exemplo. Como uma matriz de 1000 transistores não cabe no interior de um módulo fica mais difícil achar o benefício do método na área. Porém, considerando que essa matriz ocupa o espaço de dois módulos, o benefício na área seria de 167 vezes e o benefício de tempo seria de 1000 vezes.

A Secção a seguir apresenta dados da simulação do circuito pós-leiaute com a matriz de 90 transistores, analisando esses dados será visto que circuitos bem maiores são possíveis.

6 SIMULAÇÕES

Com o objetivo de testar a eficácia do circuito, o *netlist* gerado após o leiaute foi submetido a simulações que visavam testar a eficiência do leiaute, capacidade dos *buffers* e tempos de respostas. Para isso foi utilizado o simulador HSPICE.

A utilização do circuito juntamente com a degradação dos transistores também foi simulada com a ajuda do simulador apresentado em (KACZER, 2011) combinado com o HSPICE. As simulações efetuadas no circuito são mostradas nas seções abaixo.

6.1 Caracterização do Circuito

Primeiramente, os tempos de *hold* e *setup* dos *flip-flops* foram determinados obtendo uma estimativa para a geração dos sinais externos *Col* e *Row1*. Os tempos de *hold* e *setup* foram achados para diferentes *slopes* dos sinais de entrada (*CK1*, *CK2* e *Coll*). Os tempos de *setup* foram calculados como: tempo mínimo que o sinal *Coll* tem que estar estável antes da subida de *CK1* para haver a mudança (subida ou descida) em *Col2*. Já os tempos de *hold* foram achados como: tempo mínimo que o sinal *Coll* tem que ficar estável após a descida do sinal *CK2*. Todos os tempos foram simulados para $VDD = 1V$.

Tabela 6.1 Tempos de *setup* e *hold* para os *flip-flops* das colunas.

<i>Slope</i>		5 μ s	1 μ s	500ns	100ns	50ns	10ns
T_{Setup}	Subida	689ns	138ns	69n	14ns	7ns	2ns
	Descida	1,38 μ s	283ns	142ns	29ns	15ns	3ns
T_{Hold}	Subida	-6,4 μ s	-1,3 μ s	-642ns	-129ns	-64ns	-12ns
	Descida	-5,6 μ s	-1,1 μ s	-569ns	-111ns	-56ns	-11ns

De acordo com os valores encontrados não há preocupação quanto aos tempos de *hold* e *setup* na configuração proposta dos registradores das colunas, Seção 4.5. Sendo assim, o sinal de *Coll* pode ser mudado junto com *CK2*, ou seja, com tempo de *hold* igual a zero, como proposto em 4.5.

Os tempos de *hold* também foram encontrados para os *flip-flops* responsáveis pelas linhas. Ao invés de se utilizar o sinal de relógio interno *Row_Ck*, gerado pelos registradores das colunas, se calculou os tempos de *hold* relativos a descida do sinal *Ck2*, por este ser o sinal de entrada conhecido. Assim, os tempos calculados acabam sendo a soma do atraso para a geração do sinal *Row_Ck*, no último *flip-flop* das colunas, com os tempos de *hold*, dos *flip-flops* das colunas, relativos a *Row_Ck*. Como dito na seção 4.5.7 e mostrado na Figura 4.20 o T_{hold} do *flip-flop* da linha é responsável pela sobreposição do sinal de medida entre 2 dispositivos em linhas diferentes, portanto, sua minimização é importante.

Tabela 6.2 tempos de *hold* para os *flip-flops* das linhas.

<i>slope</i>		5 μ s	100ns	10ns
T_{Hold}	Subida	488,9 ns	7,86 ns	1,03 ns
	Descida	-489 ns	7,87 ns	1,04 ns

A tabela Tabela 6.3 mostra a caracterização do atraso e do *slope* nos sinais internos do circuito. Este é um conjunto de medidas, feitas na simulação do *netlist*, relevante para se caracterizar o circuito e para verificar seu correto funcionamento. Foram realizadas medidas nas interconexões dos sinais *Row_Ck* e *ColM* medindo o atraso e o *slope* para ambos os sinais propagarem da primeira linha da matriz até a última linha 30, também foram medidos atrasos e *slopes* nos *flip-flops* responsáveis pelas linhas e colunas.

Tabela 6.3 Caracterização do atraso e do *slope* nos sinais internos do circuito.

Medidas	Tipo	Tipo	Slope do sinal CK2		
			5μ	100n	10n
<i>Ck2</i> para <i>Row_Ck</i>	Atraso	Subida	489ns	7,87ns	1,04ns
<i>Row_Ck</i>	Slope	Subida	116ps	113ps	124ps
<i>Row_Ck</i> (FF1 para FF30)	Atraso	Subida	51,2ps	48,8ps	54,46ps
<i>CK2</i> para <i>ColM</i>	Atraso	Subida	535,2ns	9,39ns	1,16ns
<i>CK2</i> para <i>ColM</i>	Atraso	Descida	438,7ns	7,72ns	0,96ns
<i>ColM</i>	Slope	Subida	66,0ps	60,13ps	67,03ps
<i>ColM</i>	Slope	Descida	45,8ps	44,01ps	46,74ps
<i>ColM</i> (Célula1 para Célula30)	Atraso	Subida	31,9ps	30,2ps	30,79ps
<i>ColM</i> (Célula1 para Célula30)	Slope	Descida	19,9ps	19,32ps	19,48ps

Através dos tempos achados se deduz que a frequência máxima de operação do circuito está acima do proposto, que é cerca de 1 khz (período aproximado de 1ms, para período de medida igual a 1ms). Também se percebe que a utilização do menor *slope* de entrada possível diminui os atrasos no circuito e por consequência diminui o erro nos tempos de medida e estresses causados pela sobreposição dos sinais.

Dado os pequenos atrasos e *slopes* encontrados após a caracterização do circuito com uma matriz de 90 transistores, pode-se dizer que matrizes maiores são possíveis de serem utilizadas sem modificações maiores no circuito.

6.2 Medindo V_{th}

Para testar as influências (correntes de fuga, capacitâncias e resistências parasitas) do circuito de controle, das chaves e dos multiplexadores, na capacidade de se medir variações na corrente I_d do DUT, os métodos explicados na seção 3.1 foram aplicados também no *netlist* final do circuito. O resultado é apresentado na Tabela 6.4, juntamente com a comparação em relação às medidas feitas anteriormente (com o transistor sendo medido isoladamente). O erro calculado é devido principalmente ao *leakage* encontrado quando nenhum transistor está sendo medido e a corrente deveria ser zero. O *leakage* encontrado, de aproximadamente -5 nA para todos os DUTs, influencia no cálculo de V_{th} , adicionando-se à corrente medida e prejudicando o desempenho do *force* e *sense* alterando V_d . Porém esse *leakage* pode ser determinado, sendo medido enquanto nenhum transistor estiver sendo medido, e seu valor pode ser usado para compensar o valor encontrado na medida de I_d do DUT.

Tabela 6.4: Comparação entre os métodos de medida de V_{th} para DUTs na Matriz e Isolados.

Método	Isolado	Matriz	Erro	Erro%
CC	0,242 V	0,252 V	0,01	4,13%
ELR	0,252 V	0,255 V	0,003	1,19%
SD	0,265 V	0,272 V	0,012	4,53%

6.3 Caracterizando BTI

O funcionamento de todo o circuito foi simulado com a ajuda do simulador apresentado em (KACZER, 2011), o qual se integra ao simulador HSPICE para simular a degradação nos transistores fazendo uso da teoria de *trapping/detrapping*. O simulador muda os tempos de captura e emissão das armadilhas de acordo com a tensão aplicada na porta V_g . Assim, as tensões limiaries dos transistores se degradam e recuperam durante a simulação transiente dependendo da tensão de entrada.

O circuito foi simulado de forma que apenas os DUTs possam ser estressados, tornando assim a simulação mais rápida. Foi escolhida para ser simulada a matriz que contém os transistores do tipo PMOS com 30 nm de comprimento de canal e 150 nm e 250 nm de largura.

Como já visto, existem dois tipos de medidas a serem realizadas no transistor: (1) primeiramente, devem ser medidas as curvas $I_d \times V_g$ encontrando as tensões limiaries dos transistores previamente ao estresse, utilizando-se um dos métodos descritos em 3.1, (2) após o primeiro período de estresse a medida de ΔV_{th} é efetuada utilizando apenas um ponto medido de I_d para um V_g próximo ao V_{th} encontrado, como visto em 3.2. A automatização desse processo é dada pela criação de *look-up tabels* com os valores de $I_d \times V_g$ encontrados em (1), então ΔV_{th} é encontrado achando o valor de V_g na tabela

correspondente ao I_d medido em (2) e assim a degradação da tensão limiar é dada por: $\Delta V_{th} = V_{g,meas} - V_{g,tabela}$.

Na simulação foi utilizado VDD igual a 1 V e V_{stress} estático de 0 V e V_d igual a 0.95 V. O *slope* dos sinais externos utilizado foi de 100 ns. O período de medida utilizado foi de 1.0005 ms. A medida da corrente foi feita tirando a média no período de 1ms, começando em 250 ns e terminando em 1.00025 ms, assim tendo 250 ns, no início e fim do período, como tempo de segurança deixado para o sinal estabilizar.

Por fim, as medidas de ΔV_{th} foram efetuadas após 0.1 s, 0.4 s, 1.6 s, 6.4 s, 25.6 s, 102.4 s, 409.6 s, 1638.4 s, 6553.6 s, 26214.4 s e 104857.6 s de estresse e os desvios das tensões limiars achados são mostrados no gráfico da Figura 6.1, para os primeiros 15 transistores.

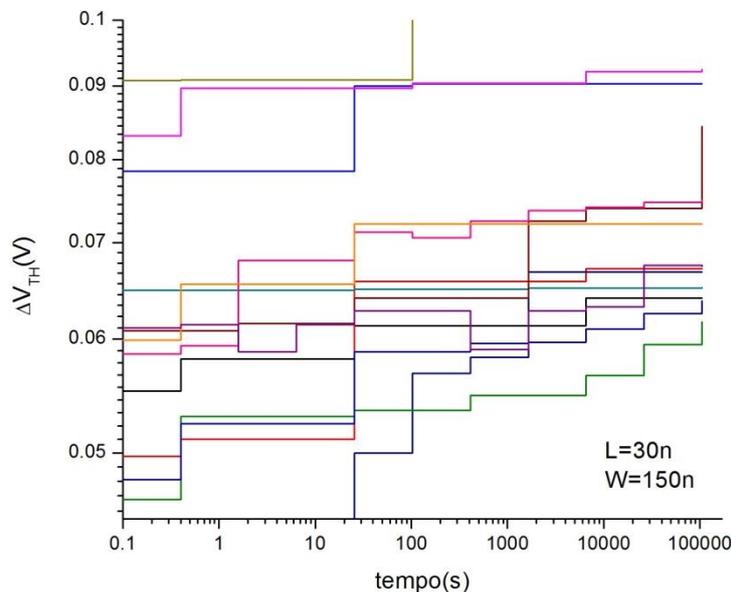


Figura 6.1 ΔV_{th} encontrado para os primeiros 15 transistores da matriz.

A leve recuperação encontrada em alguns transistores da Figura 6.1 é explicada pela recuperação que ocorre durante a medida onde $V_{g,meas}$ é menor que $V_{g,stress}$. Como I_d é medido como a média num período de 1 ms, é esperado que recuperações ocorram e contribuam para um erro na caracterização de BTI.

A Figura 6.2 mostra a corrente medida em V_{force} , para os primeiros transistores, durante a simulação do circuito, onde se pode ver a recuperação da corrente durante a medida. Os *glitches* na corrente representam o período de transição na troca de DUT a ser medido.

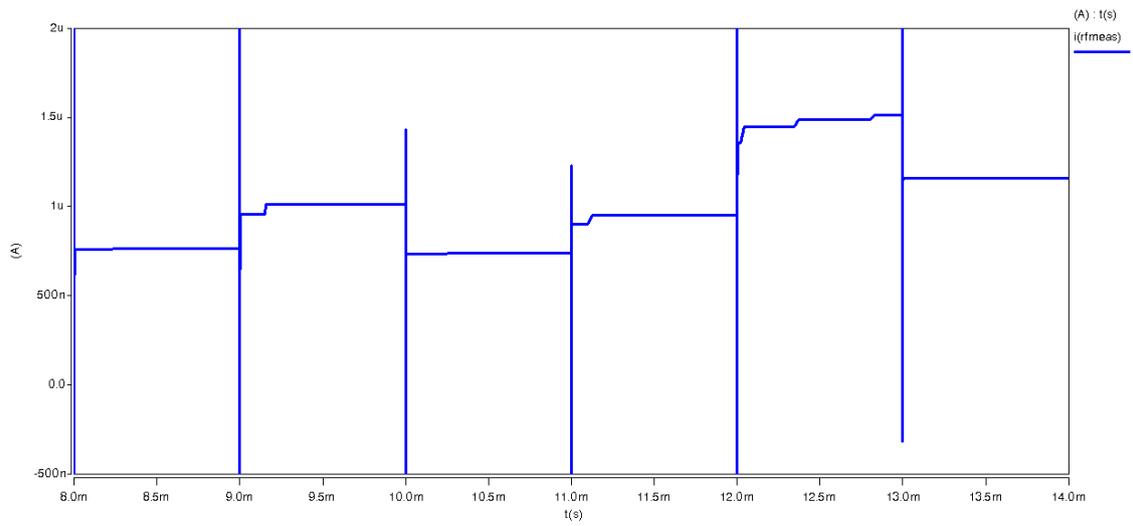


Figura 6.2 Corrente I_d medida em V_{force} .

7 CONCLUSÃO

Nesse trabalho foram descritos os mecanismos que originam o fenômeno chamado de *Bias Temperature Instability* (BTI), bem como o impacto deste fenômeno de degradação nos circuitos e a importância da correta caracterização do efeito, principalmente em transistores altamente miniaturizados, onde a grande variabilidade de BTI requer uma caracterização estatística adequada.

Também foram discutidos métodos de medidas e circuitos propostos na literatura para a caracterização de BTI, analisando-se suas vantagens e desvantagens. Assim, se chegou a conclusão que um novo circuito para a medida do efeito deveria ser projetado a fim de cumprir todos os requisitos propostos para uma adequada caracterização estatística de BTI em dispositivos altamente miniaturizados.

A fim de se atingir este objetivo, foi proposto um circuito *on-chip* que permite a caracterização paralela de uma grande amostra de transistores, permitindo o estudo dos efeitos estático e dinâmico de BTI, bem como seu comportamento ao longo do tempo de recuperação (relaxamento). O uso do método de estresse paralelo e medida sequencial, controlado por um circuito *on-chip*, permitiu diminuir drasticamente o tempo de caracterização e a área utilizada no chip, enquanto mantém boa precisão na medida. Foi realizado o layout final do circuito no nó tecnológico de 28nm do IMEC.

O layout feito contou com 12 matrizes de 90 transistores cada, distribuídas em 6 módulos, possibilitando a caracterização de BTI em 1080 dispositivos e possibilitando a caracterização 90 vezes mais rápida em uma área 40 vezes menor, comparado a caracterização de 1080 transistores individualmente. O circuito final com uma área de $30\mu\text{m} \times 76\mu\text{m}$ permite sua inclusão nas *drop-in/dice lanes*, possibilitando sua utilização na caracterização de processos de fabricação. Simulações realizadas com o circuito mostram que o tamanho da matriz pode ser aumentado, possibilitando a inclusão de mais transistores e ainda mantendo a precisão e performance do circuito.

REFERÊNCIAS

ASHRAF, N. et al, Accurate Model for the Threshold Voltage Fluctuation Estimation in 45-nm Channel Length MOSFET Devices in the Presence of Random Traps and Random Dopants, **Electron Device Letter**, v.32, n.8, p 1044 – 1046, Junho 2011.

CHEN, G. et al., Dynamic NBTI of PMOS Transistors and Its Impact on Device Lifetime, in: IRPS, 2003. **Proceedings...** [S.l.s.n], p. 169–202.

DA SILVA, M.B et al. NBTIaware technique for transistor sizing of high-performance CMOS gates, in: LATIN AMERICAN TEST WORKSHOP, 2009. **Proceedings...** [s.n], Buzios, RJ 2009. p. 1-5.

DEGRAEVE, R. et al., Review of reliability issues in high-k/metal gate stacks, in: SYMP. PHYSICAL AND FAILURE ANALYSIS OF INTEGRATED CIRCUITS, 2008. **Proceedings...** [S.l.s.n], p. 1–6.

DENAI, M. et al., On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's, in: IEDM, 2004. **Proceedings...** [S.l.s.n], p. 109–112.

GRASSER, T. et al., A Two-Stage Model for Negative Bias Temperature Instability, in: IRPS, 2009. **Proceedings...** [S.l.s.n], p. 33–4.

GRASSER, T. et al. The Time Dependent Defect Spectroscopy (TDDS) for the Characterization of the Bias Temperature Instability, in: IRPS, 2010. **Proceedings...** IEEE International, Anaheim, CA, 2010. p. 16-25.

GRASSER, T. et al, The Paradigm Shift in Understanding the Bias Temperature Instability: From Reaction–Diffusion to Switching Oxide Traps, **IEEE Electron Devices Society**, v.58, n.11, p 3652 – 3666, Setembro 2011

HUARD, V. New Characterization and Modeling Approach for NBTI Degradation from Transistor to Product Level, in: IEDM, 2007. **Proceedings...** IEEE International, Washington, DC, 2007. pp. 797–800

JEPPSON, K, SVENSSON, C. Negative bias stress of MOS devices at high electric fields and degradation of MOS devices. **Journal of Applied Physics**, v.48, 1977.

KACZER, B. et al., Ubiquitous relaxation in BTI stressing - new evaluation and insights, in: IRPS, 2008. **Proceedings...** IEEE International, Phoenix, AZ, 2008. p. 20–27.

KACZER, B. et al. NBTI from the Perspective of Defect States with Widely Distributed Time Scales, in: IRPS, 2009. **Proceedings...** IEEE International, Montreal, QC, 2009. p. 55-60.

KACZER, B. et al. Origin of NBTI Variability in Deeply Scaled pFETs, in: IRPS, 2010. **Proceedings...** IEEE International, Anaheim, CA, 2010. p. 26 - 32.

KEANE, J. et al. An on-chip NBTI sensor for measuring pMOS threshold voltage degradation, **IEEE Transactions on VLSI**, v.18, n.6, p. 947–956, Julho 2010.

KEANE, J. et al., An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB, **Journal of Solid-State Circuits**, v.45, n.4, p. 817–829, Abril 2010.

KEANE, J.; ZHANG, W.; KIM, H. An on-chip monitor for statistically significant circuit aging characterization, in: IEDM. 2010. **Proceedings...** IEEE International, San Francisco, CA, 2010. p. 1-4

KETCHEN, M. B.; BHUSHAN, M.; BOLAM, R. Ring oscillator based test structure for NBTI analysis, in: ICMTS, 2007. **Proceedings...** IEEE International, Tokyo, 2007. p. 42-47.

KIM, T. et al., Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits, in: VLSI Circuits, 2007. **Proceedings...** IEEE, Kyoto, 2007. p. 122-123.

KIM, J. et al., Ring oscillator circuit structures for measurement of isolated NBTI/PBTI effects, in: ICICDT, 2008. **Proceedings...** IEEE International, Austin, TX, 2008. p. 163-166.

MIURA, Y.; MATUKURA, Y. Investigation of Silicon-Silicon Dioxide Interface Using MOS Structure, **Japanese Journal of Applied Physics**, v.5, p. 180, 1966.

ORTIZ-CONDE, A. A review of recent MOSFET threshold voltage extraction methods, **Microelectronics Reliability**, v.42, n.4, p. 583-596, Abril 2002.

REDDY, V. et al., Impact of negative bias temperature instability on digital circuit reliability, in: IRPS, 2002. **Proceedings...** [S.l.s.n], p. 248–254.

SHEN, C. et al., Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with SiON dielectrics, in: IEDM, 2006. **Proceedings...** [s.n], San Francisco, CA, 2006 p. 1–4.

SCHLÜNDER, C. et al. A new smart device array structure for statistical investigations of BTI degradation and recovery, in: IRPS, 2011. **Proceedings...** IEEE International, Monterey, CA, 2011. p. 56-60.

RESINGER, H. et al., Analysis of NBTI degradation and recovery behavior based on ultra fast V_t -measurement, in: IRPS, 2006. **Proceedings...** IEEE International, San Jose, CA, 2006. p. 448–453.

REISINGER, H. A Comparison of Fast Methods for Measuring NBTI Degradation, **IEEE Transaction on Device and Materials Reliability**, v.7, n.4, p. 531-539, Dezembro 2007.

SCHRODER, D.K. et al., Negative bias temperature instability: What do we understand?, **Microelectronics Reliability**, v.47, n.6, p. 841-852, Junho 2007.

VATTIKONDA, R et al. Modeling and minimization of PMOS NBTI effect for robust nanometer design, in: DAC, 2006. **Proceedings...** [S.l.s.n], p. 1047-1052.

WIRTH, G.I et al, **Statistical Model for MOSFET Bias Temperature Instability Component due to Charge Trapping**, IEEE Electron Devices Society, v58, n8, p. 2743-2751, Junho 2011.

WANG, W. et al. The Impact of NBTI on the Performance of Combinational and Sequential, in: DAC, 2007. **Proceedings...** ACM, San Diego, CA, 2007. p. 364-369.

ZAFAR, S. et al. A Comparative Study of NBTI and PBTI (Charge Trapping) in SiO₂/HfO₂ Stacks with FUSI, TiN, Re Gates, in: VLSIT, 2006. **Proceedings...** [s.n], Honolulu, HI, 2006. p. 23-25.

ZHAO, W.; CAO, Y. New generation of Predictive Technology Model for sub-45nm early design exploration, **IEEE Transactions on Electron Devices**, v.53, n.11, p. 2816-2823, Novembro 2006.