



INAOE

Circuitos robustos a envejecimiento de NBTI con voltaje dual de alimentación

por:

**Ing. Freddy Alexander Forero
Ramirez**

Tesis sometida como requisito parcial para obtener el grado de:

**MAESTRO EN CIENCIAS EN LA
ESPECIALIDAD DE ELECTRÓNICA**

en el

Instituto Nacional de Astrofísica, Óptica y Electrónica

Noviembre 2015

Tonantzintla, Puebla

Director de tesis:

Dr. Víctor Hugo Champac Vilela, INAOE

©INAOE 2015

Derechos reservados

El autor otorga al INAOE el permiso de reproducir y distribuir
copias de esta tesis en su totalidad o en partes.



Resumen

La continua reducción de los dispositivos MOSFET ha permitido que los circuitos integrados VLSI alcancen una mayor densidad de integración y velocidad de operación, lo cual ha llevado a un incremento en la demanda de circuitos electrónicos. Sin embargo, en el régimen de dimensiones nanométricas el impacto de algunos mecanismos que degradan el rendimiento del dispositivo es más significativo, disminuyendo su tiempo de vida útil. Los principales mecanismos de envejecimiento que afectan el rendimiento son NBTI, HCI y TDDBD, siendo NBTI identificado como el factor predominante de envejecimiento en transistores PMOS. El efecto que tiene NBTI es el de aumentar la magnitud del voltaje umbral del transistor a través del tiempo. El incremento del voltaje umbral ocasiona que el retardo de propagación de los caminos lógicos de señal en un circuito digital se incremente. Si este retardo sobrepasa las restricciones de temporización del sistema se presentará una falla que en algunas aplicaciones como en el área automotriz, aeroespacial o médica son inaceptables. Para tolerar el incremento del retardo de propagación debido al envejecimiento de los dispositivos, convencionalmente se introduce un tiempo adicional en el período de reloj que funciona como margen de seguridad (también llamado *banda de guarda*). Dependiendo de las condiciones de operación del dispositivo como temperatura y voltaje de alimentación el efecto del NBTI se hace mayor o menor, siendo la temperatura el factor más importante en el incremento del efecto del NBTI.

La temperatura de un dispositivo es proporcional a la potencia disipada, ésta

a su vez depende del valor del voltaje de alimentación del circuito; por lo tanto al bajar el voltaje de alimentación no solo favorece la reducción consumo de potencia sino también la disminución del impacto del NBTI. Uno de los inconvenientes de bajar el voltaje de operación del circuito es que hace que el rendimiento del circuito baje al volverlo mas lento. Sin embargo, si se baja el voltaje selectivamente en determinadas compuertas, se podría diseñar un circuito digital con poco impacto en el retardo de los caminos lógicos críticos y un menor consumo de potencia, lo cual se traduce en un doble beneficio. El primero como se comento anteriormente disminuir el impacto del NBTI, el segundo es que al reducir el consumo de potencia se beneficia el tiempo de duración de la fuente de energía en sistemas portátiles o móviles.

En esta tesis de maestría, se presenta una metodología para la selección del conjunto de compuertas que trabajarían a un voltaje inferior al nominal considerando efectos de envejecimiento por NBTI. Además, esta metodología tiene en cuenta los efectos causados cuando una compuerta trabaje a un voltaje inferior y cambios de temperatura, esto se realiza penalizando el retardo total del circuito pero beneficiando el consumo de potencia y la degradación del retardo por NBTI.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT)

Por el apoyo económico otorgado a través de la beca para estudios de maestría.

(No. de Registro: 420129)

Al Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)

por abrirme las puertas para realizar los estudios de posgrado.

A mi asesor: Dr. Victor Hugo Champac Vilela

por el conocimiento y formación académica que he recibido de su parte.

A los Doctores:

Dr. Mónico Linares Aranda

Dr. Esteban Tlelo Cuautle

Dr. Alejandro Díaz Mendez

Por haberme apoyado con la revisión de esta tesis y por retroalimentarme con sus puntos de vista y opiniones.

A los trabajadores del INAOE:

Por el soporte y ayuda que prestan a los estudiantes.

*A Mar, mis papas, mis hermanos y mis amigos
Por su apoyo incondicional
en todo este tiempo*

Contenido

Resumen	I
Agradecimientos	III
Lista de Acrónimos	VIII
1. Introducción	1
1.1. Confiabilidad en circuitos integrados digitales	2
1.1.1. Confiabilidad vs fallas tempranas (Yield)	3
1.1.2. Curva de tina de baño e impacto del escalado tecnológico .	3
1.2. Mecanismos que afectan la confiabilidad	6
1.2.1. <i>Negative Bias Temperature Instability</i> (NBTI)	6
1.2.2. <i>Hot Carriers Injection</i> (HCI)	7
1.2.3. <i>Time Dependent Dielectric Break-down</i> (TDDBD)	8
1.3. Banda de guarda	8
1.4. Justificación de la tesis	10
1.5. Organización de la tesis	11
2. Envejecimiento debido a NBTI y su impacto en tecnologías na- nométricas	12
2.1. Fundamentos	13
2.2. Modelo predictivo de degradación del V_{th} debido a NBTI	19
2.3. Impacto en el rendimiento de un circuito digital	24

3. Técnicas para reducir el impacto del envejecimiento en circuitos integrados digitales	29
3.1. Incremento de la movilidad efectiva de los huecos	30
3.2. Control del vector de entrada	32
3.3. Control de Nodos Internos	33
3.4. Redimensionamiento de las compuertas	34
3.5. Ordenar los nodos de entrada en las compuertas	35
3.6. Circuitos Adaptivos	37
3.6.1. Camino lógico réplica y sensores embebidos	39
3.6.2. Predicción de error	40
3.6.3. Detección y corrección de error	42
3.7. Disminución del voltaje de alimentación	44
4. Estimación de Retardo, Consumo de Potencia y Temperatura del circuito considerando efectos de NBTI	46
4.1. Modelo de retardo de compuerta	47
4.2. Calculo de potencia	50
4.2.1. Potencia Dinámica debida a Conmutación en Capacitancias	50
4.2.2. Potencia de Corto circuito	54
4.2.3. Potencia debido a corrientes de fuga	54
4.2.4. Potencia debido a Pulsos Espurios	56
4.3. Cálculo de Temperatura	57
4.4. Algoritmo iterativo para cálculo de retardo, potencia, temperatura e impacto de NBTI	60
5. Metodología de robustez a envejecimiento basado en Dual-V_{DD}	64
5.1. Consideraciones al utilizar Dual- V_{DD}	66
5.1.1. Impacto en el Retardo Total	66
5.1.2. Impacto en el Layout	68
5.1.3. Impacto en el consumo de potencia estática	71

5.1.4. Selección de Valor de Voltaje de Alimentación Bajo	74
5.2. Metodología de selección de compuertas	75
5.2.1. Identificación de caminos lógicos topológicos y cálculo de su retardo medio de propagación	77
5.2.2. Estimación de Retardo y su degradación de los caminos lógi- cos con $V_{DD} =$ Voltaje Nominal	78
5.2.3. Prefiltrado de caminos lógicos	79
5.2.4. Cambiar $V_{DD}=V_{DDL}$ a la n-esima compuerta y estimación de retardos	80
5.2.5. Métrica de Selección de Compuertas	80
5.2.6. Validación de la Métrica	82
5.2.7. Obtención del conjunto de compuertas con la métrica más alta	84
5.3. Resultados en circuitos ISCAS	85
5.3.1. Degradación de Retardo	86
5.3.2. Consumo de Potencia	88
6. Conclusiones	90
Lista de figuras	95
Lista de tablas	96
Bibliografía	97

Lista de Acrónimos

NBTI *Negative Bias Temperature Instability*

HCI *Hot Carriers Injection*

TDDBD *Time Dependent Dielectric Break-down*

PDF *Probability Density Function*

LTDM *Long Term Degradation Model*

RDF *Random Dopant Fluctuation*

AF *Aging Factor*

LCP *Longet Critical Path*

NCP *Near Critical Paths*

SSTA *Statistical Static Timing Analysis*

STA *Static Timing Analysis*

Capítulo 1

Introducción

Durante los últimos años, el desarrollo de productos más rápidos y complejos en el campo de la electrónica se ha basado en el escalado de las dimensiones de los dispositivos, alcanzando actualmente el rango nanométrico y sub–nanométrico tal como lo indica la tendencia observada por “la ley de Moore” [1].

En este momento la mayoría de dispositivos electrónicos forman parte de un sistema en movimiento como carros, aviones, satélites y algunas tendencias apuntan a sistemas que la personas visten o llevan puestos denominados sables o en ingles *Wearables* , una gran cantidad de estos van dirigidos hacia la salud. Por lo tanto, deben ser dispositivos electrónicos pequeños y de bajo consumo de potencia para que el tiempo de duración de su fuente de alimentación sea lo largo. Además, una posible falla en los ejemplos mencionados anteriormente puede tener un alto costo tanto económico como humano, es por eso que estos dispositivos deben ser también robustos y tener un tiempo de vida útil largo.

Los circuitos digitales nanométricos, debido a su tamaño disipan potencia en un área cada vez mas pequeña, lo que lleva a operar a temperaturas altas en las cuales la degradación de los dispositivos con el tiempo es más propensa y tiene un gran impacto en el rendimiento esperado, limitando así el tiempo de vida útil de un circuito. Considerar e investigar estos efectos permite desarrollar nuevas técnicas

de diseño para asegurar el correcto funcionamiento de los circuitos digitales de alta complejidad manteniendo un rendimiento y tiempo de vida especificados, bajo ciertas condiciones físicas, tecnológicas y económicas.

1.1. Confiabilidad en circuitos integrados digitales

Es un hecho que los circuitos y sistemas electrónicos tienen un tiempo de vida finito antes de que una falla crítica en el sistema ocurra. Sin embargo, si este tipo de fallas se produce una vez superado el tiempo de vida esperado por el usuario, éste estará satisfecho y se habrán cumplido las expectativas planteadas al producto [3].

La confiabilidad (en inglés *Reliability*) es la habilidad de los dispositivos electrónicos de realizar las funciones para las que fueron diseñados cumpliendo con las especificaciones requeridas, bajo las condiciones de operación presentes, como voltaje de polarización y frecuencia de operación, durante un período de tiempo específico [4]. El propósito del estudio de la confiabilidad de los dispositivos es el de buscar alternativas que permitan asegurar que el tiempo de vida del sistema será mayor que el tiempo de uso requerido.

En el mundo de los dispositivos CMOS, el criterio de operación para algunos productos puede ser muy diferente. Algunos son utilizados raramente y por tanto requieren muy poco tiempo de vida útil. Por otra parte existen aplicaciones donde garantizar la confiabilidad es de vital importancia y a la vez es requerido un alto rendimiento por parte de los dispositivos electrónicos. Por ejemplo, en el campo de la industria automotriz o espacial, donde la electrónica digital es cada vez más común y los circuitos se encuentran operando constantemente bajo condiciones extremas, se espera un correcto funcionamiento libre de fallas durante todo el tiempo de vida útil, por ejemplo 10 años.

1.1.1. Confiabilidad vs fallas tempranas (Yield)

Confiabilidad y fallas tempranas (*Yield*) son quizás los aspectos más importantes para el desarrollo de una nueva tecnología [5]. El yield está definido como la probabilidad de falla de un dispositivo recién fabricado ($t = 0$), mientras que la confiabilidad puede ser vista como una falla funcional del dispositivo durante su tiempo de operación ($t > 0$). Un proceso de fabricación con bajo yield es inaceptable para comenzar a trabajar, pero aún un proceso con un muy alto yield (baja probabilidad de obtener defectos iniciales) y baja confiabilidad no es sostenible económicamente a largo plazo.

En la industria, la confiabilidad de un producto debe ser conocida antes de que éste sea vendido, de esta forma los costos de garantía pueden ser medidos y la satisfacción del consumidor asegurada. Cumplir con estos objetivos, implica calcular el impacto que tendrá cada mecanismo de falla y predecir el tiempo en el cual su impacto comienza a ser significativo.

1.1.2. Curva de tina de baño e impacto del escalado tecnológico

Los errores o fallas que pueden ocurrir en un circuito digital, pueden clasificarse en cuatro tipos diferentes, dependiendo del tiempo determinado en el que ocurren. La figura 1.1 es conocida como *curva de tina de baño*. Esta curva es una medición del número de fallas por unidad de tiempo (tasa de fallas) versus el tiempo de operación y es una estimación del comportamiento de la confiabilidad de un sistema electrónico fabricado.

- Fallas tempranas (*yield*): Son defectos que se deben principalmente a las variaciones o imperfecciones en el proceso de manufactura que incluso en el tiempo cero de operación pueden ocasionar errores por mal funcionamiento del circuito. La gran mayoría de este tipo de defectos son detectados en la

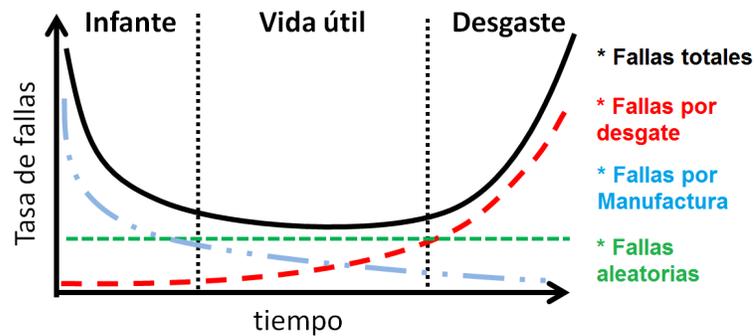


Figura 1.1: Tasa de Fallas representada en la Curva de tina de baño. [3]

etapa de pruebas o *testing* realizado a los circuitos.

- Fallas infantiles: Son aquellos defectos que no fueron detectados en la etapa de pruebas pero que el mal funcionamiento se presenta rápidamente. Comúnmente un análisis de los defectos en edad infante revela estructuras extremadamente débiles, que de alguna forma no fallaron en la pruebas iniciales de fallas tempranas. A medida que el proceso de fabricación madura para una nueva tecnología, la cantidad de estos defectos debe disminuir.

Una técnica para detectar este tipo de fallas es conocida como *burn-in* y consiste en aplicar condiciones de operación (como temperatura o voltaje) superiores a las nominales por un período de tiempo corto de manera que estos circuitos defectuosos se hagan evidentes y sean descartados. De esta forma, el consumidor recibirá sólo dispositivos que se encuentren en el tiempo útil de operación [3].

- Fallas durante el tiempo útil de operación: En esta etapa de operación, la número de fallas presentes es baja y deben estar contenidas dentro de las especificaciones establecidas para asegurar la viabilidad de la garantía establecida. Los mecanismos de falla presentes en esta etapa son de carácter aleatorio.
- Fallas por envejecimiento: A medida que pasa el tiempo, mayor cantidad

de circuitos comienzan a ser defectuosos, estas fallas se deben al desgaste, degradación o envejecimiento de los dispositivos, es decir, a su confiabilidad. El momento en el que este tipo de fallas comienza a hacerse presente depende de la tecnología propia y las condiciones de operación como temperatura, el voltaje de alimentación o la frecuencia de uso del dispositivo, entre otros factores. El objetivo de estudiar los mecanismos que afectan la confiabilidad es el de asegurar que las fallas por envejecimiento no estarán presentes hasta después de cumplir el tiempo de vida útil esperado. En la figura 1.2 se muestra un ejemplo donde luego de cierto tiempo de uso, debido a la degradación se alcanza el umbral máximo permitido de alguna especificación, pero este tiempo de uso es menor al establecido como garantía, lo cual representa una pérdida económica.

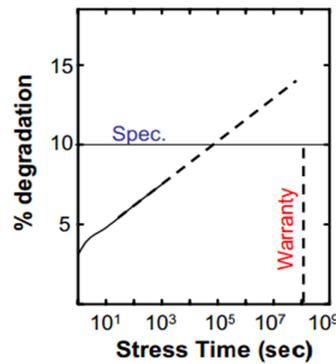


Figura 1.2: Degradación del rendimiento e incumplimiento del tiempo de vida requerido. [5]

A medida que se alcanzan los límites físicos y tecnológicos de las dimensiones en los dispositivos, tanto los defectos de manufactura como los fenómenos de degradación tienden a aumentar el número de fallas observadas y reducir la confiabilidad. Esto se observa en la tendencia marcada por las flechas en la figura 1.3.

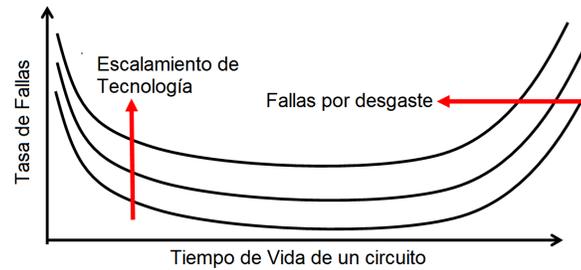


Figura 1.3: Tendencias en la curva de tina de baño

1.2. Mecanismos que afectan la confiabilidad

Existen diferentes fenómenos físicos que afectan la confiabilidad de los dispositivos, ya sea debido a los materiales del substrato y el óxido, así como las interfaces entre ellos. Los principales fenómenos de degradación que afectan la confiabilidad son NBTI, HCI y TDDBD [5]. A continuación se mencionarán las principales características de cada uno de ellos.

1.2.1. NBTI

Inestabilidad debido a voltaje de polarización negativo y Temperatura *NBTI*, ha sido identificado como un fenómeno dominante de envejecimiento en tecnologías CMOS nanométricas, el cual afecta principalmente a los transistores PMOS. Los enlaces de hidrógeno y silicio en la interfaz, pueden ser fácilmente rotos debido al campo eléctrico perpendicular y el aumento de la temperatura, generando trampas o estados en la interface. Esto trae como consecuencia un incremento del voltaje umbral del dispositivo a lo largo del tiempo [10]. La condición de polarización propicia para este efecto es cuando la compuerta tiene un voltaje negativo respecto al drenaje y fuente.

En este trabajo, se hará énfasis en el fenómeno de NBTI y será explicado con mayor detalle más adelante.

1.2.2. HCI

Inyección de Portadores Calientes *HCI* es un fenómeno donde algunos portadores, que fluyen cerca de la región de estrangulamiento del canal, ganan energía significativa debido al alto campo eléctrico horizontal en esta zona.

En la región lineal, V_{ds} está distribuido a través del canal entero y el campo eléctrico está limitado a valores menores al campo crítico de saturación de la velocidad, V_{dsat} . A medida que V_{ds} se incrementa entrando en la región de saturación, el exceso de voltaje $V_{ds} - V_{dsat}$ se distribuye en la región de estrangulamiento del canal cerca del drenador. En este punto, el campo eléctrico tiene una dependencia exponencial con la posición [3].

Como se muestra en la figura 1.4, estos portadores energéticos pueden perder su energía mediante impactos de ionización generando un par electrón–hueco y contribuyendo a la corriente de sustrato. También pueden tener suficiente energía como para romper los enlaces entre silicio e hidrógeno en la interface, generando trampas en la interface. Igualmente pueden superar la barrera de potencial en la interface e inyectarse en el óxido de compuerta donde pueden ser atrapados o pueden generar defectos en el óxido.

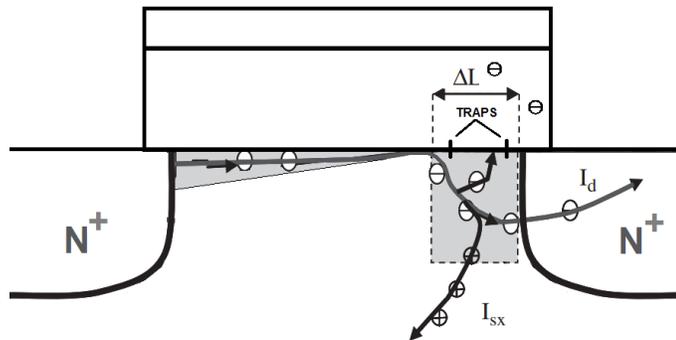


Figura 1.4: Mecanismos de Hot Carriers

1.2.3. TDDBD

El rompimiento del óxido de compuerta con el paso del tiempo se debe a que a éste es aplicado constantemente un campo eléctrico. Debido a efectos de tuneleo algunos huecos pueden quedar atrapados en el dieléctrico, mientras más carga es atrapada es posible formar un camino por el cual puede fluir corriente como se muestra en la figura 1.5. Entonces, la compuerta y el canal estarían eléctricamente unidas, dañando completamente el concepto de funcionamiento.

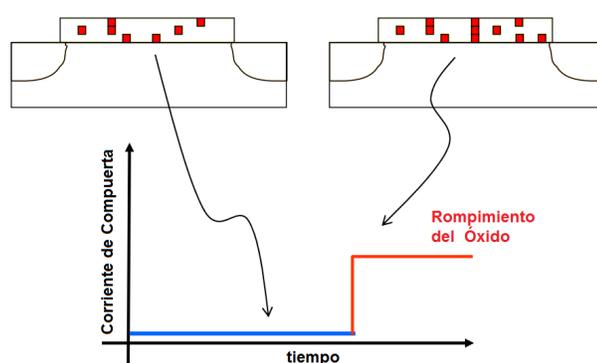


Figura 1.5: TDDBD

1.3. Banda de guarda

La mayoría de los circuitos digitales, se basan en la estructura mostrada en la figura 1.6. Un bloque con lógica combinacional recibe una o varias señales de entrada, realiza su operación en base a estas entradas y entrega el resultado a la siguiente etapa de lógica combinacional. El resultado de cada bloque es almacenado por Flip–Flops, los cuales mantienen el estado almacenado en un ciclo anterior hasta que el reloj indique el momento de almacenar la operación resultante. El tiempo en que el bloque combinacional tarda en recibir la señal de entrada, procesarla y entregar la función resultante al FF correspondiente determina el período de reloj necesario.

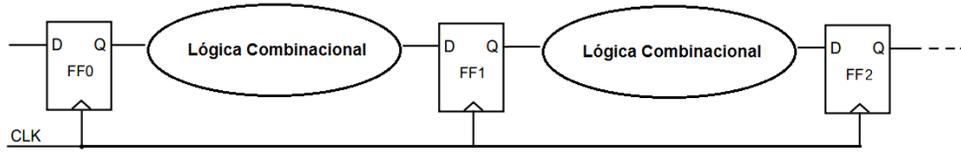


Figura 1.6: Estructura lógica de un circuito digital

Sin embargo, el período de reloj necesario depende de la variabilidad del retardo de propagación de la lógica combinacional, que puede ser de diferente tipo: Permanentes como las variaciones de proceso, de degradación en el tiempo debido a los mecanismos de envejecimiento mencionados anteriormente o transitoria que son efectos temporales como caídas en alimentación, ruido por acoples capacitivos, radiación etc. Esta variabilidad impide estimar de forma determinista el retardo de propagación del bloque de lógica combinacional y el período de reloj necesario. Para asegurar que un período de reloj seleccionado, proporcione el suficiente tiempo de propagación a los caminos lógicos del circuito es necesario dejar un lapso de tiempo extra al período nominal, para que aún bajo el peor de los casos de estos efectos de variación de retardo, el circuito pueda operar correctamente. A este tiempo extra se le conoce como margen de seguridad o banda de guarda. La figura 1.7 ilustra este concepto, donde se ha dejado una banda de guarda (zona limitada por líneas verdes) exclusiva para tolerar el incremento del retardo del camino lógico crítico de un bloque de lógica combinacional debido al envejecimiento.

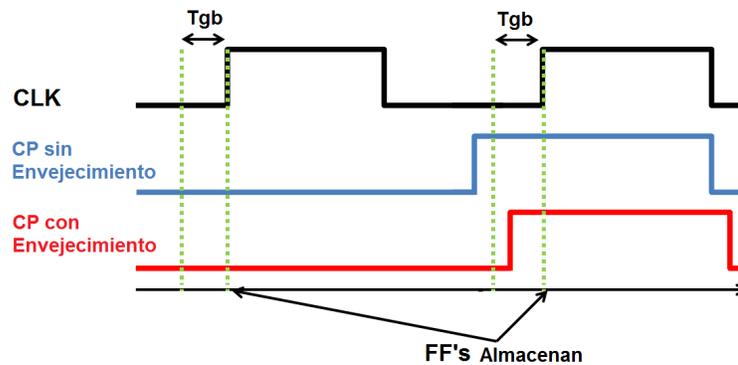


Figura 1.7: Margen de Seguridad para garantizar la correcta operación del circuito

1.4. Justificación de la tesis

Actualmente la tendencia de los circuitos va encaminada a aplicaciones de bajo consumo de potencia debido a que la mayoría de las aplicaciones son o se necesitan ser portátiles. Además algunas de estas aplicaciones van encaminadas a sectores como la salud y automotriz donde una falla puede tener un alto costo. Este es un nuevo paradigma en donde el circuito no solo tiene que ser eficiente sino también presentar un correcto funcionamiento de varios años. De igual forma es bien sabido que las aplicaciones de baja potencia no solo pretenden ahorrar energía sino también que sean aplicaciones que no necesitan medios de refrigeramiento complejos y puedan ser usados por usuarios cerca de sus cuerpos sin que les genere un compromiso físico debido al calor del circuito.

Tomando las dos ideas anteriores y contextualizando hacia la generación de circuitos robustos a NBTI, se aplica la técnica Dual-VDD para bajar la degradación del retardo de los caminos lógicos de un circuito y también bajar el consumo de potencia del circuito el cual está fuertemente relacionado con la temperatura de operación. Al bajar la temperatura de operación del circuito el impacto del NBTI disminuye de manera sustancial. Por lo cual resulta que el estudio de esta técnica a profundidad resulta útil e importante para generar circuitos tolerantes a envejecimiento.

Sin embargo la técnica Dual-VDD presenta varios inconvenientes uno de ellos es utilizar convertidores de nivel los cuales incrementan el consumo de área de silicio haciendo al circuito más caro, por lo cual aquí también se estudia que impactos tiene no utilizar estos conversores de nivel. Otro inconveniente es la selección apropiada de compuertas que trabajaran con un voltaje de alimentación bajo. Con base en lo anterior se propone una metodología que tiene en cuenta el efecto de temperatura, disipación de potencia, y forma de patrones geométricos (*Layout*) para hacer esta elección de tal forma que el circuito presente la mayor robustez frente al envejecimiento debido a NBTI

1.5. Organización de la tesis

En esta introducción se ha dado una primera perspectiva de la importancia de la confiabilidad en los circuitos integrados digitales, los principales fenómenos que afectan la confiabilidad y el impacto de introducir márgenes de seguridad para tolerar este tipo de variaciones en el retardo de propagación de los caminos críticos de un circuito. En el capítulo 2 se hará un mayor enfoque en el mecanismo de envejecimiento analizado en esta tesis, el cual es el NBTI. Se revisará la física de este fenómeno, el modelo matemático que lo describe y su impacto sobre el rendimiento de un circuito digital. En el capítulo 3 se hace una revisión de las principales técnicas para reducir el impacto del envejecimiento por NBTI, a nivel adaptativo como consideraciones de diseño. En el capítulo 4, se plantean las bases para realizar un análisis estático de tiempos considerando el efecto del NBTI. En este capítulo se presenta el modelo lineal de compuerta, a partir del cual se desarrolla la metodología para el cálculo del valor retardo de un camino lógico, también se explica los diferentes consumos de potencia que tiene un circuito y el modelo térmico utilizado para calcular su temperatura de funcionamiento, por último se describe el algoritmo utilizado para calcular la temperatura del circuito a partir de la potencia que disipa. En el capítulo 5, se explicará la propuesta de una metodología de selección de compuertas que trabajaran con un voltaje bajo, y los determinados impactos de utilizar este voltaje bajo. En este capítulo también se incluyen los resultados obtenidos. Finalmente, en el capítulo 6 se exponen las conclusiones del trabajo de tesis.

Capítulo 2

Envejecimiento debido a NBTI y su impacto en tecnologías nanométricas

En los transistores PMOS ha sido identificado que su voltaje umbral cambia gradualmente a través del tiempo. Este cambio es ocasionado por el estrés que se genera en el óxido de compuerta debido al voltaje aplicado y su efecto se intensifica con la temperatura. La figura 2.1 muestra la condición de polarización donde el impacto de éste fenómeno es significativo, a esta condición se le llama *condición de stress*. La compuerta tiene una tensión negativa respecto al drenaje y fuente de un transistor PMOS simultáneamente. Este efecto es conocido como *Negative Bias Temperature Instability* o NBTI.

En el diseño VLSI, una predicción precisa de la degradación del rendimiento debido a NBTI es muy importante, ya que permite a los diseñadores determinar posibles soluciones para garantizar el tiempo de vida útil especificado.

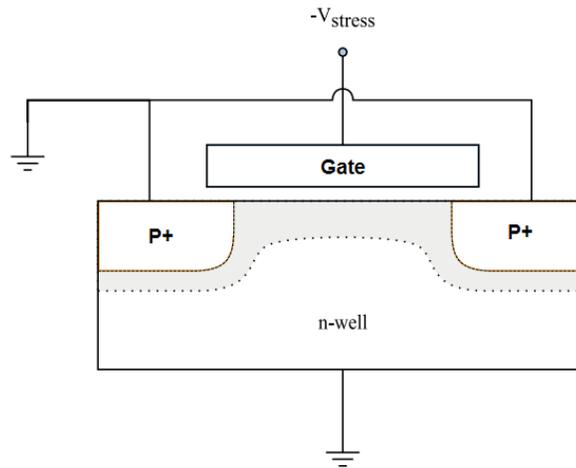


Figura 2.1: Condición de polarización donde el efecto de NBTI se presenta

2.1. Fundamentos

El fenómeno de NBTI fue observado por primera vez en los años 60, donde los transistores PMOS eran los dispositivos dominantes en los procesos de fabricación [5]. El efecto se genera debido a los enlaces incompletos (*dangling bonds*) en la interface Si/SiO_2 que actúan como estados o trampas de huecos en la zona prohibida.

El átomo de silicio posee cuatro electrones de valencia y por tanto requiere también cuatro enlaces para completar su banda de valencia. Dentro de la estructura cristalina, cada átomo de silicio completa su banda de valencia estableciendo estos cuatro enlaces con sus átomos vecinos. Sin embargo, en la superficie del cristal de silicio no todos los enlaces pueden completarse como se muestra en la figura 2.2(a).

Luego del proceso de oxidación, gran parte de estos enlaces son completados con los átomos de oxígeno (figura 2.2(b)). Sin embargo, aún varios enlaces han quedado incompletos, para lo cual se realiza un proceso llamado *pasivación* con átomos de hidrógeno, formando enlaces $Si-H$ como se muestra en la figura 2.2(c), mejorando de esta forma la calidad en la interface del dispositivo. Sin embargo,

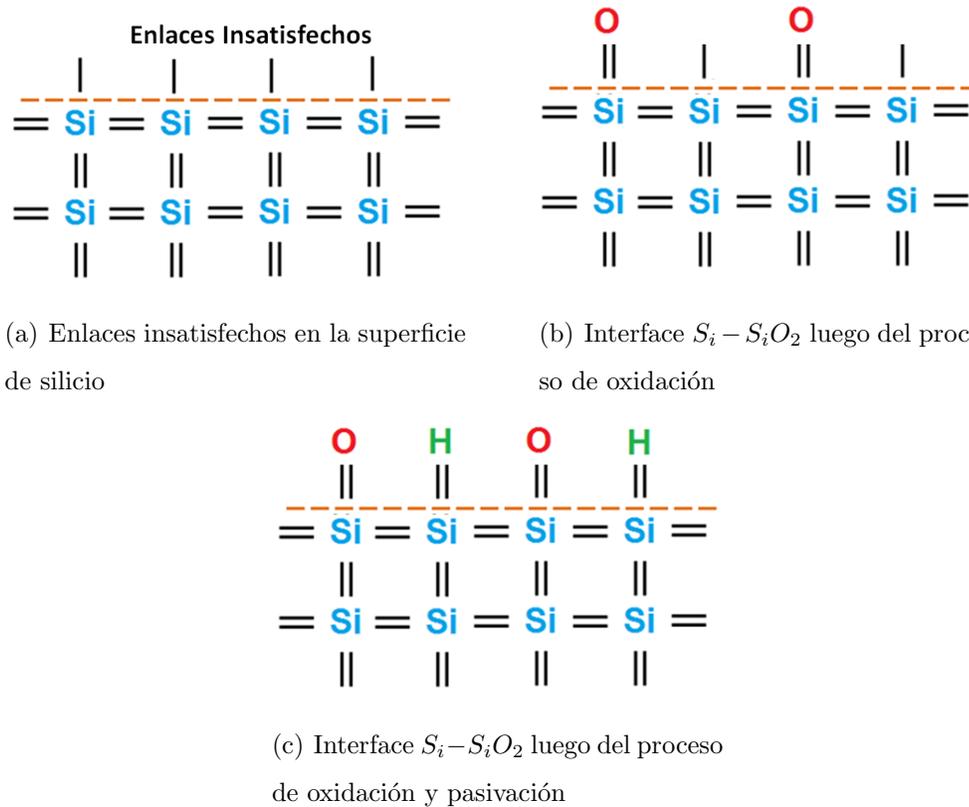


Figura 2.2: Interface del silicio y óxido de compuerta

son precisamente estos enlaces de hidrógeno y silicio los que causan el efecto del NBTI, ya que debido a su baja energía de acople pueden ser rotos fácilmente por los elevados campos eléctricos y la alta temperatura, reactivando las trampas en la interface [11].

Aunque el mecanismo exacto de generación de trampas en la interface es aún discutido, en [7] sugieren que algunos huecos en la capa de inversión llegan al óxido por mecanismos de tuneo, son atrapados e interactúan con los enlaces $S_i - H$ que tienen baja energía de acople (enlaces débiles) de manera que por el alto campo eléctrico y temperatura pueden ser rotos fácilmente (ver figura 2.3).

Al romperse los enlaces $S_i - H$, los átomos de hidrógeno se difunden de la interface hacia el óxido de compuerta, dejando un enlace incompleto que actúa como trampa de huecos (figura 2.4). Una característica importante de la degra-

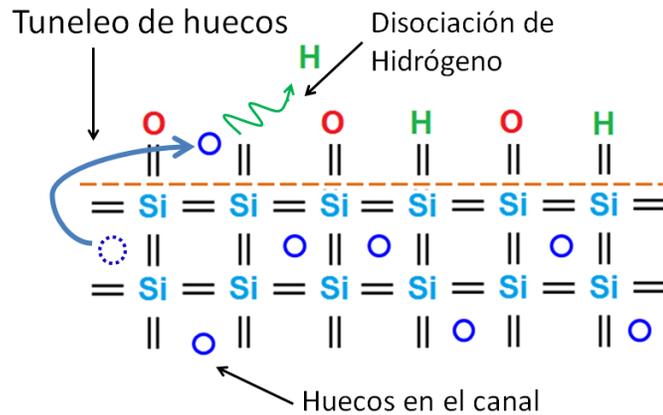


Figura 2.3: Tuneleo de huecos y rompimiento de enlaces $S_i - H$ en la interface (Reacción).

dación causada por NBTI, es la capacidad de recuperarse parcialmente cuando las condiciones de estrés (Campo eléctrico, temperatura, etc.) son disminuidas, ya que algunos enlaces $S_i - H$ son restablecidos como se ilustra en la figura 2.5.

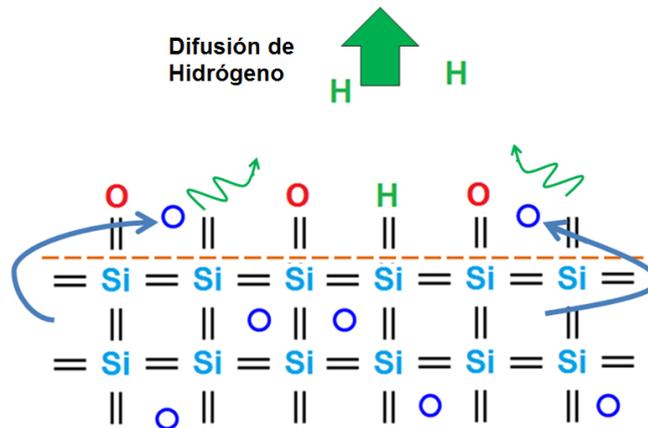


Figura 2.4: Difusión de los átomos de hidrógeno en el óxido (Difusión).

La figura 2.6 muestra la dependencia de la generación de trampas en la interface durante el tiempo de estrés para varios valores de temperatura y campo eléctrico. Como se puede observar, el cambio en las trampas en la interface incrementa con ambos parámetros y muestran una dependencia en forma de ley de potencia con el tiempo (t^n). En la relación con la temperatura, es evidente que

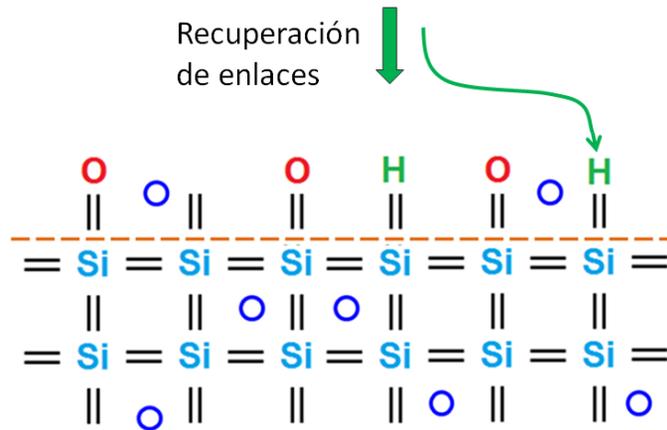


Figura 2.5: Pasivación de enlaces con átomos de hidrógeno que se recuperan (recuperación) .

la tendencia cambia luego de cierto punto en el tiempo (t_{break}) [7] a partir del cual la pendiente es mayor y que esta pendiente a su vez depende de la temperatura. Sin embargo, para el caso del campo eléctrico, tanto la pendiente como el punto de cambio de la tendencia son independientes del campo aplicado. El tiempo t_{break} está determinado por el momento en el cual el frente de difusión de Hidrógeno alcanza la interface entre el óxido de compuerta y el poly, ya que en éste la constante de difusión es mayor.

La generación de trampas en la interface en un transistor PMOS contribuyen con carga positiva a la banda de valencia, teniendo como consecuencia una desviación negativa en el voltaje umbral del transistor (se hace más negativo). La relación entre la densidad de trampas en la interface del silicio y óxido (N_{it}) y el cambio en voltaje umbral puede representarse como:

$$\Delta V_{th} = -\frac{q}{C_{ox}} \Delta N_{it} \quad (2.1)$$

Entonces, el voltaje umbral es proporcional a la generación de trampas en la interface. La figura, 2.7 muestra la magnitud del cambio del voltaje umbral del transistor en función del tiempo, durante la etapa de estrés y recuperación. Debe

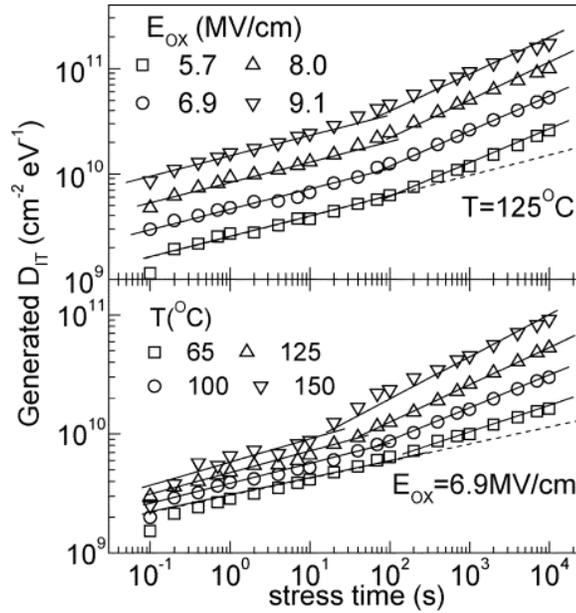


Figura 2.6: Dependencia del NBTI con el campo eléctrico y temperatura. [7]

notarse que aunque el tiempo dado a cada etapa es el mismo, la totalidad de los enlaces de $Si-H$ no son restablecidos. Por lo que el efecto general será aumento (en magnitud) del voltaje umbral. Si la proporción entre el tiempo en el que el dispositivo se encuentre en etapa de estrés aumenta respecto al tiempo en la etapa de recuperación, es de esperar que la desviación total del voltaje umbral sea aún mayor. Debe resaltarse de esta figura que el mayor incremento del voltaje umbral al comienzo de las etapas de estrés (recuperación) se debe a que el proceso de difusión del hidrógeno domina la generación (pasivación) de las trampas en la interface. Sin embargo, se llega a un momento en el cual la concentración de átomos de hidrógeno que pueden difundirse de la interface al óxido (del óxido a la interface) disminuye, haciendo que el cambio del voltaje umbral sea más lento [6].

Una característica importante del impacto del NBTI en el V_{th} , es que el aumento no depende de la frecuencia de conmutación (ver figura 2.8). Esto debido a que los mecanismos de difusión en la etapa de estrés son simétricos con respecto

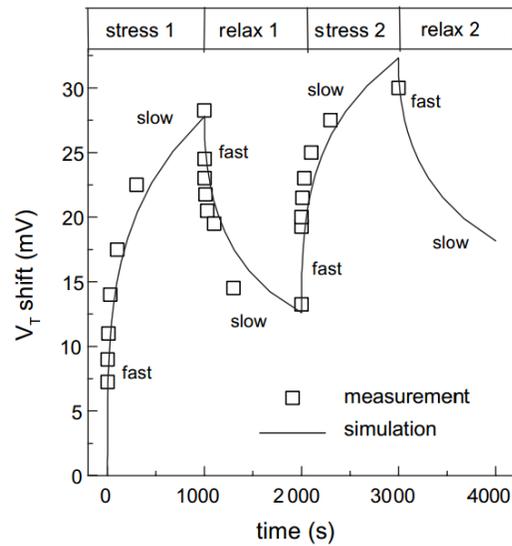


Figura 2.7: Desviación del voltaje umbral como consecuencia del NBTI. [6]

a los de la etapa de recuperación. La simetría de estos efectos se debe a que la cantidad de enlaces que son recuperados por la pasivación con el hidrógeno sólo dependen del mecanismo de difusión y no del campo eléctrico que sí es diferente entre una etapa y otra [8], esto debido a que el hidrógeno que se difunde tiene carga neutral. Para analizar la confiabilidad de un circuito digital, resultaría más realista considerar los efectos dinámicos del NBTI y no el peor caso de degradación estático.

Como consecuencia del incremento del voltaje umbral, la corriente de encendido de un transistor bajo una misma tensión de compuerta disminuye. Esto causa una reducción del tiempo de propagación de las compuertas digitales, ya que la capacidad de carga de los nodos por parte de los transistores PMOS disminuye. Si el incremento en retardo es significativo, podría ocasionar un retardo que viola las especificaciones de frecuencia de operación y un circuito dejaría de funcionar correctamente. Con el escalado tecnológico, el campo eléctrico que causa este fenómeno se ha incrementado, al igual que la temperatura debido al alto consumo de potencia, convirtiendo al NBTI en uno de los fenómenos de degradación que más afecta a los dispositivos.

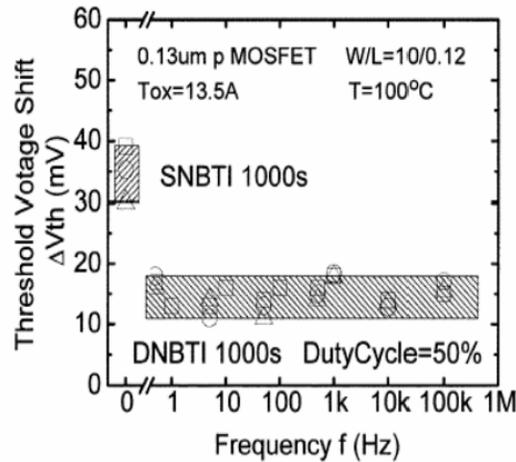


Figura 2.8: Dependencia del NBTI con la frecuencia de operación. [9]

2.2. Modelo predictivo de degradación del V_{th} debido a NBTI

En [12] y [13] desarrollan un modelo para determinar un borde superior para la desviación de V_{th} debido a NBTI (como se muestra en la figura 2.9), basándose en el modelo físico de reacción y difusión [6] que describe matemáticamente los conceptos expuestos anteriormente. En este modelo, el cambio del voltaje umbral depende de principalmente de las condiciones dinámicas de operación como el voltaje de alimentación (V_{dd}), la temperatura (T) y la probabilidad de estrés (α) que puede ser vista como un factor que determina la proporción del tiempo en el que el transistor será afectado por NBTI y el tiempo en el que se encontrará en estado de relajación ($\alpha = t_{stress}/(t_{stress} + t_{rec})$). Esta aproximación resulta más realista que considerar que los dispositivos se encuentran todo el tiempo bajo degradación estática y permite realizar una mejor estimación del impacto del NBTI en el rendimiento de un circuito digital. Este modelo es llamado *Long Term Degradation Model* (LTDM). El LTDM describe el aumento del voltaje umbral debido al efecto de NBTI en un transistor PMOS por la ecuación 2.2.

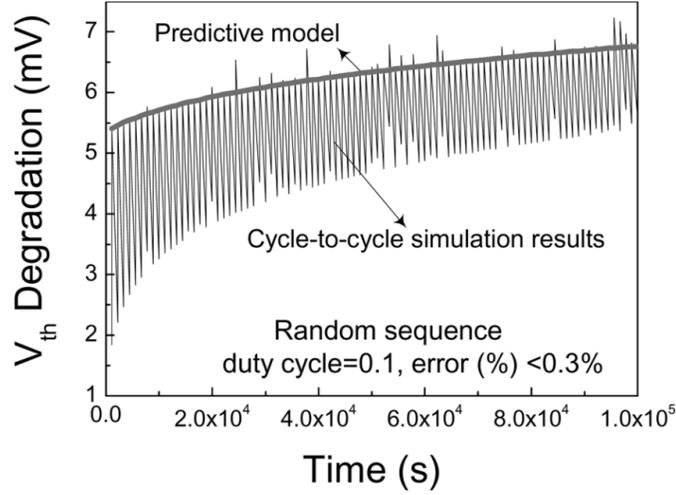


Figura 2.9: LTDM como borde superior de la degradación del V_{th} [14]

$$|\Delta V_{th_{NBTI}}| = \left(\frac{\sqrt{K_v^2 \alpha T_{clk}}}{1 - \beta_t^{1/2n}} \right)^{2n} \quad (2.2)$$

En esta ecuación, T_{clk} , α y n son el período de reloj, la probabilidad de estrés y una constante exponencial de tiempo con el valor de 1/6, respectivamente. K_v describe la dependencia con el voltaje de alimentación V_{dd} , la temperatura T , el espesor del óxido t_{ox} y otros parámetros asociados con el efecto de NBTI [13]. β_t refleja el impacto del tiempo de degradación.

En las figuras 2.10(a) y 2.10(b) se muestra el comportamiento de la degradación del V_{th} durante 10 años, obtenida mediante el LTDM para diferentes valores de probabilidad de estrés y de voltaje umbral inicial representado como una desviación debida a variaciones de proceso. Del comportamiento con respecto a la probabilidad de estrés, se puede observar que cuando la probabilidad es de cero (el transistor nunca esta en condición de estrés), la degradación por NBTI lógicamente es nula, a medida que se aumenta la probabilidad, mayor es el cambio en el voltaje umbral. De esta figura también es posible intuir que el cambio del voltaje umbral es menos sensible a cambios en la probabilidad cuando la probabilidad cambia de un valor mayor a uno menor por ejemplo de 0.9 a 0.1. Esto se debe a

2.2. MODELO PREDICTIVO DE DEGRADACIÓN DEL VTH DEBIDO A NBTI

que si el valor de α es alto y el transistor estará un mayor tiempo en etapa de estrés y la densidad de enlaces $Si - H$ que pueden romperse disminuyen con el tiempo. En el caso Por el contrario en el caso cuando se pasa de una probabilidad baja a una alta, siempre hay gran cantidad de enlaces que pueden romperse cuando el transistor se encuentre en estrés.

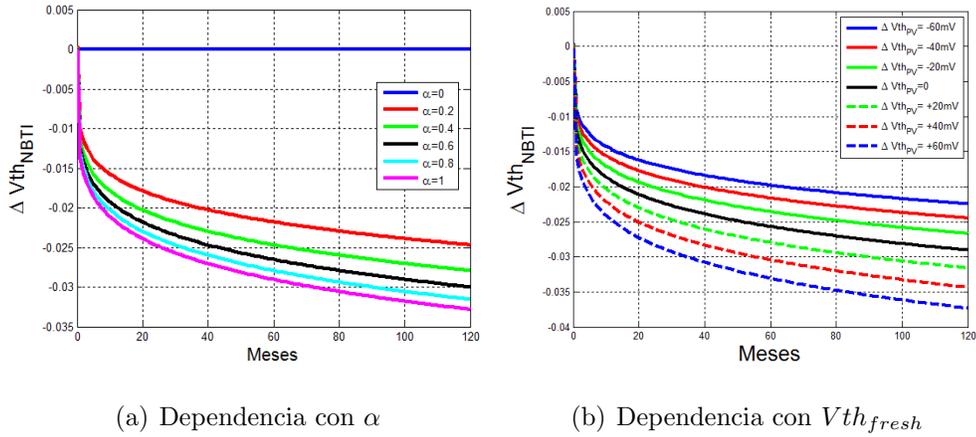


Figura 2.10: Curvas obtenidas al aplicar el LTDM

De la figura 2.10(b), teniendo en cuenta que el V_{th} de un transistor PMOS es negativo, es importante resaltar que transistores con valores de V_{th} iniciales mayores en magnitud ($\Delta V_{th_PV} < 0$) tendrán una menor degradación debido a NBTI. Como consecuencia, la diferencia inicial entre los voltajes de umbral de dos dispositivos disminuye con el tiempo gracias al envejecimiento. Debido a esta dependencia con el voltaje umbral inicial, el envejecimiento del circuito interactúa con las variaciones de proceso estáticas y se convierte en un fenómeno de carácter estadístico.

Como se mencionó anteriormente el término K_v de la ecuación del LTDM contiene la dependencia del NBTI con la tensión de alimentación y temperatura. En la ecuación 2.3 se puede observar los elementos que comprenden el factor K_v , en esta ecuación se identifica el voltaje de *overdrive* $|V_{gs} - V_{th_fresh}|$ en un transistor MOSFET, donde V_{th_fresh} es el voltaje de umbral en tiempo igual 0. El voltaje de *overdrive* esta relacionado con la cantidad campo eléctrico que

2.2. MODELO PREDICTIVO DE DEGRADACIÓN DEL VTH DEBIDO A NBTI

se aplica en la compuerta, el cual en términos de degradación por NBTI rompe enlaces de la interface Si-H y también arrastra huecos a esta interface. V_{gs} que es el voltaje entre la compuerta y la fuente del transistor, en el caso de un transistor PMOS $V_{gs} = -V_{DD}$ y entre más alto V_{DD} mas intenso es el campo eléctrico por lo cual el NBTI se incrementa, en la ecuación el término K_v aumenta. La disminución del impacto de la tensión en el NBTI se puede observar en la figura 2.11, cuando se baja la tensión de alimentación de 1.2V a 1.0V el incremento del voltaje de umbral por NBTI disminuye en un 50 %, estos resultados se obtienen a una temperatura constante de $120^{\circ}C$.

$$\uparrow K_v = \left(\frac{qtox}{\epsilon_{ox}} \right)^3 K^2 C_{ox} (\uparrow |V_{gs} - V_{th_{fresh}}|) \sqrt{C} \exp\left(\frac{2E_{ox}}{E_0}\right) \quad (2.3)$$

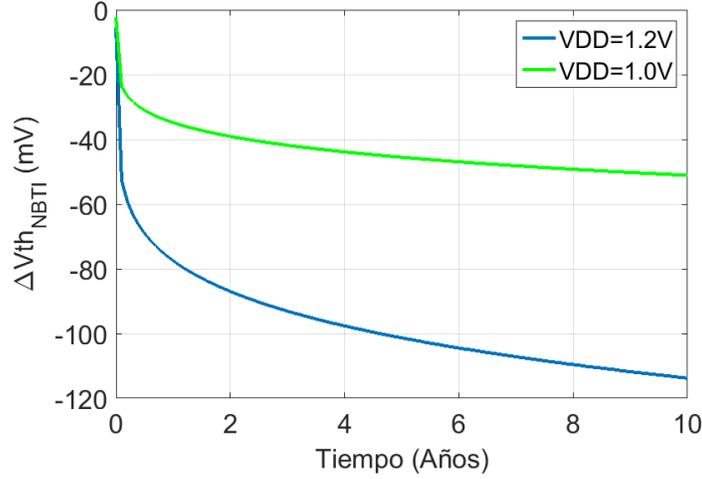


Figura 2.11: Variación del Voltaje de umbral debido a NBTI para diferentes voltajes de operación

El factor C que se describe en la ecuación 2.4 que esta dentro de la ecuación de K_v contiene el efecto de la temperatura. Al aumentar la temperatura mas enlaces de Si-H son rotos lo cual incrementa el efecto del NBTI, el impacto de la temperatura en el NBTI es muy significativo como se aprecia en la ecuación el comportamiento del NBTI frente a la temperatura es exponencial, por lo cual

disminuir la temperatura a la cual va operar el circuito es uno de los métodos mas eficientes para evitar el NBTI, en la figura 2.12 se observa que al disminuir la temperatura de 120°C a 90°C (-25 %) el impacto del NBTI disminuye en un 40 % mientras al disminuir la temperatura a la mitad esto es 60°C el impacto del NBTI disminuye en un 75 %, estos resultados se obtienen a una tensión de alimentación constante de 1.2V.

$$C = \frac{\exp(-E_a/kT)}{T_0} \quad (2.4)$$

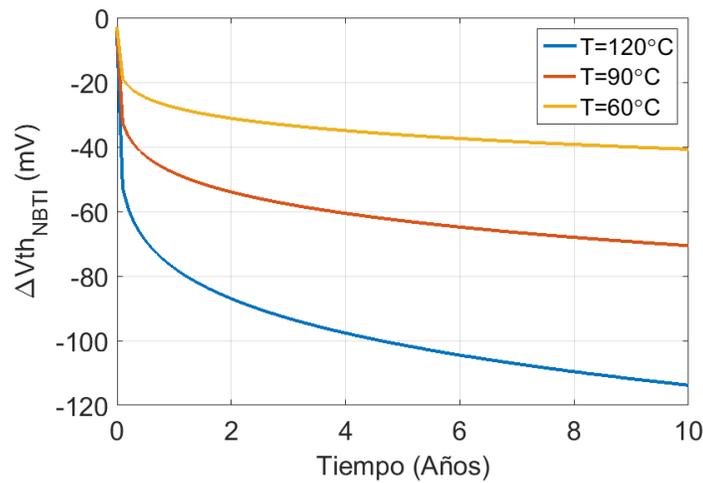


Figura 2.12: Variación del Voltaje de umbral debido a NBTI para diferentes temperaturas de operación

Pero tomar una temperatura alta con un voltaje de alimentación alto es un caso pesimista y como se verá más adelante en el capítulo 4 hay una relación entre la temperatura a la cual opera la compuerta y la tensión de alimentación, por lo cual el impacto del NBTI puede bajar aún más, esto se puede observar en la figura 2.13a donde la variación del voltaje de umbral debido a NBTI esta en función de la temperatura y la tensión de alimentación. Si se toma la tensión de alimentación igual 1.0V se obtiene la figura 2.13b, en la cual a una temperatura de 60°C la reducción del impacto del NBTI ha bajado a un 84 %.

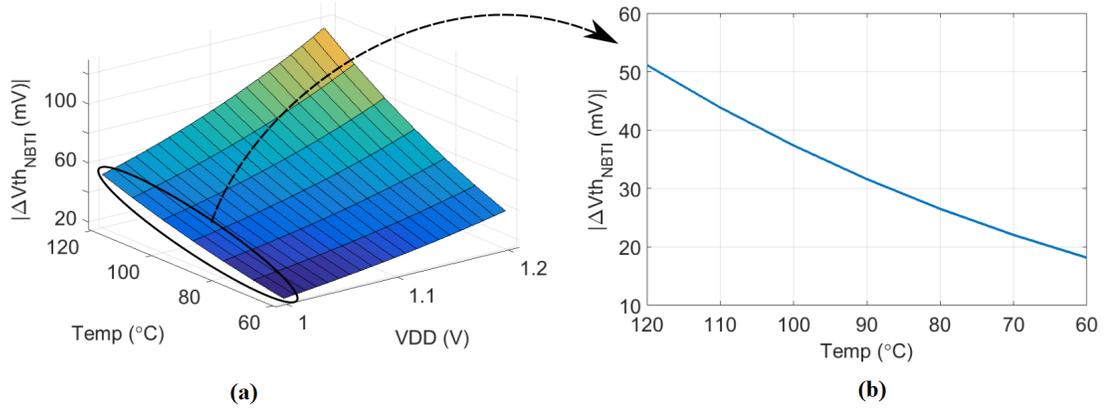


Figura 2.13: Variación del Voltaje de umbral debido a NBTI para diferentes temperaturas y tensiones de operación

2.3. Impacto en el rendimiento de un circuito digital

El incremento del V_{th} causado por las trampas en la interface generadas por el efecto de NBTI, tiene una relación directa con el retardo de propagación de las compuertas digitales en un camino lógico. Este incremento puede ser visto a través del concepto de *sensibilidad*, que expresa el impacto que tiene la desviación de un parámetro de cada transistor de la compuerta, en el rendimiento de la misma [17]. El efecto del NBTI puede calcularse sumando la contribución que tiene la desviación de cada uno de los parámetros que se degradan con el tiempo de cada transistor por separado en el rendimiento nominal. Suponiendo que únicamente el voltaje umbral presenta degradación debido a NBTI, el retardo de una compuerta digital puede expresarse como [2.5]:

$$D_{age} = D_o + \sum_{k=1}^K S_{V_{th},k}^D \Delta V_{th_{k,nbti}} \quad (2.5)$$

En la anterior ecuación, K es el número de transistores en la compuerta y D_o es el retardo nominal de compuerta. El impacto de la degradación del V_{th} de cada transistor k en el tiempo de propagación, es el producto del coeficiente de

2.3. IMPACTO EN EL RENDIMIENTO DE UN CIRCUITO DIGITAL

sensibilidad del retardo debido a variaciones en el V_{th} del transistor $S_{V_{th},k}^D$ con la cantidad de desviación del parámetro $\Delta V_{th,k,nbti}$.

Los coeficientes de sensibilidad pueden determinarse como la derivada del retardo respecto al voltaje umbral en el punto nominal:

$$S_{V_{th},k}^D = \left. \frac{\partial D}{\partial \Delta V_{th,k}} \right|_{\Delta V_{th}=0} \quad (2.6)$$

Es importante señalar en este punto, que la variación del V_{th} de la ecuación anterior, es puramente determinista.

La figura 2.14 muestra el cambio en el tiempo del retardo de propagación de una compuerta inversora obtenido con el modelo anterior. En la figura se observa que el incremento en el retardo es proporcional al cambio en el V_{th} visto en la figura 2.10(a).

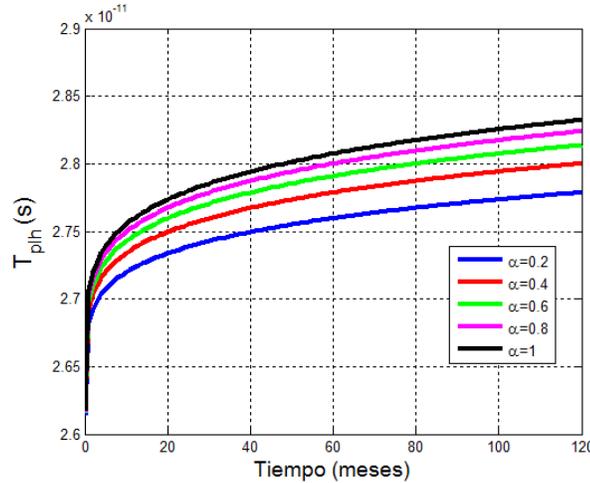


Figura 2.14: Variación del retardo de subida de una compuerta inversora

El impacto del NBTI en el retardo de un camino lógico, puede calcularse sumando cada una de las contribuciones de las compuertas por separado:

$$D_{camino\logico,age} = \sum_{i=1}^N D_{age,i} \quad (2.7)$$

2.3. IMPACTO EN EL RENDIMIENTO DE UN CIRCUITO DIGITAL

Donde N es el número de compuerta en el camino lógico y $D_{age,i}$ es el retardo considerando efectos de envejecimiento de la i -ésima compuerta en el camino lógico.

En un circuito digital, la probabilidad de las señales de entrada principales puede tomar cualquier valor entre 0 y 1, esto implica que para cualquier combinación de valores, la degradación del retardo del circuito puede tomar un valor diferente. Como se muestra en la figura 2.15 para dos circuitos diferentes llamados “parity” y “9symml” [14], aplicar un patrón de entrada diferente causa una degradación de retardo diferente. Además la diferencia de retardo obtenida entre la degradación con un patrón de entrada y otro puede aumentar considerablemente a través del tiempo [14]. Debido a la dificultad de conocer con exactitud el valor de probabilidad de las señales de entrada, es necesario analizar el impacto en el rendimiento bajo las peores condiciones de probabilidad que generarían una máxima degradación.

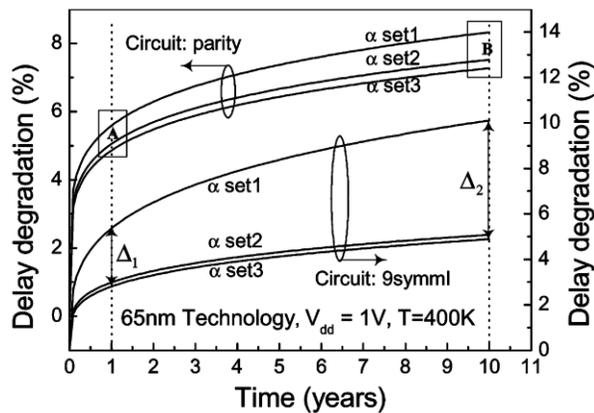


Figura 2.15: Degradación del retardo para diferentes conjuntos de probabilidad en las entradas principales de un circuito digital. [14]

Además, debe tenerse en cuenta que un patrón de entrada podría afectar en gran parte el retardo de un camino lógico determinado (supongamos un camino lógico no crítico) mientras que el impacto que tiene en otro (supongamos el camino lógico crítico) podría ser muy poco. En este caso, un camino lógico que, bajo un

2.3. IMPACTO EN EL RENDIMIENTO DE UN CIRCUITO DIGITAL

análisis sin envejecimiento, no se consideraba como camino lógico crítico, al pasar el tiempo puede convertirse en él o viceversa [14]. La figura 2.16 muestra este efecto en el circuito ISCAS C17 donde un camino no crítico (I_{10}) es más sensible al patrón de entrada aplicado que el camino crítico inicial I_9 . Entonces cuando se realiza un análisis de retardo no sólo debe considerarse el camino lógico crítico, sino también el conjunto de caminos lógicos que potencialmente podrían convertirse en caminos críticos.

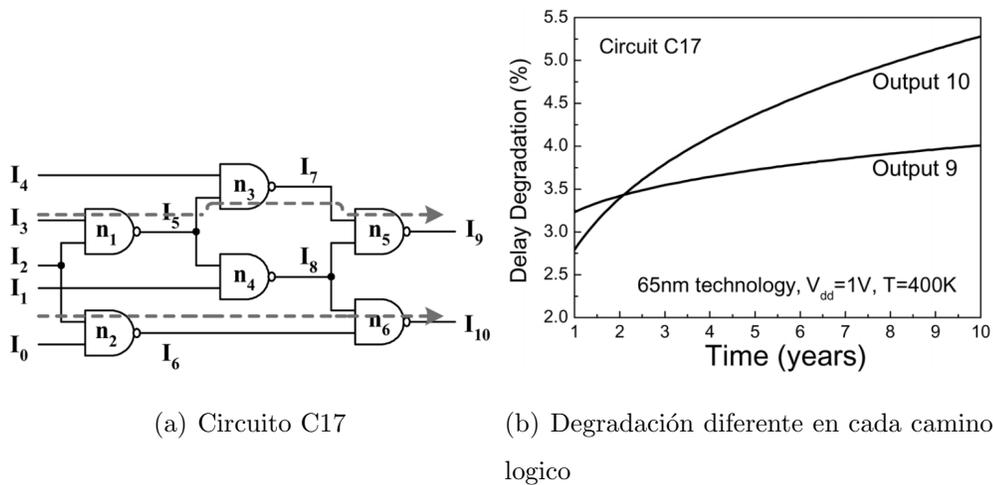


Figura 2.16: Degradación del retardo diferente para dos caminos del circuito C17. [14]

En [17] y [23] mencionan que debido a NBTI, el retardo del camino lógico crítico (CP) de un circuito digital puede incrementarse alrededor de un 10%. Si éste incremento es tal que el retardo del CP es mayor que el período completo de reloj, podría ocurrir una falla en el sistema como se muestra en la figura 2.17. Inicialmente, cuando el circuito no se ha envejecido y el CP realiza una operación, la transición que debe ser almacenada en los registros o FFs ocurre bastante antes del flanco positivo de reloj (donde los datos son almacenados). Sin embargo, a medida que pasa el tiempo, el circuito se envejece y la señal digital tarda más tiempo en propagarse y podría realizar la transición luego del flanco de reloj. Entonces, el dato almacenado no sería el correcto y ocurriría un error de

operación.

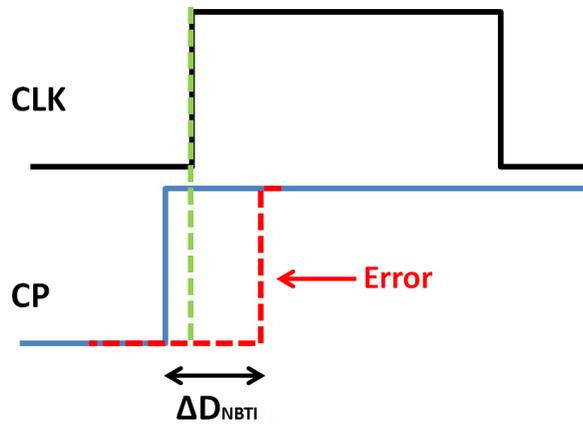


Figura 2.17: Error de operación debido al envejecimiento

Como se mencionó en el capítulo 1, para asegurar que el circuito opere correctamente durante todo su tiempo de vida útil esperado, usualmente se introduce una banda de guarda adicional en el período de reloj inicial, de forma que en los primeros años de operación el circuito opera más rápido de lo necesario, pero al pasar el tiempo, opera justo sobre la especificación de retardo. La banda de guarda requerida puede ser entre un 10 % o 20 % del retardo inicial del circuito [10] [21] [14].

Capítulo 3

Técnicas para reducir el impacto del envejecimiento en circuitos integrados digitales

En los capítulos anteriores, se han explicado los fundamentos del efecto de NBTI y cómo se ve reflejado este fenómeno ocasionando un incremento en el retardo de propagación de las señales digitales a través de una compuerta y un camino lógico. De éstos análisis se concluyó que el incremento de retardo debido a NBTI depende principalmente de la probabilidad de que los transistores de una compuerta se encuentren bajo la condición de estrés, de la temperatura de operación y el voltaje de alimentación.

Para evitar que ocurran errores de propagación, éste incremento del retardo es considerado al ajustar el período de reloj inicial de un circuito digital, introduciendo la banda de guarda. Sin embargo, esto ocasiona que no se aproveche al máximo la capacidad de los dispositivos en su primera etapa del tiempo de vida.

Las principales técnicas o estrategias para disminuir el impacto del NBTI en un circuito digital y así reducir la banda de guarda requerida pueden clasificarse según el nivel en el que se realizan:

- Técnicas a nivel transistor: Consisten en modificar las propiedades físicas de los materiales utilizados en el proceso de fabricación, con el objetivo de que el impacto del NBTI sea menos significativo en las características eléctricas del dispositivo.
- Técnicas a nivel circuital: Se basan en estrategias ya sea para que el incremento del retardo debido a la degradación sea menor y así minimizar el margen de seguridad requerido o para optimizar el rendimiento a través del tiempo según el envejecimiento de los dispositivos. Las técnicas a nivel circuito pueden enfocarse en optimizar el desgaste durante la etapa de operación activa o durante los tiempos en el que el circuito no es utilizado y se encuentra en estado de standby.

A continuación se revisarán algunas de estas técnicas haciendo un mayor enfoque en las estrategias adaptivas que hacen parte de las estrategias a nivel circuito en estado activo de operación.

3.1. Incremento de la movilidad efectiva de los huecos

En [19] sugieren que la disminución de la corriente del transistor debido al aumento de V_{th} por NBTI, puede ser compensado si la movilidad de los huecos se incrementa a medida que el dispositivo se envejece. El fundamento de esta idea parte de que la corriente en la región lineal de un transistor PMOS puede ser expresada como $|I_D| \sim \mu_{eff}(|VG| - |V_{th}|)$. Por tanto, la variación en la corriente, el voltaje y la movilidad puede ser relacionada por:

$$\frac{\Delta I_D}{I_{D0}} = \frac{\Delta \mu_{eff}}{\mu_{eff0}} - \frac{|\Delta V_{th}|}{VG - V_{th0}} \quad (3.1)$$

Donde los parámetros con subíndice (0) corresponden al valor inicial sin degradación.

3.1. INCREMENTO DE LA MOBILIDAD EFECTIVA DE LOS HUECOS

A medida que el transistor se degrada, las trampas en la interface son generadas y la magnitud del voltaje umbral del transistor aumenta, contribuyendo negativamente al cambio en la corriente a través del segundo término de la ecuación 3.1. Sin embargo, debido a la generación de carga positiva en la interface, el campo eléctrico efectivo tampoco puede permanecer constante ($|E_{eff}| = \frac{Q_{dep}+Q_{inv}/3}{\epsilon_{Si}}$), de hecho debe disminuir, ya que las trampas en la interface disminuyen la carga de la capa de inversión y de la región de agotamiento. Entonces, el cambio en la corriente puede ser compensado aprovechando la relación entre el campo eléctrico y la movilidad.

Los autores proponen el uso de transistores de silicio tenso (*Strained Silicon*) ya que en ellos la pendiente entre la movilidad y el campo eléctrico es negativa como muestran en los resultados experimentales de la figura 3.1. En este caso, una reducción en el campo eléctrico debido a NBTI, ocasiona un aumento en la movilidad efectiva que podría compensar el incremento de la magnitud del V_{th} y reducir el cambio total de la corriente. En la figura se muestran curvas para dos materiales hipotéticos cuya relación de movilidad y campo eléctrico efectivo tiene la pendiente suficiente para compensar el envejecimiento.

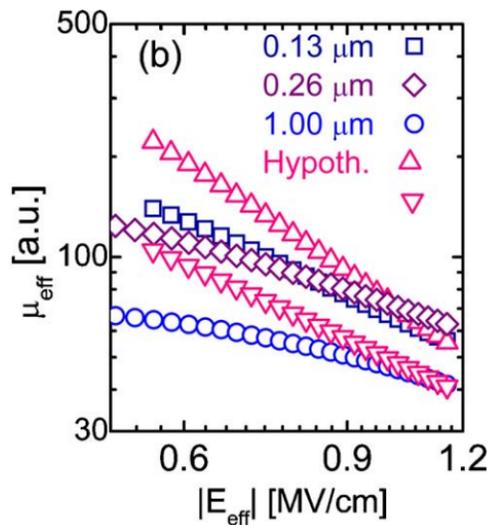


Figura 3.1: Curvas de movilidad vs campo eléctrico en silicio forzado [19]

La desventaja de esta estrategia sería que utilizar este tipo de materiales implicaría un alto costo en el proceso de fabricación.

3.2. Control del vector de entrada

Esta técnica consiste en aplicar el vector de entrada al circuito digital que genere una menor degradación debido a NBTI cuando el circuito se encuentre inactivo o en standby. Para determinar un vector óptimo evitando realizar barridos exhaustivos, es posible utilizar un generador de vectores aleatorio y seleccionar el vector más adecuado. Debido a que el vector óptimo para disminuir el envejecimiento por NBTI puede no ser el mejor en cuanto a consumo de potencia, en [20] proponen analizar los vectores en un diagrama de pareto como el mostrado en la figura 3.2 y de esta forma seleccionar el vector que mejor optimice ambos valores. En esta técnica, la penalización requerida en área y tiempo por las memoria o flip-flops que almacenen el vector óptimo en las entradas principales puede ser despreciada en circuitos grandes.

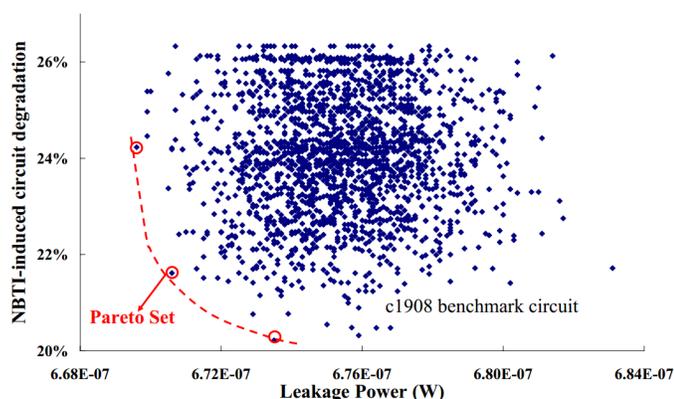


Figura 3.2: Conjunto de pareto para disminuir efecto de NBTI y consumo de potencia simultáneamente durante el período de stanby [20]

Aunque la técnica de control de los vectores de entrada permite disminuir el impacto del NBTI, su eficiencia puede no ser suficiente en circuitos grandes [23]

[24] ya que sólo permite controlar ciertos nodos internos.

3.3. Control de Nodos Internos

Otra estrategia para la disminución del efecto de NBTI durante el período de tiempo en reposo (standby) de un circuito digital, consiste en reemplazar la estructura de algunas compuertas digitales e insertar una señal extra que controle el estado de algunos nodos internos, especialmente los de entrada de aquellas compuertas más afectadas por el NBTI durante el período de operación dinámico. La figura 3.3 ilustra el concepto de esta técnica. Los autores de [23] y [24] utilizan la estructura de la figura 3.3a para controlar el nodo a la salida de las compuertas tipo NAND-AND, mientras que la de la figura 3.3b se utiliza para controlar las compuertas tipo NOR-OR.

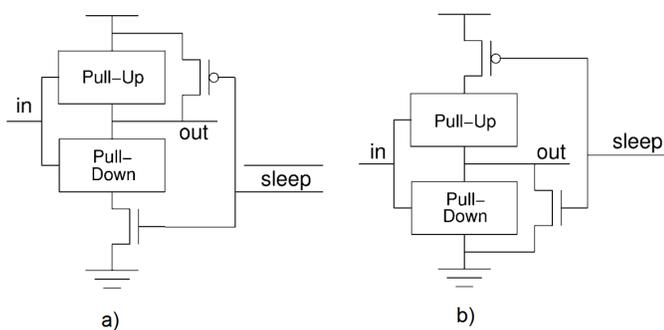


Figura 3.3: Técnica de reemplazo de compuertas para controlar el estado de algunos nodos internos

Al activar la señal de *sleep* (únicamente en el estado de standby), se cambia el estado de algunos nodos internos, llevando ciertas compuertas (que tienen estos nodos como fan-in) de la condición de estrés a la de relajación o recuperación. Bajo esta metodología, una compuerta de dos entradas, se convierte en una de tres, por tanto, no se requiere estrictamente una estructura de compuertas diferentes a las de las celdas estándar. Sin embargo, la técnica intrduce penalizaciones en retardo, consumo de potencia y área. La penalización en retardo, es en parte

compensanda a largo plazo por la reducción del impacto del NBTI. El consumo de potencia dinámico durante el lapso de tiempo de operación activo o en standby es despreciable ya que la señal de sleep permanece a un único estado lógico. Por otra parte, se incrementa la complejidad en la etapa de ruteado [24]. Tanto en [23] como en [24] presentan algoritmos para seleccionar adecuadamente las compuertas que deben ser reemplazadas y optimizar la reducción del efecto de NBTI junto con la técnica de control del vector de entrada.

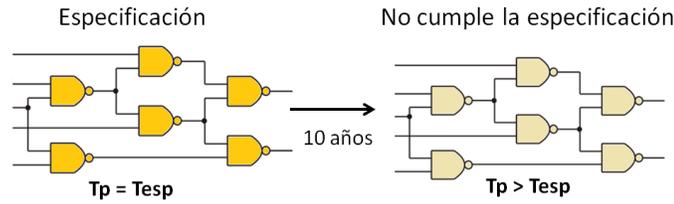
3.4. Redimensionamiento de las compuertas

El sobredimensionamiento de las compuertas es una forma de cumplir la especificación de retardo, ya que los dispositivos funcionarán más rápido durante su tiempo de operación [18], de forma que cuando envejezcan, aún cumplan la mínima restricción de tiempo. La metodología consiste, como se ilustra en la figura 3.4, en no realizar el diseño pensando en cumplir la especificación nominal, sino teniendo en cuenta el porcentaje de incremento del retardo que se tendrá durante el tiempo de vida útil esperado.

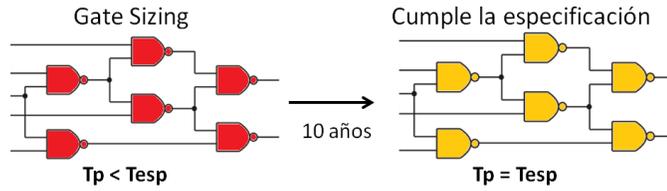
En este tipo de técnicas, resulta importante un cálculo adecuado de la probabilidad de las señales de entrada en cada compuerta digital y una selección precisa de las compuertas que serán redimensionadas ya que de lo contrario la penalización en área podría ser significativa.

En [21] asignan los valores de probabilidad de las entradas principales y éstos son propagados a través de todo el circuito, de esta forma evitan realizar el sobredimensionamiento para un peor caso de envejecimiento (donde se consideraría que la probabilidad de que todos los transistores se encuentren en condiciones de estrés es de la unidad, $\alpha = 1$). Sin embargo, esta suposición no es viable, ya que los valores de probabilidad de entrada, en general no pueden ser estimados adecuadamente para cualquier circuito combinacional.

En [18] sugieren optimizar las compuertas más cercanas a las salidas princi-



(a) Diseño para la especificación



(b) Sobre dimensionamiento para cumplir la especificación luego de 10 años

Figura 3.4: Sobre dimensionamiento de las compuertas

pales, ya que la señal digital que llega a estas compuertas presenta una mayor degradación en el slew rate (debido a el incremento del V_{th} de las compuertas anteriores) y por tanto, su impacto en el incremento del retardo de éstas últimas compuertas será mayor.

Por otra parte, en [22] proponen una estrategia para posicionar las compuertas en un camino lógico según su incremento del retardo y aquellas que tengan una degradación mayor a un porcentaje umbral, son optimizadas.

3.5. Ordenar los nodos de entrada en las compuertas

Como se explicó en el segundo capítulo, el modelo de degradación del V_{th} por NBTI, depende de la probabilidad de que el transistor se encuentre bajo condiciones de estrés (α). En una compuerta inversora o compuerta de tipo NAND, donde los transistores PMOS se encuentran en paralelo y su fuente está conectada direc-

3.5. ORDENAR LOS NODOS DE ENTRADA EN LAS COMPUERTAS

tamente al riel de alimentación, la probabilidad de que el transistor se encuentre en condiciones de estrés es igual a la probabilidad de que la señal de entrada se encuentre en nivel lógico de cero (ver figura 3.5).

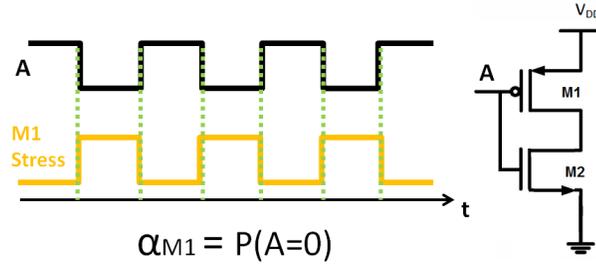


Figura 3.5: Condiciones de estrés en una compuerta inversora

Por otra parte, en compuertas donde hay dos o más transistores PMOS en serie (o *stack*), para que un transistor esté bajo condiciones de estrés, todos los demás transistores en una posición superior a él, también deben estarlo. Entonces, la probabilidad de que un transistor esté bajo estrés corresponde a la probabilidad de que su señal de entrada esté en cero lógico y que todos los demás transistores arriba de él estén bajo estrés. En la figura 3.6 se observa que aunque en algunos intervalos de tiempo, la señal B tiene el valor de cero lógico, el transistor *M2* no se encuentra polarizado en la condición de estrés, debido a que la tensión en su fuente no corresponde a *Vdd*, ya que el transistor *M1* se encuentra apagado.

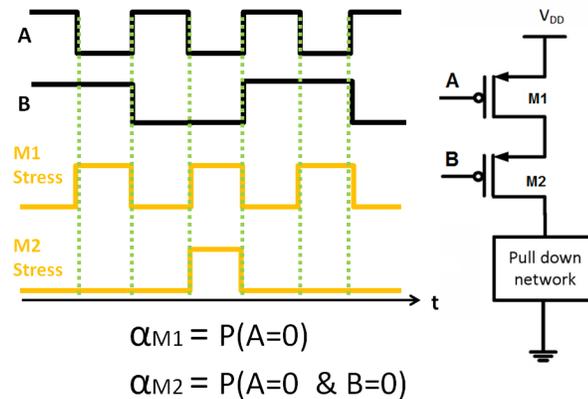


Figura 3.6: Condiciones de estrés en una compuerta NOR de dos entradas

Entonces, la idea de esta técnica es la de manipular las probabilidades de estrés, ordenando los nodos de entrada de forma que el efecto del NBTI en las compuertas críticas pueda ser reducido [27].

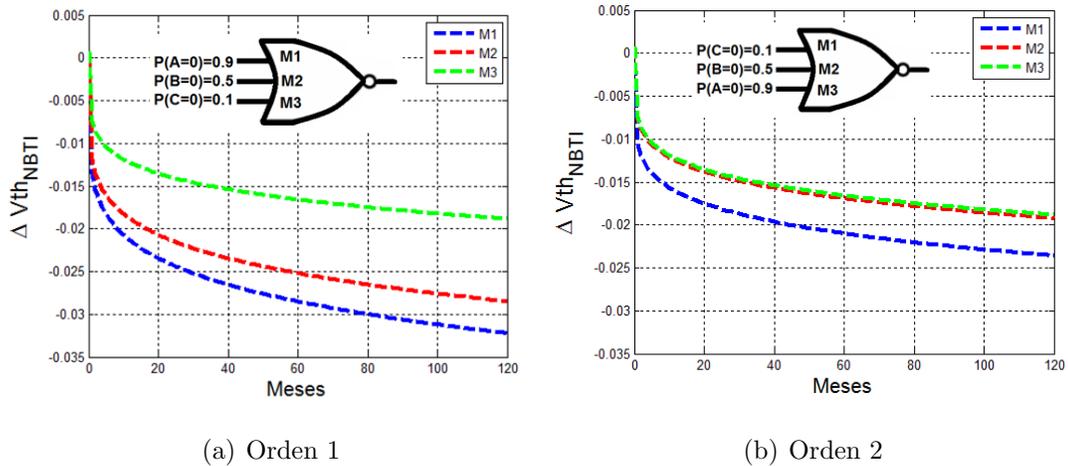


Figura 3.7: Ordenamiento de los nodos de entrada para disminuir el envejecimiento en una compuerta NOR3

La figura 3.7 muestra la disminución del efecto del NBTI en los transistores M2 y M3 obtenida en una compuerta NOR de tres entradas, tan sólo con realizar un adecuado ordenamiento de los nodos de entrada en función de su probabilidad. De esta figura también es obvio, que la degradación de los transistores arriba del stack siempre será mayor que los de abajo. Nótese además que el orden de los nodos adecuado es de menor a mayor probabilidad comenzando por el transistor PMOS conectado directamente a V_{dd} (M1).

3.6. Circuitos Adaptivos

La estrategia de utilizar circuitos adaptivos, busca reducir la banda de guarda requerida realizando una sintonización de los parámetros del sistema (frecuencia o voltaje) a medida que el circuito se desgasta con el tiempo.

El proceso de sintonización, puede llevarse a cabo de forma fuera de línea

(*offline*) que consiste en realizar el proceso de sensado de envejecimiento únicamente cuando el circuito se encuentre en reposo (viable en sistemas con múltiples núcleos), o en los momentos de encendido (por ejemplo cada vez que se enciende un automóvil). Este tipo de sintonización puede ser imperceptible para el usuario y pueden aplicarse gran cantidad de patrones a la entrada para sensar el retardo de cada camino lógico deseado [29]. Sin embargo, existen aplicaciones que se encuentran operando constantemente y detener su funcionamiento para realizar un test de envejecimiento resulta inaceptable. Para este tipo de aplicaciones es útil realizar un monitoreo en línea (*on-line*) de forma paralela al tiempo de operación, haciendo un seguimiento más exhaustivo.

Las variables del sistema que se pueden modificar según el envejecimiento de un circuito son:

- Voltaje de alimentación: La corriente que fluye por los dispositivos es proporcional al voltaje de alimentación utilizado. Entonces, para una especificación de frecuencia dada, se disminuye el voltaje de alimentación cuando el circuito se encuentra sin envejecimiento y opera rápidamente, reduciendo el consumo de potencia. Respectivamente, el voltaje de alimentación se aumenta a medida que éste se envejece.
- Frecuencia de Operación: De forma análoga al caso del voltaje de alimentación, se aumenta la frecuencia de operación cuando el circuito se encuentra joven, teniendo un máximo rendimiento, y luego se disminuye a medida que el circuito se desgasta. De esta forma, se obtiene un rendimiento óptimo durante todo el ciclo de vida útil.
- Efecto cuerpo: Esta idea consiste en aprovechar el efecto cuerpo para compensar el incremento del voltaje umbral.

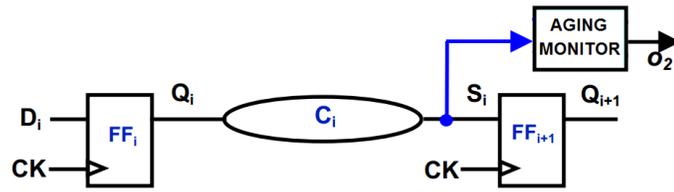
Las técnicas adaptivas requieren el uso de sensores o monitores para estimar que tanto se ha degradado el retardo de los caminos críticos del circuito debi-

do al envejecimiento y así poder realizar la sintonización requerida. Al medir el envejecimiento en un camino lógico mediante estos sensores, es posible reducir significativamente la banda de guarda requerida, ya que a medida que el camino lógico de interés se envejece, el cambio en el retardo es detectado por los sensores y estos envían una señal de alerta a los controladores de los parámetros de operación mencionados anteriormente.

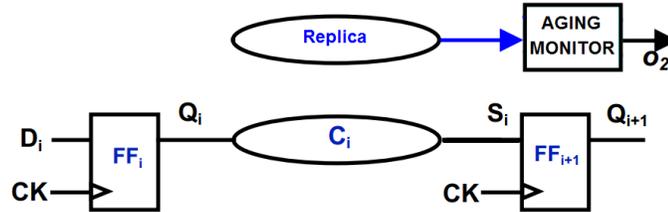
3.6.1. Camino lógico réplica y sensores embebidos

Los sensores pueden ser insertados ya sea directamente a la salida del camino lógico de interés (de forma embebida) o a la salida de un camino lógico réplica aislado que emula en lo posible las características del camino lógico de interés. La figura 3.8 ilustra estas posibilidades. El camino lógico réplica, es fácil de implementar y puede ser diseñado de forma que pueda calibrarse luego del proceso de fabricación (aumentando los costos de test), para así considerar posibles diferencias debido a variaciones de proceso [30]. Sin embargo, la degradación del camino lógico crítico de interés depende altamente de la probabilidad de las señales de entrada a la lógica combinatorial, y conectar éstas al camino lógico réplica trae complicaciones de ruteado e integridad de señal. Además, el camino lógico crítico de un circuito digital puede cambiar a través del tiempo y la réplica no puede emular correctamente a todos. Para solucionar los inconvenientes mencionados, al utilizar un camino lógico réplica, debe garantizarse que su retardo sea mayor y que éste se degrade más rápido que cualquier otro camino lógico en el circuito [30] [29].

Por otra parte, insertar sensores directamente a la salida de los caminos lógicos críticos del circuito, no requiere de calibraciones post layout [30], ya que el retardo es medido directamente del camino lógico de interés. Además permiten reducir al máximo la banda de guarda requerida, ya que cualquier incremento de retardo debido no sólo al envejecimiento será correctamente detectado por los



(a) Sensor Embebido



(b) camino lógico réplica + Sensor

Figura 3.8: Tipos de sensores

sensores. Para reducir el consumo de potencia, resulta útil activar los sensores únicamente cuando el camino lógico crítico se active, ya que múltiples caminos pueden compartir una misma salida principal pero sólo algunos son potencialmente críticos. Igualmente, garantizar que el sensor se activa únicamente cuando el camino lógico crítico lo hace permite alternativas como el control de la velocidad, tanto aumentando como disminuyendo el período de operación cuando no se han detectado transiciones tardías por parte de los caminos lógicos críticos durante largos períodos de tiempo [29]. Sin embargo, las desventajas de utilizar esta forma de sensado es que cargan directamente el camino lógico crítico, teniendo una penalización en el retardo de propagación y además es más complejo el proceso de insertar los sensores y realizar una correcta selección de los caminos lógicos que deben ser monitoreados que en el caso del camino lógico réplica.

3.6.2. Predicción de error

Esta técnica consiste en predecir el punto operacional donde el camino lógico crítico del circuito falla en cumplir las especificaciones de retardo (cuando el retar-

do del camino lógico crítico está cerca del flanco de reloj) y garantizar la correcta operación añadiendo márgenes de seguridad (mucho menores a las requeridas en un circuito convencional) para predecir el punto de falla [31].

La figura 3.9 ilustra la estructura general de un sensor de predicción de error y su funcionamiento puede explicarse con las señales de la figura 3.10. Cuando el camino lógico crítico del circuito se encuentra sin envejecimiento (CP without aging), realiza una transición con suficiente tiempo de antelación al flanco positivo del reloj, donde los flip-flops muestrean y retienen el dato propagado, de esta forma, ambos FF almacenan el mismo valor y el circuito continúa operando de forma convencional.

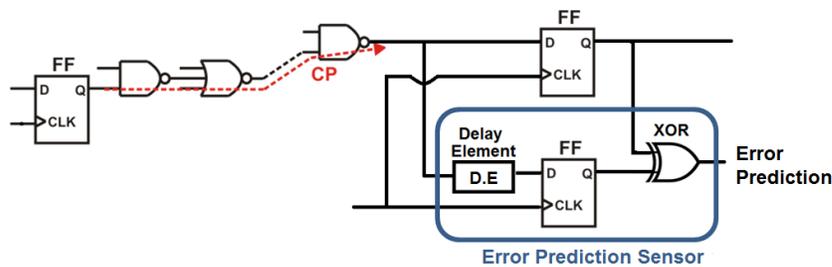


Figura 3.9: Sensor de predicción de error

Cuando el retardo del camino lógico crítico se incrementa debido al envejecimiento (CP with aging), la transición que realiza está más cerca del flanco de reloj (más cerca de generar una falla), si el tiempo entre el momento en el que ocurre esta transición y el flanco de reloj es menor que el retardo definido como margen por el elemento de retardo del sensor (Delay Element), la señal retardada a la salida de este elemento realizará su transición después del flanco de reloj y esta transición ya no será almacenada en el FF del sensor. Entonces, las señales retenidas en el FF principal y en FF del sensor serán diferentes y al compararse en la compuerta XOR, activarán la señal de alerta (Error Prediction), indicando que el período de reloj debe aumentarse.

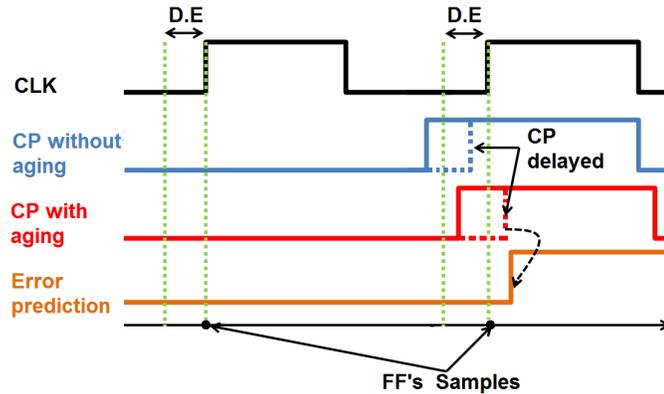


Figura 3.10: Funcionamiento del sensor de predicción de error

3.6.3. Detección y corrección de error

El concepto base de estos esquemas es el de escalar los parámetros del sistema sólo hasta el punto en el que un error de tiempo ha ocurrido [33] [34]. Una vez el error es detectado, debe realizarse una lógica de recuperación del dato correcto, para que el circuito pueda continuar operando correctamente. Al permitir que el circuito falle, pueden reducirse al máximo las márgenes de seguridad requeridas para tolerar los peores casos de retardo de propagación, obteniendo un rendimiento significativamente mayor, incluso que las técnicas de predicción de error [30]. Aunque este tipo de técnicas pueden tolerar variaciones de retardo de todo tipo (alta frecuencia o baja frecuencia), existe un compromiso entre el beneficio de una menor frecuencia de operación y el tiempo extra requerido por los procesos de recuperación de error.

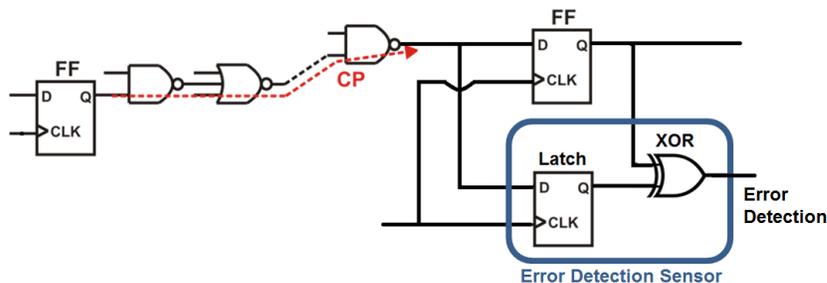


Figura 3.11: Sensor de detección de error

La figura 3.11 muestra la estructura general de este tipo de sensores y su funcionamiento puede ser entendido por el diagrama de tiempos de la figura 3.12. Cuando el camino lógico crítico se encuentra sin envejecimiento (CP without aging), el circuito opera de forma convencional (nótese que en este caso, la transición del CP puede ocurrir muy cerca del flanco de reloj, desde luego, manteniendo el tiempo de disposición (setup) y de sostenimiento (hold) requeridos por el FF). A medida que el retardo del camino lógico se incrementa debido al envejecimiento de los dispositivos, eventualmente su retardo de propagación puede superar incluso el período de reloj, realizando una transición luego del flanco positivo del reloj (CP with aging), la cual no será correctamente almacenada en el FF principal. Sin embargo, durante el período de tiempo en el que la señal de reloj se encuentra en alto, el latch (también llamado “*shadow latch*”) se encuentra en estado de transparencia y éste sí almacena la señal tardía. De esta forma, el dato presente en el FF principal será diferente al dato en el latch, activando la señal de error al ser comparados en la compuerta XOR.

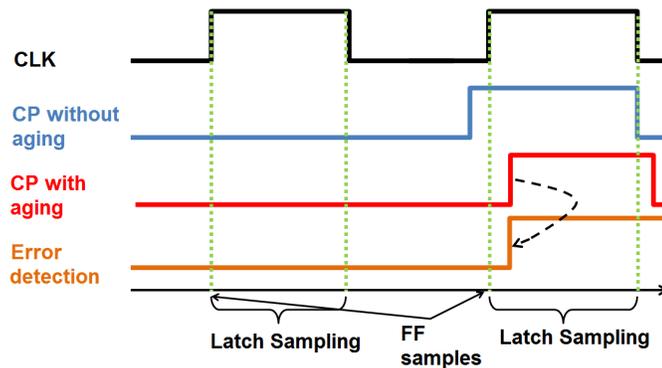


Figura 3.12: Funcionamiento del sensor de detección de error

Una vez la señal de error es activada, la lógica de recuperación del dato debe llevarse a cabo. Para esto usualmente se repite nuevamente la operación realizada a una menor frecuencia de operación o mayor voltaje de alimentación [30].

3.7. Disminución del voltaje de alimentación

Como su nombre lo indica, consiste en aplicar un voltaje de alimentación por debajo de la nominal a la cual trabajan las compuertas, de esta forma se reduce el efecto de NBTI ya que el campo eléctrico entre la compuerta y la fuente disminuye, lo cual disminuye la cantidad de enlaces Si-H que se pueden romper, por lo tanto también disminuye el impacto del NBTI. Es importante mencionar que al disminuir el voltaje de alimentación también disminuye el consumo de potencia de una compuerta, por lo cual también se disminuye la temperatura a la cual opera la compuerta la cual depende de la potencia disipada, esto hace que disminuya aun más el impacto del NBTI, es decir que al disminuir el voltaje de alimentación el impacto del NBTI se ataca por dos lados, directamente por el voltaje e indirectamente por la temperatura.

En la gráfica 3.13 se observa cómo disminuye la degradación del voltaje de umbral al trabajar con un voltaje menor al nominal. En esta técnica se selecciona las compuertas candidatas a tener un voltaje de alimentación menor de tal manera que no se impacte el retardo de manera apreciable. Sin embargo, para utilizar esta técnica requiere insertar rieles extra de alimentación.

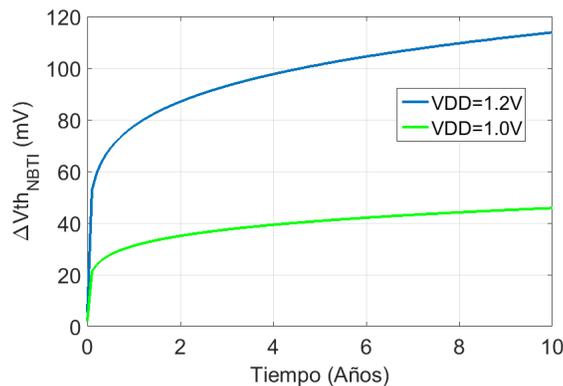


Figura 3.13: Disminución del efecto NBTI al bajar el voltaje de alimentación

Otro aspecto de esta técnica es su impacto en el patrón geométrico *Layout*, ya que cada pozo del transistor PMOS se polariza a su respectiva fuente en este

3.7. DISMINUCIÓN DEL VOLTAJE DE ALIMENTACIÓN

caso por reglas de diseño los pozos tipo n de dos compuertas que trabajan con diferente voltaje de alimentación tendrían que estar separados lo cual penaliza mas el área, para evitar esto se proponen *Layouts* compactos en los cuales todos lo pozos se polarizan al voltaje de alimentación mas alto, esto tiene el inconveniente de incrementar el voltaje de umbral en los transistores PMOS haciendo aún más lentos los transistores, aunque este aspecto es beneficioso en términos de los efectos del NBTI, ya que al aumentar el voltaje de umbral tambien se disminuye el impacto del NBTI. Sin embargo hacer lento el circuito no es algo muy llamativo en con respecto al rendimiento, pero con una criterio de selección de compuertas adecuado se puede minimizar esta perdida en el retardo.

Capítulo 4

Estimación de Retardo, Consumo de Potencia y Temperatura del circuito considerando efectos de NBTI

Como se mencionó en el capítulo 2, la temperatura es uno de los factores que más impacta en el efecto del NBTI, por lo cual tomar una alta temperatura y constante de operación significa tomar un peor caso con respecto a degradación, por lo cual al realizar una estimación de la temperatura se puede obtener una aproximación más exacta de este fenómeno. La temperatura está muy relacionada con la disipación de potencia y se han desarrollado varias herramientas que calculan estas temperaturas con una gran cantidad de consideraciones e interacciones. En las primeras secciones de este capítulo se explicará cómo se calcula la potencia disipada por un circuito y a partir de ésta los modelos que se aplican para determinar la temperatura a la cual opera. En la siguiente sección se explicará el procedimiento para obtener el retardo de los caminos lógicos de un circuito considerando efectos temperatura, tensión de alimentación y envejecimiento por NBTI

de forma conjunta. Por último se explicará la relación entre consumo de potencia, temperatura y retardo, y cómo esto ayuda a obtener un valor del impacto NBTI mas preciso.

4.1. Modelo de retardo de compuerta

Generar los modelos de retardo de compuerta es considerado el primer paso en el análisis ya que de éstos representan las *celdas de compuerta* que serán utilizadas en el *Static Timing Analysis* (STA). En un STA convencional, cada celda es generada mediante la caracterización del retardo de cada compuerta como función del nodo de entrada y de su capacitancia de fan-out, pero en este trabajo se tiene en cuenta la temperatura y la variación de voltaje de umbral, esta última sera utilizada para calcular el impacto del NBTI en el retardo

El retardo de una compuerta puede ser representado como una combinación lineal de aquellos parámetros de los cuales que depende (ver ecuación 4.1). Este modelo es llamado *modelo lineal de retardo de compuerta*.

$$D = D_0 + \sum_{k=1} S_k \cdot X_k \quad (4.1)$$

Donde D_0 representa el retardo de la compuerta bajo condiciones nominales, S_i es el coeficiente de sensibilidad del retardo a variaciones en en el parámetro k y X_k es la magnitud de dicha variación del parámetro.

Típicamente los parámetros cuya variación (X_k) presenta un mayor impacto (mayor sensibilidad) en el retardo de compuerta son el ancho de canal (W), la longitud de canal (L), el espesor del óxido (T_{ox}) y el voltaje umbral (V_{th}). En este trabajo se centra en los cambios de voltaje de umbral los cuales se deben a NBTI, por lo cual solo se calcula la sensibilidad de este parámetro.

La caracterización se llevó a cabo mediante simulaciones en HSPICE. La figura 4.1 ilustra un esquemático del procedimiento de simulación realizado para

caracterizar el retardo nominal de cada compuerta. Para cada nodo de entrada de la compuerta bajo caracterización, se aplica un pulso de entrada cuyo slew rate (SR_{in}) se varía dentro de un rango determinado y su amplitud toma dos valores: un voltaje bajo (V_{DD}) y otro voltaje alto (V_{DDH}). De la misma forma se varía la alimentación de la compuerta lo cual emulará los voltajes de alimentación del Dual-VDD. En el nodo de salida de la compuerta se introduce una capacitancia variable que emulará cada uno de los posibles valores de capacitancia de fan-out vistos por la compuerta. Por último se varía la temperatura a la cual opera la compuerta.

Para cada combinación de SR_{in} , V_{in} , V_{DD} , CL y temperatura se obtiene el valor correspondiente de retardo de propagación de alto a bajo (t_{phl}) y de bajo a alto (t_{plh}). De esta forma el retardo de propagación de la compuerta estará en función de las cinco variables antes mencionadas. Igualmente se determinan los respectivos slew rates y tensiones a la salida (SR_{ohl} , SR_{oh} , V_{out}) ya que éstos corresponderán a los SR y tensiones a la entrada con los que se deberá evaluar el retardo de las compuertas subsiguientes en el circuito, cuando se realice el STA [?]. Debe mencionarse que mientras se caracteriza el retardo de un nodo de entrada determinado, los demás se polarizan al valor de tensión que sensibilice a la compuerta por el nodo de entrada considerado (V_{DD} para las compuertas tipo NAND y AND, GND para las compuertas tipo NOR y OR, etc.).

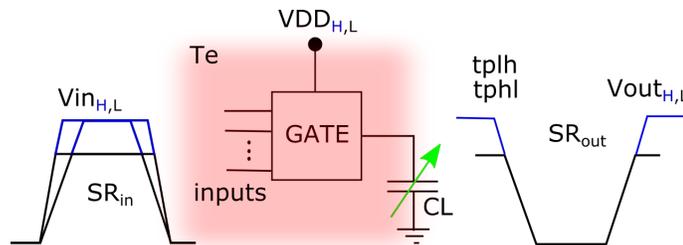
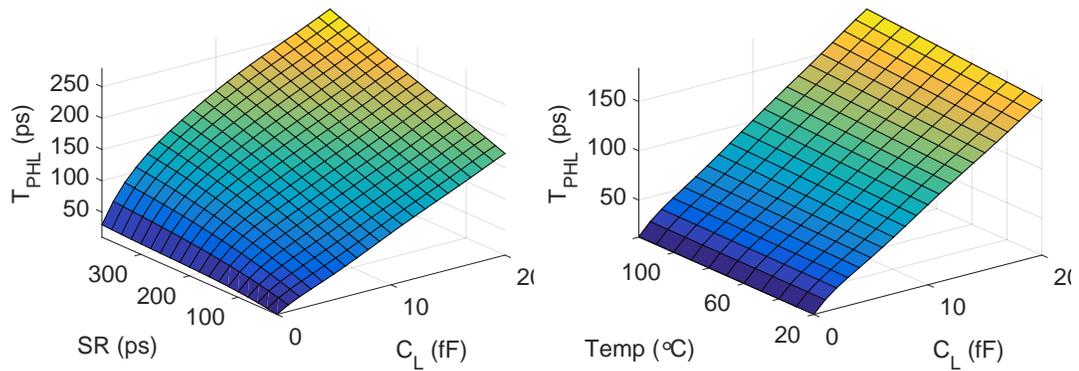


Figura 4.1: Caracterización del retardo de propagación de una compuerta digital

La figura 4.2 muestra un ejemplo de la matriz de datos obtenida para el tiempo de propagación de alto a bajo de una compuerta inversora en función de CL y

SR in para una VDD y temperatura constantes. Para reducir el espacio de memoria que requiere almacenar cada uno de los puntos caracterizados en la herramienta de cómputo y también poder utilizar valores continuos de capacitancia de carga y SR a la entrada (la caracterización se realiza para valores discretos), la matriz obtenida es aproximada mediante un polinomio de tres variables.

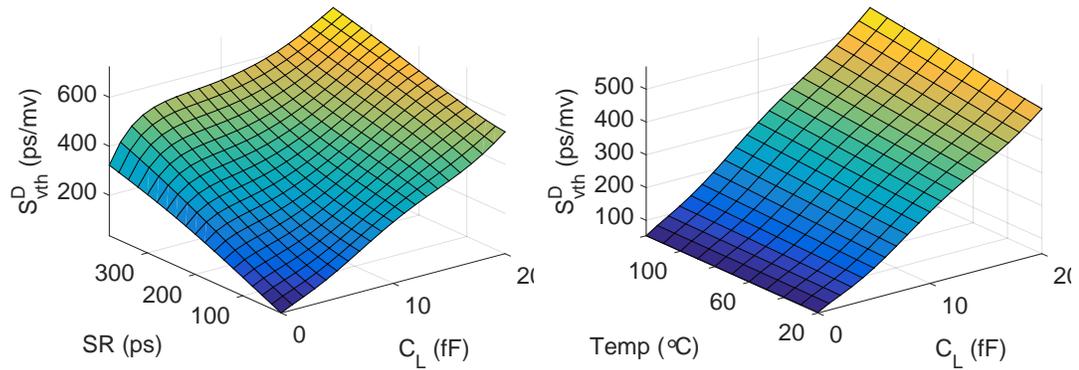


(a) Matriz de T_{PHL} vs SR y C_L en un inversor (b) Matriz de T_{PHL} vs C_L y Temperatura en un inversor con Slew Rate constante

Figura 4.2: Retardo de un inversor en función de multiples variables

Para el voltaje de umbral, ya que cada transistor puede presentar un valor diferente de envejecimiento, la sensibilidad en el retardo de compuerta fue caracterizada para cada transistor por separado obteniendo un valor correspondiente para cada uno de los transistores presentes en la compuerta. Al igual que para el retardo, cada una de las matrices de caracterización obtenidas (ver figura 4.3) fue aproximada por un polinomio de tres variables y la sensibilidad es calculada como la derivada de este polinomio respecto al parámetro correspondiente evaluada en el valor del parámetro nominal. Ya que cada transistor puede presentar un valor diferente tanto de envejecimiento, la sensibilidad en el retardo de compuerta fue caracterizada para cada transistor por separado obteniendo un valor correspondiente a cada uno de los transistores presentes en la compuerta.

Una vez realizada esta caracterización, el circuito puede ser analizado a nivel de compuerta, donde cada una está descrita por su retardo nominal y sus respectivas



(a) Matriz de S_{vth}^D vs SR y C_L en un inversor (b) Matriz de S_{vth}^D vs C_L y Temperatura en un inversor con Slew Rate constante con temperatura constante

Figura 4.3: Sensibilidad del retardo a cambio de voltaje de umbral de un inversor en función de multiples variables

sensibilidades a cada uno de los parámetros.

4.2. Calculo de potencia

La potencia disipada por un circuito digital se debe a diferentes fenómenos y su valor esta cambiando continuamente debido a las diferentes cargas de trabajo que maneja y el estado del circuito (standby, alto rendimiento, ahorro de batería, etc.), por lo cual los cálculos de potencia no son deterministas y solo se puede hacer una estimación. En la figura 4.4 se observa las clases de potencia que disipa un circuito digital.

4.2.1. Potencia Dinámica debida a Conmutación en Capacitancias

En todas las compuertas al cargar la capacitancia de carga C_L a través de la red PMOS, el voltaje del capacitor pasa de 0V a VDD, esto significa que de la fuente de alimentación entrega una cantidad de energía. Parte de esta energía es

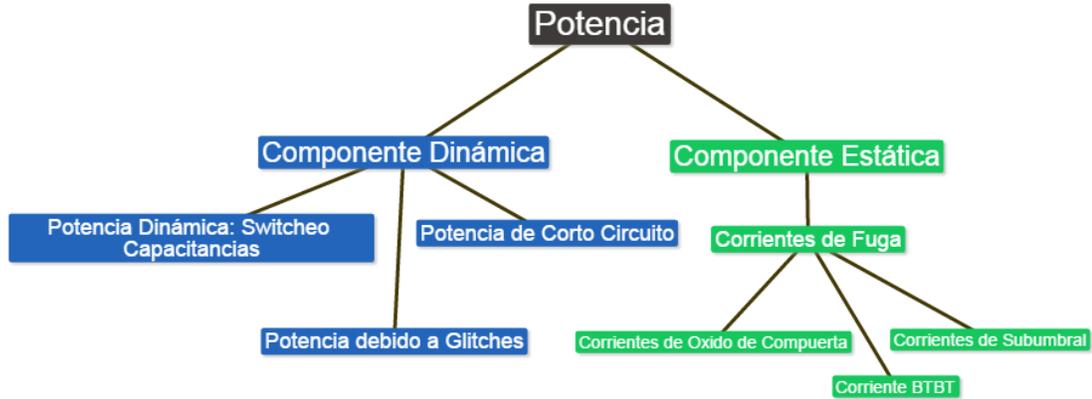


Figura 4.4: Diagrama de Potencias Disipadas por un Circuito Digital

disipada por la red PMOS durante la misma carga la otra parte queda almacenada en la capacitancia de carga C_L . Durante la transición de alto a bajo el capacitor se descarga y la energía que almacenaba es disipada por la red NMOS. Esta cantidad de energía consumida puede ser calculada de manera precisa, suponga un inversor con carga C_L , en la figura 4.5 observamos el voltaje de entrada en un periodo T el cual el inversor realiza dos transiciones de alto a bajo y otra de bajo a alto, también observamos el voltaje en el capacitor y la corriente de la fuente.

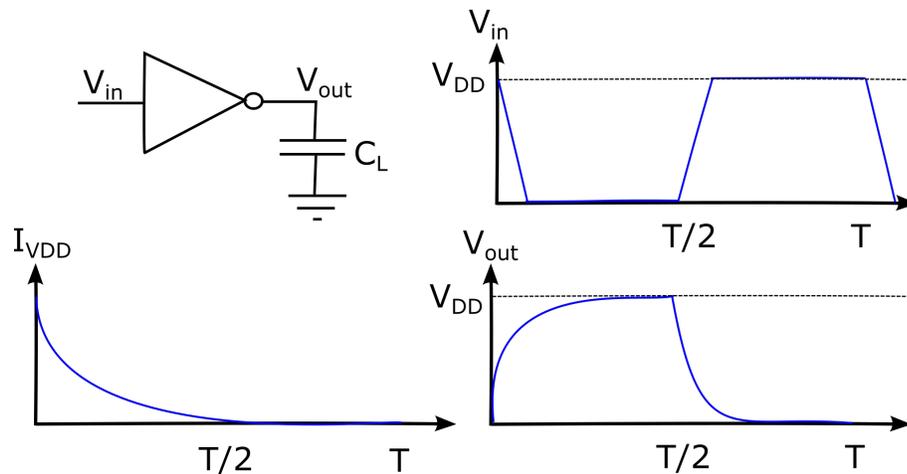


Figura 4.5: Corrientes y Voltajes de un inversor con carga C_L

La energía entregada por la fuente en un periodo T es igual a:

$$E_{VDD} = \int_0^T V_{DD} i_{VDD}(t) dt \quad (4.2)$$

La corriente de la fuente es igual a la corriente del capacitor cuando se carga la cual es va de 0 a T/2 y es igual a:

$$i_{C_L} = C_L \frac{dv_{out}}{dt} \quad (4.3)$$

Por lo tanto tenemos que la energía entregada por la fuente es igual a

$$\begin{aligned} E_{VDD} &= \int_0^{T/2} V_{DD} C_L \frac{dv_{out}}{dt} dt \\ &= V_{DD} C_L \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2 \end{aligned} \quad (4.4)$$

De igual forma la energía almacenada en el capacitor es igual a

$$\begin{aligned} E_C &= \int_0^{T/2} v_{out} i_{VDD}(t) dt = \int_0^{T/2} v_{out} C_L \frac{dv_{out}}{dt} dt \\ &= C_L \int_0^{V_{DD}} v_{out} dv_{out} = \frac{C_L V_{DD}^2}{2} \end{aligned} \quad (4.5)$$

Por lo tanto energía disipada por el PMOS es igual a la energía entregada por la fuente menos la energía almacenada en el capacitor es decir $C_L V_{DD}^2/2$, en la siguiente transición de bajo a alto el NMOS descarga la energía que queda en el C_L y la disipa, por lo tanto la energía disipada por el inversor en la carga y descarga de C_L es igual a $C_L V_{DD}^2$

La potencia disipada es este periodo de tiempo se denomina potencia dinámica la cual es igual a:

$$P_{dyn} = \frac{E_{inv}}{T} = E_{inv}f_{clk} = C_L V_{DD}^2 f_{clk} \quad (4.6)$$

Cuando una compuerta esta en un circuito su conmutación de 0 a 1 depende de la probabilidad de la señal de su entrada, por lo tanto el anterior factor se debe multiplicar por un factor denominado factor de actividad $\alpha_{0 \rightarrow 1}$

$$P_{dyn} = \alpha_{0 \rightarrow 1} C_L V_{DD}^2 f_{clk} \quad (4.7)$$

Este factor es igual a la probabilidad de que la señal pase 0 a 1 es decir la probabilidad de que este en cero por la probabilidad que este en 1, en la tabla 4.1 se muestra el valor de factores de actividad para diferentes compuertas

Tabla 4.1: Factores de Actividad de Compuertas

Compuerta	factor de actividad ($\alpha_{0 \rightarrow 1}$)
INV	$sp(1 - sp)$
NAND2	$(1 - sp_1 sp_2)(sp_1 sp_2)$
NAND3	$(1 - sp_1 sp_2 sp_3)(sp_1 sp_2 sp_3)$
NAND4	$(1 - sp_1 sp_2 sp_3 sp_4)(sp_1 sp_2 sp_3 sp_4)$
NOR2	$(1 - sp_1)(1 - sp_2)[1 - (1 - sp_1)(1 - sp_2)]$
NOR3	$(1 - sp_1)(1 - sp_2)(1 - sp_3)[1 - (1 - sp_1)(1 - sp_2)(1 - sp_3)]$
NOR4	$(1 - sp_1)(1 - sp_2)(1 - sp_3)(1 - sp_4)[1 - (1 - sp_1)(1 - sp_2)(1 - sp_3)(1 - sp_4)]$

Para un circuito la potencia dinámica sera igual a la sumatoria de las potencias dinámicas de cada compuerta y la frecuencia esta determinada por el retardo del camino lógico mas crítico, ya que este determina el valor de la frecuencia de reloj, en esta fórmula se agrega el VDD de cada compuerta ya que esto sera útil para circuitos con varias tensiones de trabajo.

$$P_{dyn_{total}} = \sum_{i=1}^N (\alpha_{0 \rightarrow 1} C_{L_i} V_{DD_i}^2) \frac{1}{D_{crit-path}} \quad (4.8)$$

4.2.2. Potencia de Corto circuito

Se presenta cuando las transiciones de entrada superan el voltaje de umbral de un transistor (red PMOS o NMOS) pero aun la red opuesta (red NMOS o PMOS respectivamente) no han bajado del voltaje de umbral, por lo cual las dos redes están encendidas y hay un camino directo entre la fuente de alimentación y la tierra, esto se ilustra en la figura 4.6. Esta corriente es limitada por la capacidad de corriente de los transistores es decir que depende de las dimensiones de los transistores.

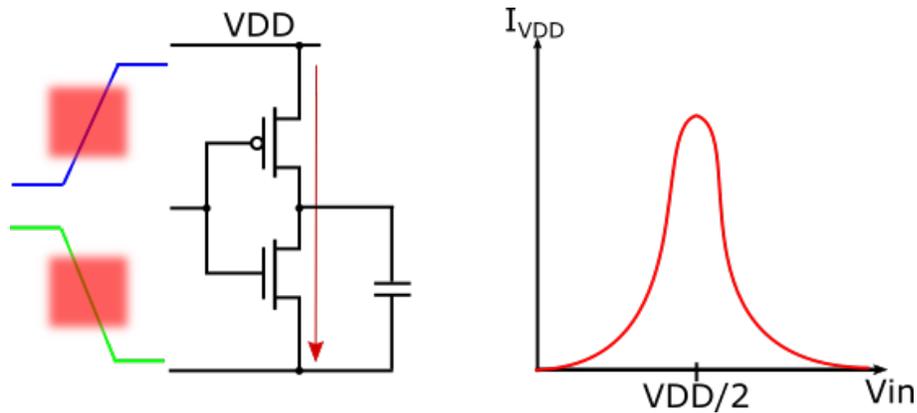


Figura 4.6: Diagrama de la corriente de corto circuito en un Inversor

4.2.3. Potencia debido a corrientes de fuga

Las corrientes de fuga se presentan cuando las entradas de las compuertas están en un valor estable y se deben a varios fenómenos en el transistor [44]. Estas tienen la característica de tener un incremento exponencial con la temperatura y son dependientes del dimensionamiento de los transistores y el valor de la tensión de alimentación; actualmente tienen un impacto muy alto en tecnologías nanométricas debido a que los efectos de tunelamiento son mas evidentes a estas escalas. Las tres causas de las corrientes de fuga son:

1. **Corrientes de Subumbral:** Es la corriente presente un transistor entre

fuente y drenaje cuando este se encuentra apagado. En este trabajo es de gran impacto debido a que no se utilizaran convertidores de nivel para que una compuerta con baja tensión de alimentación maneje una de alta tensión de alimentación y sera la dominante en el total de corriente de fuga. En el modelo BSIM esta corriente es descrita por la ecuación 4.10.

$$I_{sub} = \mu_0 \frac{W_{eff}}{L_{eff}} v_T^2 \sqrt{\frac{q\varepsilon_s N_{cheff}}{2\Phi_s}} (1 - e^{-\frac{V_{ds}}{v_T}}) \cdot \exp\left[\frac{(V_{gs} - v_T - \gamma V_{sub} + \eta V_{ds})}{nv_T}\right] \quad (4.9)$$

Donde μ_0 es la movilidad de los portadores, W_{eff} y L_{eff} son el ancho y largo efectivos del canal del transistor, v_T es el voltaje térmico, ε_s es la permitividad del silicio, N_{cheff} es el dopaje efectivo del canal y Φ_s es el potencial de superficie. V_{gs} , V_{ds} , V_{sb} , V_t son los voltajes compuerta-fuente, drenaje-fuente, fuente-cuerpo y de umbral respectivamente, γ es el coeficiente del efecto cuerpo linealizado y η es el coeficiente del fenómeno Drain Induced Barried Lowering (DIBL) y n es el coeficiente de balanceo de subumbral.

2. **Corrientes de Óxido de Compuerta:** En las actuales tecnologías la disminución del grosor del óxido de compuerta ha sido tan significativa como la disminución de la longitud de canal, como resultado, esta corriente se ha incrementado mucho mas rápido que las corrientes de subumbral. Algunos estudios sugieren que una diferencia en el T_{ox} de 2 Angstroms incrementa el valor de esta corriente en un orden de magnitud.
3. **Corrientes de BTBT:** Esta corriente llamada de polarización inversa drenaje-substrato y fuente-substrato de Banda de Unión a Banda de Tunelamiento (junction **B**and to **B**and **T**unneling), se deben a la polarización inversa de las uniones PN inherentes del transistor esto causa una corriente significativa en estas uniones debido al tunelamiento de electrones de la banda de valencia de la región-p a la banda de conducción de la región-n.

Para estimar la potencia debido a estas tres corrientes se realiza una aproximación estadística [45] debido a que su valor depende los valores de la entrada a la compuerta o vectores de entrada que es igual a:

$$I_{leak}^{avg} = \sum_{i \in V_{inputs}} P_i I_{leak_i} \quad (4.10)$$

Donde P_i es la probabilidad es la probabilidad de que el vector de entrada el cual pertenece al grupo de todos los vectores de entrada posibles V_{inputs} sea igual al vector i -esimo e I_{leak_i} es la corriente por fugas de la compuerta debido a ese vector i -esimo de entrada. Las anteriores corrientes se estimaran por medio de HSpice y se evaluaran para diferentes temperaturas, con esto se obtiene un polinomio en función de la temperatura para cada vector de entrada, lo cual permite calcular un valor mas preciso de estas corrientes en función de la temperatura de la compuerta y la probabilidad del vector de entrada, por lo tanto la anterior ecuación queda como:

$$I_{leak}^{avg}(Temp) = \sum_{i \in V_{input}} P_i I_{leak_i}(Temp) \quad (4.11)$$

4.2.4. Potencia debido a Pulsos Espurios

Los Pulsos Espurios son valores a la salida de una compuerta que temporalmente no corresponden al valor estado estable de la salida , esto se debe a que las entradas de una compuerta no cambian al mismo tiempo generando una transición no deseada que disipa potencia al cargar y descargar la capacitancia de carga de dicha compuerta y de las compuertas subsiguientes que sean sensibles a esta transición. En la figura 4.7 se ilustra con una AND2 que cambia sus entradas A y B de "01" → "10" respectivamente, el valor lógico para las dos combinaciones a la salida es "0" pero si la señal B llega con un retraso la entradas valdrán "11" generando un "1" a la salida.

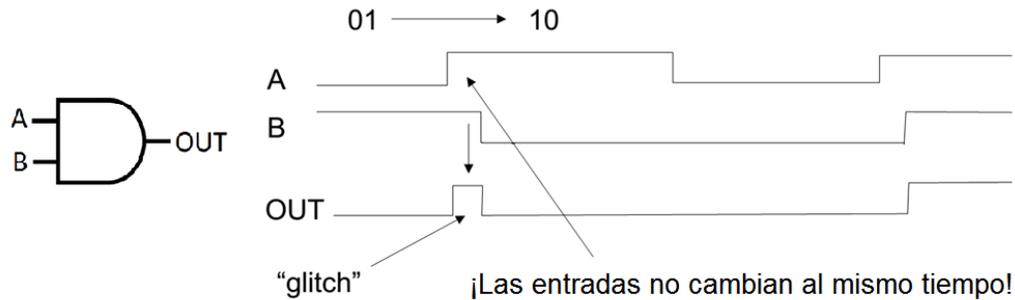


Figura 4.7: Pulsos Espurios en una compuerta AND2

4.3. Cálculo de Temperatura

La estimación de temperatura en circuitos integrados es en este momento una parte esencial del flujo de diseño, ya que define algunas características para una apropiada operación del circuito, como son el empaquetado, si necesita disipador de calor y el tamaño del mismo. Todas las herramientas actuales [46] calculan la temperatura del circuito en función de la potencia que disipa generando un perfil de temperatura con el cual se puede predecir qué zonas del circuito van a generar más calor, esto es útil para determinar el efecto del NBTI de manera mas precisa sin tener que prever un peor caso de temperatura y de esta manera atacarlo sin necesidad de sobrecostos en área o retardo.

En el modelo de transferencia de calor por conducción básico se tienen dos cuerpos a diferentes temperaturas y un sólido que los une de longitud x y sección transversal de área A por el medio del cual hay una transferencia de energía en forma de calor en un determinado tiempo (esto es potencia), lo cual se ilustra en la figura 4.8 y esta determinado por la ecuación de la ley de Fourier:

$$\frac{\Delta Q}{\Delta t} = \frac{kA}{x}(T_1 - T_2) \quad (4.12)$$

Donde k es la conductividad termal del material que compone al sólido, la cual viene expresada en $W/m^{\circ}C$. A este modelo de conducción se puede hacer su analogía eléctrica, en la cual el flujo de calor o potencia corresponde a la corriente,

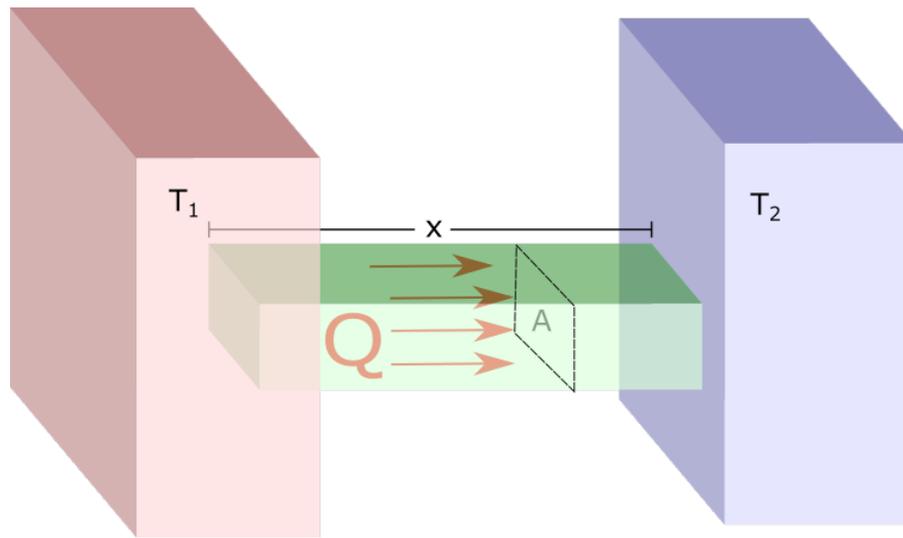


Figura 4.8: Conducción de calor a través de un sólido

las temperaturas a voltajes y el término x/kA es análogo a una resistencia R_{th} denominada resistencia térmica la cual tiene magnitudes $^{\circ}C/W$, por lo cual la ecuación 4.15 podemos reescribirla como:

$$P_Q = \frac{1}{R_{th}}(T_1 - T_2) \quad (4.13)$$

El circuito que corresponde a la ecuación anterior se muestra en la figura 4.9, como podemos ver la resistencia térmica de un cuerpo es proporcional a la longitud e inversamente proporcional al área.

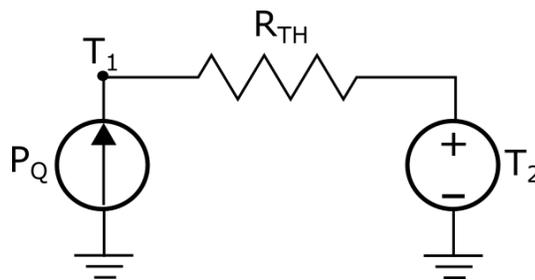


Figura 4.9: Circuito Equivalente de la conducción de calor a través de un sólido

El modelo anterior es aplicable a circuitos integrados digitales en los cuales

queremos saber la temperatura de una compuerta la cual disipa una potencia determinada, en este caso se asume que el circuito o chip esta compuesto del silicio, un difusor y un disipador de calor como se muestra en la figura 4.10, se asume que cada compuerta tiene un área de silicio determinada que es igual al área total del circuito sobre el número de compuertas $A_{gate} = A_{circuito}/N$, como el silicio es poco eficiente al transmitir el calor, la temperatura de una compuerta no interactúa con la temperatura de las compuertas cercanas, en cambio tanto el difusor y como el disipador son buenos conductores de calor y expelen de manera eficiente el calor generado por el circuito, por lo tanto el área donde disipan la potencia es el área total del circuito $A_{circuito}$. En la tabla 4.2 se muestra el valor de cada una de los conductividades térmicas de las capas del circuito, y la ecuación que se utiliza para calcular su respectiva resistencia térmica, donde t_{si} , t_{dif} y t_{dis} son los espesores de cada capa. En este trabajo esto valores son variables debido a que cada circuito tiene un área y disipación de potencia diferente.

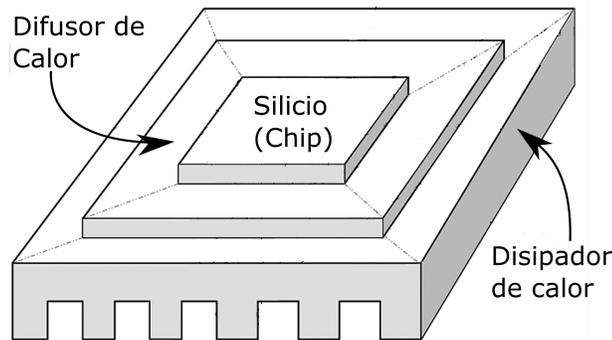


Figura 4.10: representación de los materiales que componen un CI

La temperatura de cada compuerta se puede calcular por medio del circuito que se observa en la figura 4.11 y es igual a:

$$T_{gate} = P_{gate}Rth_{chip-difusor} + P_{total}(Rth_{disipador-ambiente} + Rth_{difusor-disipador}) + T_{amb} \quad (4.14)$$

4.4. ALGORITMO ITERATIVO PARA CÁLCULO DE RETARDO, POTENCIA, TEMPERATURA E IMPACTO DE NBTI

Capa del CI	k ($W/m^{\circ}C$)	R_{th}
Silicio (Chip)	100	$\frac{t_{si}}{kA_{gate}}$
Difusor	400	$\frac{t_{dif}}{kA_{circuito}}$
Disipador	400	$\frac{t_{dis}}{kA_{circuito}}$

Tabla 4.2: Conductividad Térmica y Resistencia Térmica respectiva

Donde P_{gate} es la potencia que disipa cada compuerta y es igual a la suma de la potencia dinámica mas la potencia por fugas y P_{total} es la potencia disipada por todo el circuito, es decir la suma de todas las potencias disipadas de todas las compuertas del circuito. En la figura 4.11 se observa que la potencia total equivale a la suma de todas las fuentes de corriente y esta corriente total pasa por la resistencia térmica del empaque y el disipador.

$$\begin{aligned}
 P_{gate} &= P_{dinamica,gate} + P_{fugas,gate} \\
 P_{total} &= \sum_{i=1}^N P_{gate_i} = \sum_{i=1}^N (P_{dinamica,gate_i} + P_{fugas,gate_i}) \quad (4.15)
 \end{aligned}$$

4.4. Algoritmo iterativo para cálculo de retardo, potencia, temperatura e impacto de NBTI

Como se ha observado en las anteriores secciones el retardo de un circuito define la potencia que disipa, esta a su vez define la temperatura al cual trabajan las compuertas de dicho circuito, y la temperatura afecta tanto al retardo como la potencia estática que consume el circuito. Cuando se trabaja con varias fuentes

4.4. ALGORITMO ITERATIVO PARA CÁLCULO DE RETARDO, POTENCIA, TEMPERATURA E IMPACTO DE NBTI

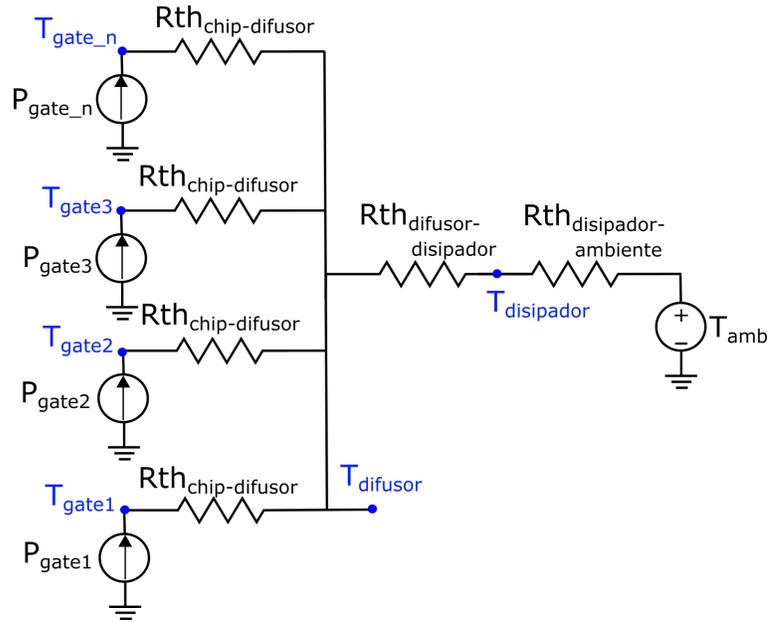


Figura 4.11: Circuito equivalente para cálculo de temperatura de un CI

de alimentación se afectan las tres anteriores características del circuito, las cuales impactan el efecto del NBTI, pero también hay que tener en cuenta que el voltaje de alimentación influye sobre el efecto del NBTI directamente. Estas interacciones se ilustran en la figura 4.12.

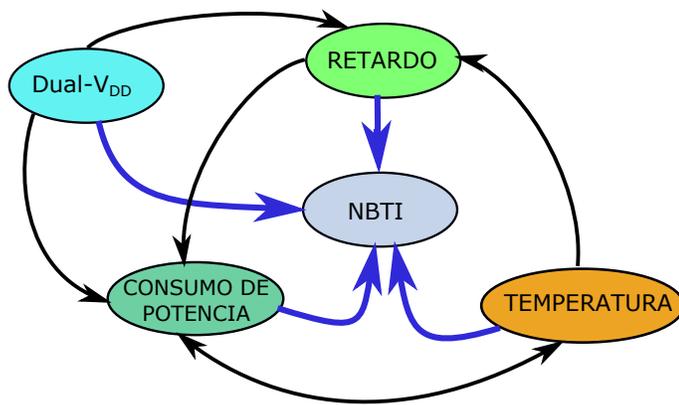


Figura 4.12: Interacciones presentes en un circuito digital Dual-VDD

Debido a estas interacciones la metodología que se propone en el capítulo 5 cuenta con un algoritmo iterativo con el cual se obtiene una temperatura de

4.4. ALGORITMO ITERATIVO PARA CÁLCULO DE RETARDO, POTENCIA, TEMPERATURA E IMPACTO DE NBTI

trabajo de las compuertas, este algoritmo se ilustra en la figura 4.13, primero se asume que el circuito opera a una temperatura inicial con la cual se procede a calcular el retardo, el retardo degradado y potencia estática a esta temperatura, como segundo paso con el retardo del LCP se calcula el consumo de potencia dinámica del circuito, en el tercer paso con la potencia total consumida por el circuito y cada compuerta se calcula las temperaturas de cada compuerta. este ciclo es repetido hasta que las temperaturas de las compuertas converjan a un valor con este valor de temperaturas que trabajan las compuertas se obtiene impacto de NBTI no tan pesimista y más cercano a la realidad.

4.4. ALGORITMO ITERATIVO PARA CÁLCULO DE RETARDO, POTENCIA, TEMPERATURA E IMPACTO DE NBTI

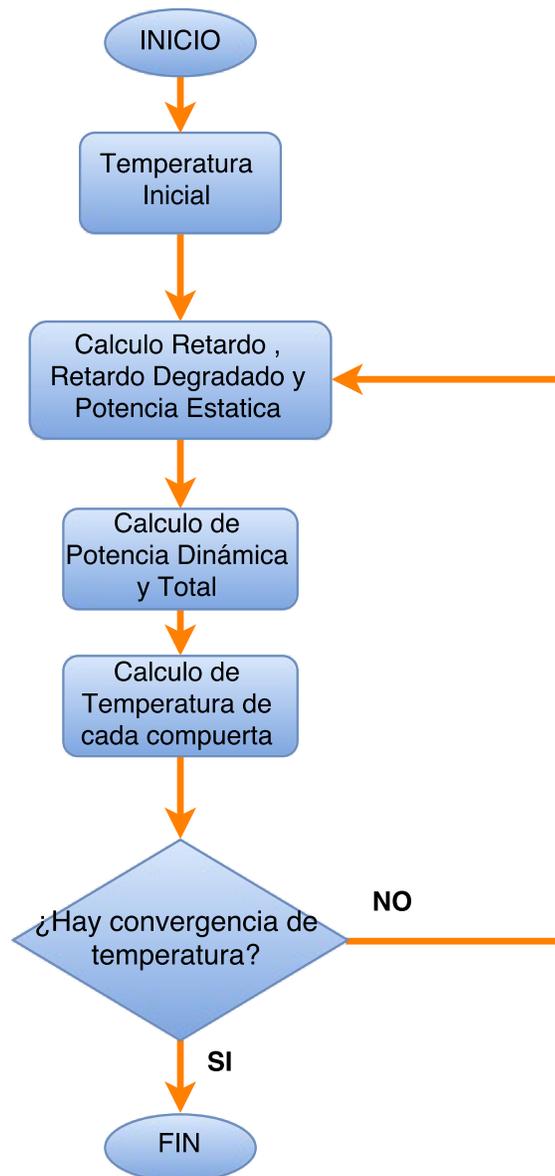


Figura 4.13: Algoritmo iterativo para cálculo de retardo, potencia y temperatura

Capítulo 5

Metodología de robustez a envejecimiento basado en Dual- V_{DD}

Al utilizar dos fuentes de voltaje (Dual- V_{DD}) en circuitos digitales, se busca reducir el voltaje a las compuertas que de tal manera que se impacte poco el retardo del circuito pero que al mismo tiempo haya una reducción significativa de la potencia.

Existen dos formas de organización de la compuertas con Dual- V_{DD} los cuales son:

1. Escalamiento de voltaje agrupado (CVS) [41] en el cual las compuertas con bajo voltaje no pueden operar a las de alto voltaje debido a las corrientes estáticas altas que se generarían, esta restricción disminuye la cantidad de compuertas a seleccionar, esto se ilustra en la figura 5.1.
2. Escalamiento de voltaje agrupado Extendido (ECVS) [42] utiliza convertidores de nivel los cuales permiten evadir la restricción anteriormente mencionada como se observa en la figura 5.2. Estos convertidores de nivel [43] penalizan el área y pueden llegar penalizar el retardo del circuito.

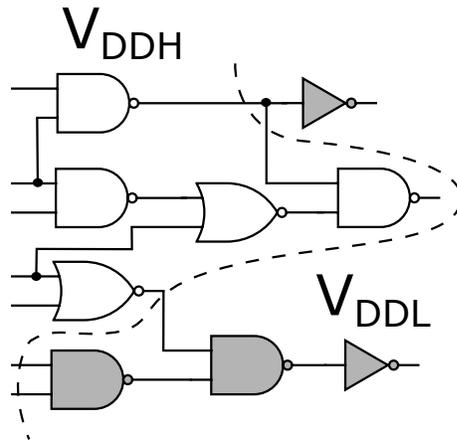


Figura 5.1: Ejemplo de Escalamiento de voltaje agrupado (CVS)

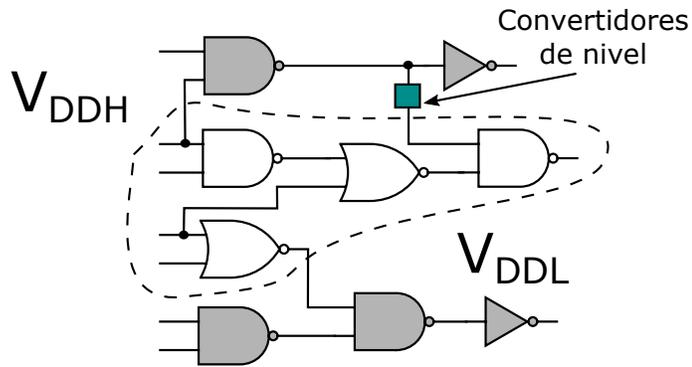


Figura 5.2: Ejemplo de Escalamiento de voltaje agrupado Extendido (ECVS)

Sin embargo, el operar con voltajes de alimentación no tan alejados en magnitud en ECVS es posible no insertar convertidores de nivel para no penalizar el área, con un incremento mínimo en la potencia estática debida a corrientes de subumbral que suceden cuando una compuerta alimentada con voltaje bajo maneja una compuerta alimentada por voltaje alto.

Con esta técnica ocurren fenómenos interesantes con respecto a la degradación del retardo y consumo de potencia que son estudiados y explicados en la primera sección de este capítulo; lo anterior se realiza para poder utilizar esta técnica de la manera mas eficaz posible para mitigar los efectos del NBTI a 10 años.

En la segunda sección se explica la metodología de selección de compuertas

y el algoritmo utilizado para realizar este procedimiento, en la tercera sección se explica la metodología de selección de compuertas candidatas a trabajar con bajo voltaje y por ultimo en la sección final se muestran y analizan los resultados de esta técnica aplicada a varios circuitos ISCAS.

5.1. Consideraciones al utilizar Dual- V_{DD}

Al utilizar la técnica Dual-VDD se reduce el voltaje de alimentación de algunas compuertas, esto implica varias consideraciones además de la disminución del impacto del NBTI y de la potencia, también afecta el retardo y por último en el layout las cuales se explicaran en esta sección.

5.1.1. Impacto en el Retardo Total

En la sección 2 se menciona que el retardo total degradado en el *modelo lineal de retardo de compuerta* esta conformado por dos elementos que son el retardo fresco, es decir el retardo del circuito en el tiempo igual a 0 y la degradación del retardo, este último elemento esta compuesto por dos factores que son la sensibilidad del retardo a variaciones del voltaje de umbral y la degradación del voltaje de umbral debido a NBTI. Estos tres elementos son función del voltaje de alimentación y de la temperatura, por lo cual la ecuación del retardo total la podemos reescribir como la ecuación 5.1. Donde $S_{D,Vth,k}$ es la sensibilidad del retardo k-esimo transistor a cambios de voltaje de umbral y $\Delta Vth_{k,nbti}$ es el incremento del voltaje de umbral a del k-esimo transistor debido a NBTI.

$$\begin{aligned}
 D_{aged} &= D_0(V_{DD}, T) + \Delta D_{aged}(V_{DD}, T) \\
 &= D_0(V_{DD}, T) + \sum_{k=1}^K S_{D,Vth,k}(V_{DD}, T) \Delta Vth_{k,nbti}(V_{DD}, T)
 \end{aligned}
 \tag{5.1}$$

La principal implicación de reducir el voltaje de alimentación en una compuerta digital es que su retardo fresco (D_0) aumenta, esto se debe a que las corrientes de los transistores disminuyen y por lo tanto los tiempos de carga y descarga de las capacitancias de carga se hace mas largo. En la figura 5.3 se ilustra como el retardo fresco de un inversor, es decir sin efectos de envejecimiento, aumenta cuando su voltaje de alimentación disminuye, en este caso si se disminuye de $1.2V$ a $1.0V$ y $0.9V$ el retardo aumenta en un 35% y en 67% respectivamente.

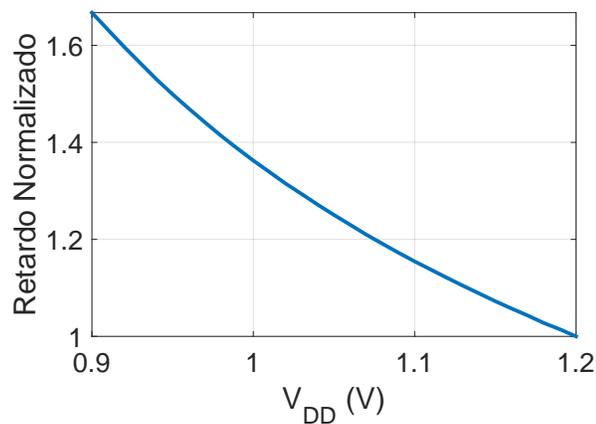
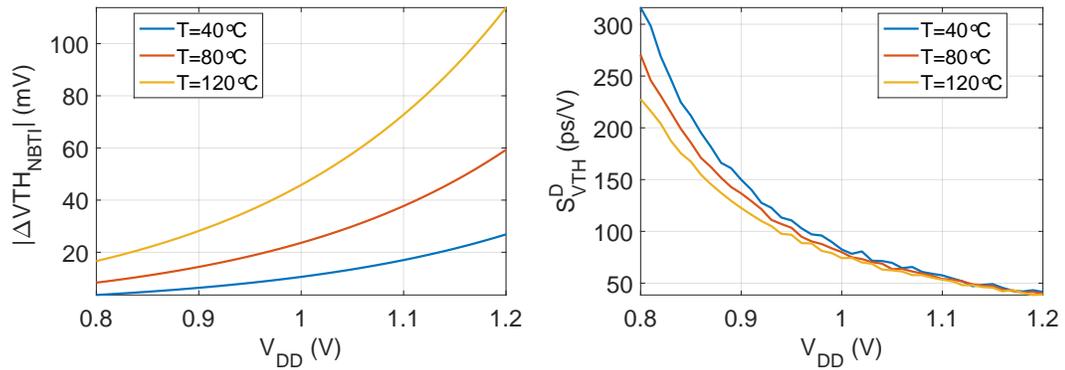


Figura 5.3: Comportamiento del retardo fresco vs el voltaje de alimentación

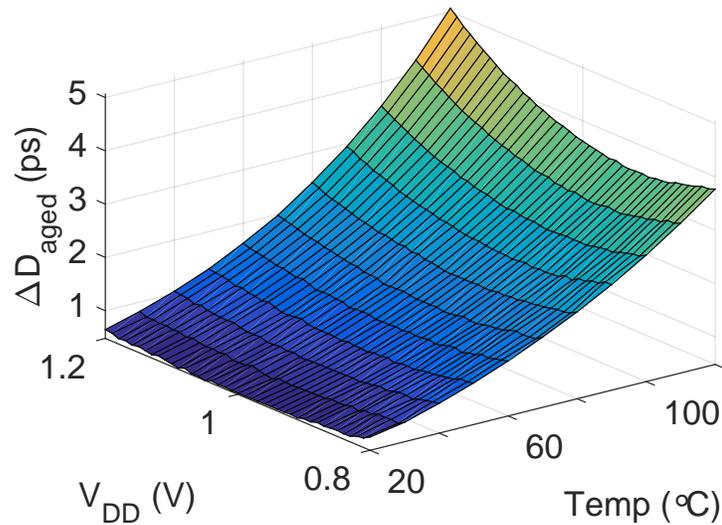
Como se mencionó en el capítulo 2 la degradación del voltaje de umbral debido a NBTI disminuye al bajar tanto el voltaje de alimentación como la temperatura de una compuerta. Al contrario pasa con la sensibilidad como se observa en la figura 5.4(b) que al disminuir el voltaje de alimentación y la temperatura, ésta tiende a incrementarse.

El comportamiento de ΔD_{age} tiende a disminuir como se observa en la figura 5.4(c) a pesar del incremento de la $S_{D,vth}$ al reducir el voltaje de alimentación y la temperatura, los efectos de la disminución de los dos factores combinados antes mencionados impactan más en el decremento de la degradación del voltaje de umbral, por ejemplo si pasamos de un voltaje de alimentación de $1.2V$ a uno de $1.0V$ y la temperatura de operación se baja de $120^{\circ}C$ a $80^{\circ}C$, la sensibilidad aumenta 2 veces, pero la degradación del voltaje de umbral disminuye 4.4 veces,



(a) $\Delta V_{th_{NBTI}}$ vs V_{DD}

(b) $S_{D,vth}$ vs V_{DD}



(c) ΔD_{aged} vs V_{DD}

Figura 5.4: Cambio de las componentes de los factores de ΔD_{aged}

por lo cual la degradación de retardo disminuye $(4.4/2) = 2.2$ veces.

5.1.2. Impacto en el Layout

Los transistores PMOS son dispositivos de cuatro terminales, Compuerta, Drenaje, Fuente y Cuerpo, normalmente la terminal del cuerpo del transistor va polarizado al voltaje de alimentación como se observa en la figura 5.5(a) donde el contacto de metal va sobre el pozo, pero al tener dos o mas voltajes de alimenta-

ción si se polariza cada PMOS a su respectiva fuente independiente, por reglas de diseño estos pozos deben estar separados una distancia mínima lo cual hace que el área del Layout se incremente como se observa en la figura 5.5(b), en el cual los transistores 2 y 3 están en serie y su polarización se redujo a un V_{DDL} .

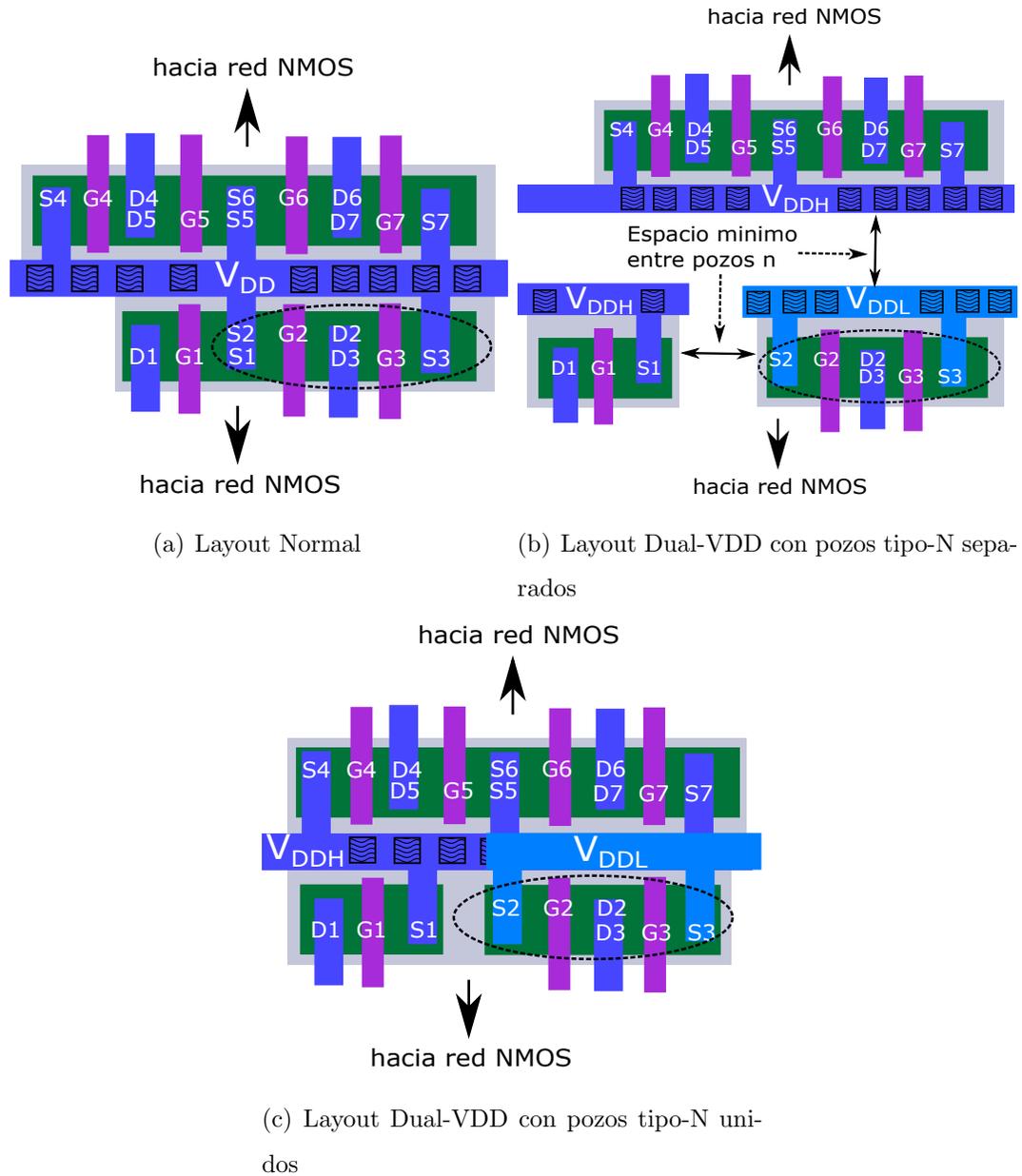


Figura 5.5: Layout transistores PMOS Dual-VDD

La solución para tener un Layout compacto es polarizar todos los pozos al

voltaje más alto, permitiendo unir los pozos de los transistores PMOS 2 y 3, y no penalizar el área como se observa en la figura 5.5(c), el costo en área que se incrementa es el de tener el transistor 1 separado del 2 y el 3 y además la interconexión extra que se debe hacer con metal mas alto, esta interconexión extra se observa con mayor claridad en la figura en 3D 5.6.

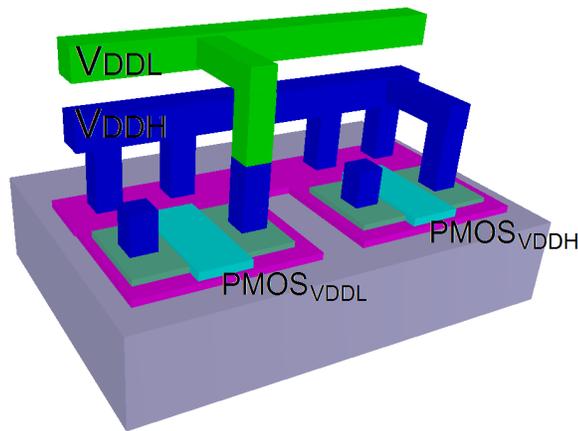


Figura 5.6: Interconexiones para PMOS para Layout Dual VDD

Al unir los pozos al voltaje de alimentación más alto, el voltaje de umbral se incrementa, este efecto se produce por la polarización inversa entre el cuerpo y la fuente, esto se explica a partir de la ecuación 5.2 en la cual el voltaje de umbral esta en función del voltaje cuerpo-fuente, si el voltaje $V_{SB} = 0$ los dos radicales son iguales y el término que multiplica al factor de cuerpo (γ) es 0 y el voltaje de umbral del PMOS es igual a V_{T0} el cual es de valor negativo, pero si $V_{BS} > 0$ se tiene que el primer radical es mayor que el segundo radical y por lo tanto el factor que multiplica al efecto cuerpo (γ) es positivo, al final la suma de los valores negativos del voltaje de umbral y el efecto cuerpo dan un valor mas alto de voltaje de umbral.

$$V_{thp} = V_{T0} - \gamma \left(\sqrt{-V_{SB} - \phi_B} - \sqrt{-\phi_B} \right) \quad (5.2)$$

En la gráfica 5.7 se muestra el incremento del voltaje de umbral en función del

voltaje de alimentación de la compuerta, la línea azul es cuando el cuerpo está polarizado al voltaje de alimentación; se observa que el voltaje de umbral apenas cambia $10mV$, pero cuando el cuerpo permanece polarizado a un voltaje de $1.2V$ el voltaje de umbral aumenta $40mV$.

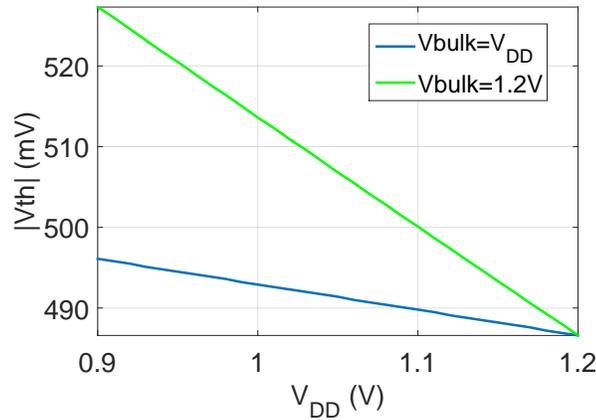


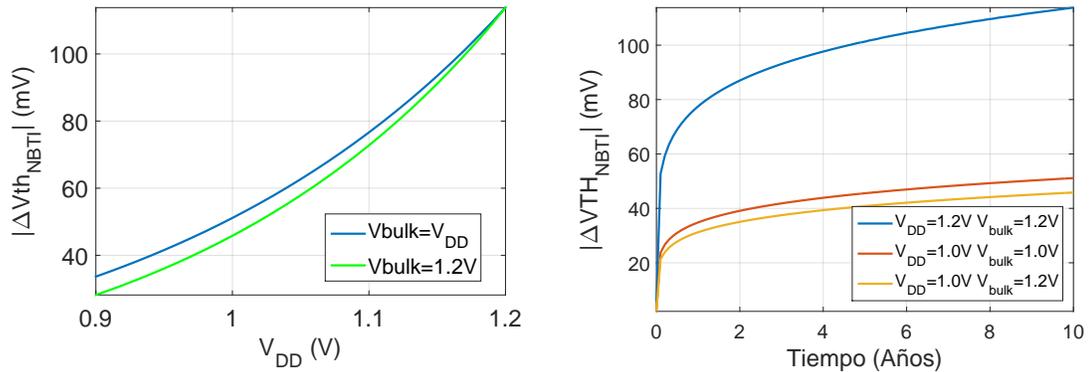
Figura 5.7: V_{th} para diferentes voltajes de VDD y V_{bulk}

En las gráficas 5.8(a) y 5.8(b) se muestra el efecto del impacto con y sin el efecto cuerpo sobre el NBTI al bajar el voltaje de alimentación de $1.2V$ a $1.0V$ visualizado en función del voltaje de alimentación y en el tiempo respectivamente, el NBTI disminuye debido a que al aumentar el voltaje de umbral el campo eléctrico en la compuerta disminuye.

Por último este incremento en el voltaje de umbral por el efecto cuerpo hace que se penalice más el retardo de una compuerta cuando se compara cuando el cuerpo está polarizado a la fuente de alimentación, en el caso de un inversor el retardo para el caso de $1.0V$ se incrementa el retardo en un 6% mientras que para $0.9V$ se incrementa en 14%, esto se ilustra en la figura 5.9.

5.1.3. Impacto en el consumo de potencia estática

Una precaución que hay que tener al utilizar la técnica Dual-VDD es que una compuerta con voltaje de alimentación bajo no puede manejar una con voltaje de



(a) Degradación de V_{th} para diferentes voltajes de V_{DD} y V_{bulk} (b) Degradación de V_{th} para diferentes voltajes de V_{DD} y V_{bulk} en función del tiempo

Figura 5.8: Impacto en la degradación del V_{th} debido al efecto cuerpo

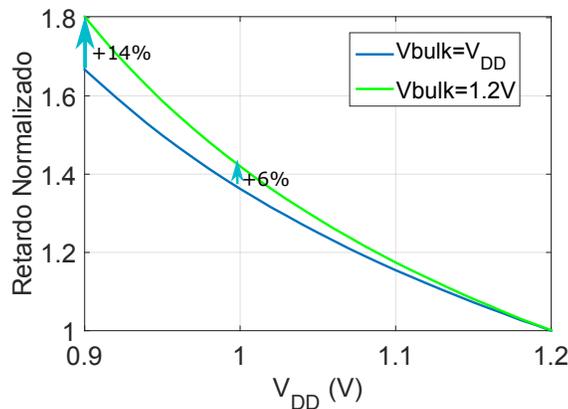


Figura 5.9: Retardo Normalizado para diferentes voltajes de V_{DD} y V_{bulk}

alimentación alto, esto se debe a que cuando la compuerta anterior este en "1" lógico, la compuerta siguiente tiene un voltaje en la entrada de sus transistores PMOS $V_{GS} = V_{DDH} - V_{DDL}$ los cuales no estarán completamente apagados y por lo cual se presenta corrientes de subumbral altas, esto se visualiza en la figura 5.10 para el caso de dos inversores.

Para evitar lo anterior en los diseños Dual-VDD se utilizan convertidores de nivel asíncronos, un ejemplo se observa en la figura 5.11, los cuales a partir de una salida de voltaje bajo arrojan una voltaje alto con implicaciones mínimas en

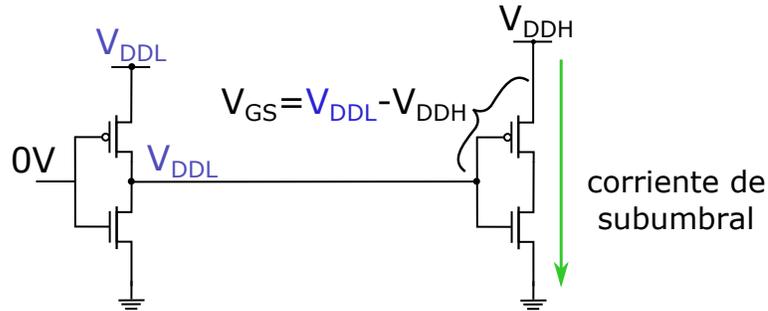


Figura 5.10: Corriente Estática que se presenta cuando una compuerta con V_{DDL} maneja una compuerta con V_{DDH}

el retardo y la potencia, sin embargo estos convertidores suponen un consumo de área significativo al agregar 7 transistores y una capacitancia extras

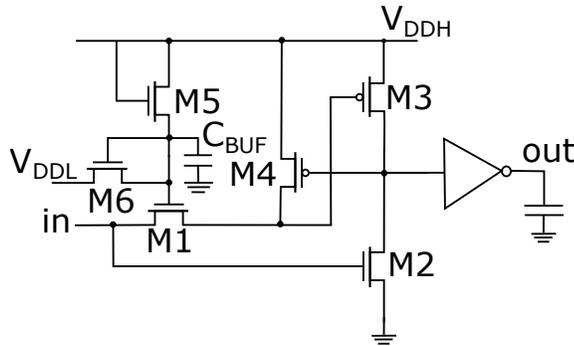


Figura 5.11: Conversor de nivel asíncrono [43]

Sin embargo en algunos trabajos [47] proponen utilizar transistores PMOS con un voltaje de umbral mas alto y así disminuir esta corriente de subumbral sin necesidad de colocar estos convertidores de voltaje. En este trabajo se propone no alejar demasiado los valores de los voltajes de alimentación bajo y alto para de esta forma minimizar el impacto de la corriente de subumbral.

En la gráfica se observa 5.12 las corrientes de subumbral de la compuerta $NAND4$ cuando todas sus entradas están "1" lógico y su voltaje de alimentación es de $1.2V$, debido a sus dimensiones de transistores NMOS y capacitancias parásitas grandes, le permite tener la mayor corriente de subumbral de todas las

compuertas en este estado. En $1.2V$ la corriente de subumbral es de $0.9nA$, mientras que para $1.0V$ la corriente de subumbral es de $92nA$, esto es un incremento de 111 veces y para $0.9V$ es de $811nA$, esto es un aumento de 972 veces.

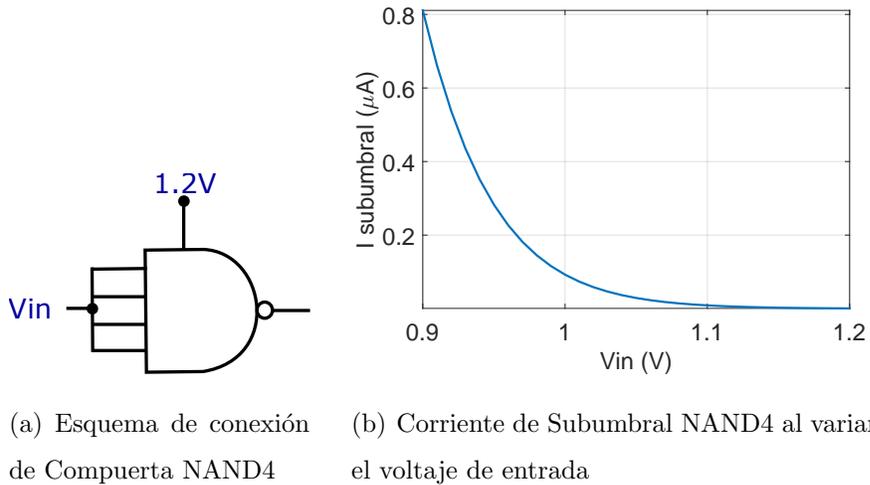


Figura 5.12: Corriente de subumbral de una compuerta NAND4

5.1.4. Selección de Valor de Voltaje de Alimentación Bajo

Como se mencionó en secciones anteriores el valor del voltaje de alimentación bajo impacta en aspectos como el retardo, el consumo de potencia tanto estático como dinámico, y la degradación del retardo por NBTI en sus dos factores S_{Vth} y el ΔVth_{NBTI} , por esto una apropiada selección de valor es importante.

En la gráfica 5.13 se ilustra la potencia de subumbral de una compuerta NAND4 en función del voltaje de entrada cuando todas sus entradas están en "1" lógico, en $0.9V$ es aproximadamente $1\mu W$, mientras que en $1.0V$ la potencia es del orden de décimas de μW . Normalmente las compuertas manejan potencias dinámicas del mismo orden el de la potencia de subumbral presente a $0.9V$ por lo cual al utilizar esta voltaje de alimentación no se ganaría en potencia consumida,

Con respecto al retardo como se observó en la sección 5.1.1 que para un inversor y las demás compuertas el retardo de aumenta en un 80% al disminuir el voltaje

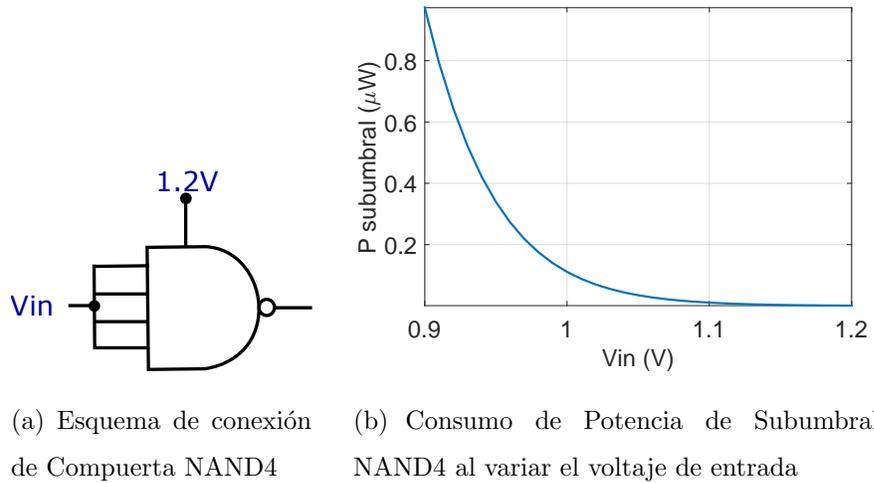


Figura 5.13: Potencia de subumbral de una compuerta NAND4

de alimentación a $0.9V$, mientras que para $1.0V$ el retardo se incrementa en un 35 %.

Por las anteriores razones las cuales son: bajo incremento en la potencia de subumbral y en el retardo por compuerta en este trabajo se utiliza $1.0V$ como voltaje de alimentación bajo para utilizar en Dual-VDD.

5.2. Metodología de selección de compuertas

En este trabajo se propone una metodología para la selección de compuertas teniendo en cuenta los efectos que se presentan al reducir el voltaje de alimentación en el retardo, su degradación debido a NBTI y además el consumo de potencia mencionados en la sección anterior, de tal forma que se obtenga el conjunto de compuertas las cuales trabajaran con un voltaje de alimentación de $1.0V$, para que un determinado circuito tenga el mejor desempeño con respecto a los tres anteriores aspectos. Además esta metodología hace una estimación de la temperatura de trabajo del circuito a partir de la potencia que consume, esto para tener un cálculo más precisó del efecto del NBTI el cual se ve afectado de manera sustancial por temperaturas altas y con esto tener una estimación menos pesimista

5.2. METODOLOGÍA DE SELECCIÓN DE COMPUERTAS

sobre su impacto en la degradación del retardo.

La estrategia que se ha implementado se ilustra en la figura 5.14 en la cual partiendo de la descripción a nivel compuerta del circuito se detectan todos los caminos de señal o caminos lógicos existentes desde una entrada principal a una salida principal. Sobre estos caminos lógicos se realiza un pre-filtrado para evitar altos costos en el tiempo de cómputo, después se realizan los cálculos de retardo y retardo degradado teniendo en cuenta la relación de temperatura y potencia de los caminos lógicos pre-filtrados. A continuación se procede a bajar el voltaje de cada una de las compuertas pertenecen a estos caminos lógicos para calcular una métrica que permita seleccionar las compuertas que trabajaran con este voltaje de tal forma que se obtiene un nuevo circuito que necesita menos banda de guarda y consumirá menos potencia.

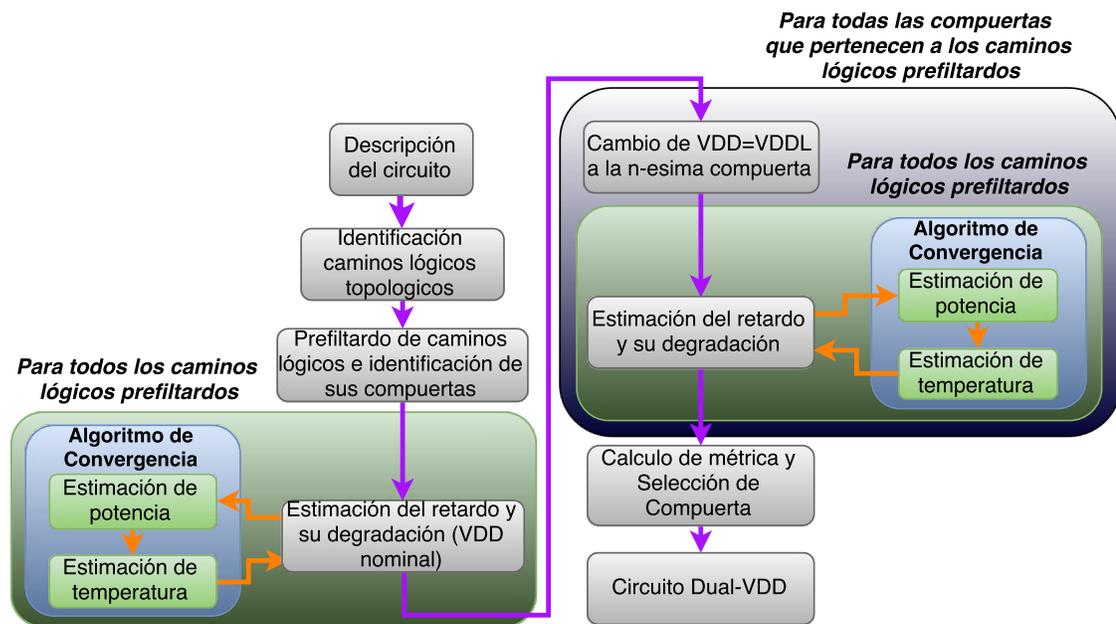


Figura 5.14: Diagrama de flujo de la metodología propuesta

5.2.1. Identificación de caminos lógicos topológicos y cálculo de su retardo medio de propagación

Para identificar los caminos lógicos topológicos se parte de una compuerta que maneje una salida principal (fin del camino lógico) y se recorre hacia atrás el trayecto de una posible señal digital, manteniendo como prioridad que la señal entre por el primer nodo de fan-in. Una vez encontrada una entrada principal, se ha detectado un camino lógico y mediante una bandera que consiste en marcar ciertos nodos por donde no puede volver a pasar la señal, se indica que ese camino recorrido ya ha sido detectado para evitar repetirlo. Entonces, se retorna a la salida principal para repetir el procedimiento cumpliendo las restricciones de las marcas. Una vez que todos los caminos lógicos que terminan en una salida principal se han encontrado, se cambia hacia la siguiente salida principal y se repite el procedimiento.

La figura 5.16 ilustra la metodología de identificación de los dos primeros caminos lógicos para un circuito prueba. Al detectar el camino lógico azul, se ha marcado la entrada principal correspondiente al nodo 1 de la compuerta $X1$, por tanto para el segundo camino lógico, la señal se ha propagado por el nodo 2 de la misma y ha pasado por el nodo 1 de la compuerta $X0$. El tercer camino lógico detectado, marcaría el segundo nodo de la compuerta $X0$, haciendo que se marque a su vez el segundo nodo de la compuerta $X1$ y el primero de la compuerta $X2$ ya que los nodos de las compuertas precedentes estarían completados. En este momento se pasaría a repetir el análisis para el segundo nodo de la compuerta $X2$.

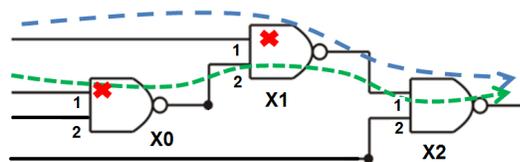


Figura 5.15: Identificación de caminos lógicos en un circuito prueba

Una vez detectados todos los caminos existentes en el circuito, se evalúa la capacitancia de carga en cada uno de los nodos del circuito sumando la capacitancia de entrada de cada una de las compuertas manejadas por un mismo nodo.

Con esta información y asignando un valor de slew-rate y tipo de transición (alto a bajo, bajo a alto) de la señal digital a las entradas principales del circuito, se evalúan los polinomios de retardo, slew rate a la salida y sensibilidad correspondientes a una señal que se propaga por los nodos de cada camino lógico encontrado. El slew rate a la salida de una compuerta es utilizado como slew-rate a la entrada de la compuerta subsiguiente.

5.2.2. Estimación de Retardo y su degradación de los caminos lógicos con $V_{DD} = \text{Voltaje Nominal}$

En este paso se toma el circuito con todos los voltajes de entradas principales y de alimentación iguales a 1.2 Voltios y para todos los caminos lógicos se calcula el retardo y su degradación con una temperatura inicial de $100^{\circ}C$, después se selecciona el camino lógico más lento con su respectiva degradación (*Longest Critical Path LCP_{aged}*), con el retardo de este camino lógico se puede determinar la frecuencia máxima de trabajo del circuito la cual es igual a $1/D_{LCP_{aged}}$. Con la frecuencia de trabajo del circuito se procede a calcular la potencia dinámica de cada una de las compuertas y la potencia dinámica total del circuito. También con la temperatura inicial se calcula la potencia debida a corrientes de subumbral de cada una de las compuertas y la total del circuito.

A partir de la potencia total del circuito y de cada una de las compuertas se procede a calcular las temperaturas de cada compuerta con el modelo térmico explicado en la sección 3. Las nuevas temperaturas son actualizadas y se procede a recalculer el retardo del LCP y la nueva frecuencia de operación.

Los anteriores pasos se realizan de manera iterativa hasta que el valor de la potencia total y temperaturas del circuito convergen a un valor.

Con este último valor de temperaturas se calcula el retardo de todos los caminos lógicos, además también se obtiene la nueva especificación de retardo la cual es el retardo del LCP más un 10 % del retardo del mismo ($D_{max} = 1.1D_{LPC_{age}}$), esta especificación sera el retardo máximo al cual puede tener el circuito y se tendrá en cuenta cuando se calcule la métrica de selección de compuertas en pasos posteriores.

5.2.3. Prefiltrado de caminos lógicos

Debido a que el tiempo de cómputo es alto al realizar el análisis de todas las compuertas y todos los caminos lógicos, se debe tomar solo una porción de caminos lógicos lentos, es decir aquellos con más retardo, para que el tiempo de cómputo se disminuya significativamente, para esto se escogen los caminos lógicos que tengan un retardo entre un 20 % o 40 % menos del retardo del camino lógico más lento *Longest Critical Path* encontrado en la anterior etapa. El valor del porcentaje depende de la cantidad de compuertas del circuito, cuando la cantidad de compuertas es muy grande (> 300) se escoge el 20 %, y cuando la cantidad de compuertas es pequeña (< 300) se escoge el 40 %. Esto se debe porque al tener menos compuertas hay menos opciones de compuertas a seleccionar y tal vez no se alcance la meta con el retardo máximo D_{max} . También en este paso se identifican todas las compuertas que pertenecen a este grupo de caminos lógicos seleccionados, y no afectar los retardos de los caminos lógicos no seleccionados, estas compuertas son las candidatas a trabajar con un voltaje de alimentación menor.

En la figura 5.16 se muestra un ejemplo de cómo se seleccionan los caminos lógicos, en este caso los tres últimos caminos lógicos son seleccionados debido a que sus retardos se encuentran por encima del límite de selección establecido.

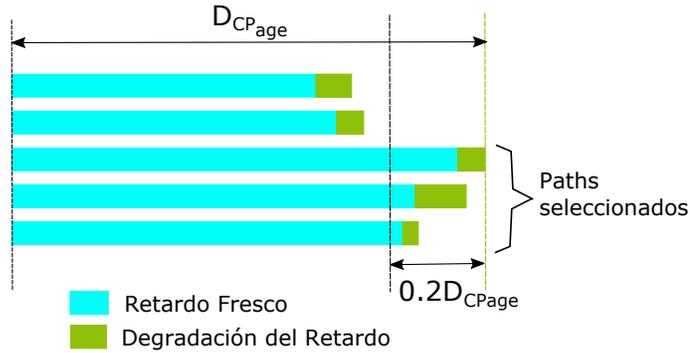


Figura 5.16: Selección de caminos lógicos

5.2.4. Cambiar $V_{DD}=V_{DDL}$ a la n-esima compuerta y estimación de retardos

De las compuertas seleccionadas en el anterior paso se reduce el voltaje de alimentación de una compuerta a la vez y se calcula con la convergencia de temperaturas y de potencia el retardo y la degradación del retardo de los caminos lógicos seleccionados.

5.2.5. Métrica de Selección de Compuertas

Para escoger las compuerta que van a trabajar con un voltaje de alimentación bajo de tal manera que el consumo de potencia del circuito se disminuya, el retardo no sea impactado significativamente y además la degradación del retardo se minimice en los caminos lógicos del circuito, se plantea la siguiente métrica.

$$M = \Delta P \sum_{j \in CP} \Delta(\Delta D_{age,j}) slack_j \quad (5.3)$$

Donde ΔP es la disminución porcentual de la potencia debida al bajar el voltaje de alimentación a una compuerta con respecto a la potencia del circuito inicial.

$$\Delta P = 100 \frac{(P_{Dual-VDD} - P_{VDDnom})}{P_{VDDnom}} \quad (5.4)$$

$\Delta(\Delta D_{age,j})$ es la disminución porcentual de la degradación del retardo de camino lógico j-esimo que pertenece a los caminos lógicos seleccionados debido a reducir el voltaje de alimentación de una compuerta con respecto a la degradación del retardo de este camino lógico j-esimo del circuito inicial, este valor tomado los LCP´s del circuito con voltaje nominal y del circuito Dual-VDD representa el porcentaje de reducción de la banda de guarda al utilizar esta metodología.

$$\Delta(\Delta D_{age,j}) = 100 \frac{(\Delta D_{age,j}^{Dual-VDD} - \Delta D_{age,j}^{VDD-nom})}{\Delta D_{age,j}^{VDD-nom}} \quad (5.5)$$

$slack_j$ o tiempo de holgura es la diferencia de tiempo entre de la especificación de retardo máxima ($D_{max} = 1.1D_{LCP_{aged}}^{VDD-nom}$) y el camino lógico j-esimo debido a bajar el voltaje de alimentación de una compuerta.

$$slack_j = D_{max} - D_{age,j}^{Dual-VDD} \quad (5.6)$$

Lo que se busca al reducir el voltaje de alimentación de una compuerta en un circuito es que la disminución de potencia sea lo más grande posible. De igual manera se busca que la disminución de la degradación del retardo en todos lo caminos lógicos sea lo mas alta posible, al multiplicar la disminución de la degradación del retardo de cada camino lógico por su respectivo tiempo de holgura se le da un peso mas alto a la compuerta que se le baja el voltaje de alimentación impacte menos en el retardo de todos los caminos lógicos. Como ΔP y $\Delta(\Delta D_{age,j})$ son negativos y tiempo de holgura es positivo la métrica tiene un signo positivo y se escogerá la compuerta que en su resultado dé el mayor valor de la métrica.

En la figura 5.17 se ilustra los comportamientos que han tenido los caminos lógicos antes y después de reducir el voltaje de alimentación de algunas compuestas, por ejemplo en los caminos lógicos 1,2 y 5, el retardo fresco se incrementa poco y la degradación baja, el camino lógico 3 no presenta ningún cambio debido a que no se ha movido ninguna fuente de ese camino lógico y por ultimo en el camino lógico 4 se incrementa de manera notoria el retardo fresco lo cual lo con-

vierte en el nuevo camino crítico. En esta figura también se ilustra lo que es la nueva especificación de retardo, en este caso se obtiene a partir del camino lógico 3 el cual era el camino lógico crítico y también se ilustra lo que son los tiempos de holgura que para los caminos lógicos 3 y 4 son valores pequeños, por lo cual el valor de su métrica puede ser bajo.

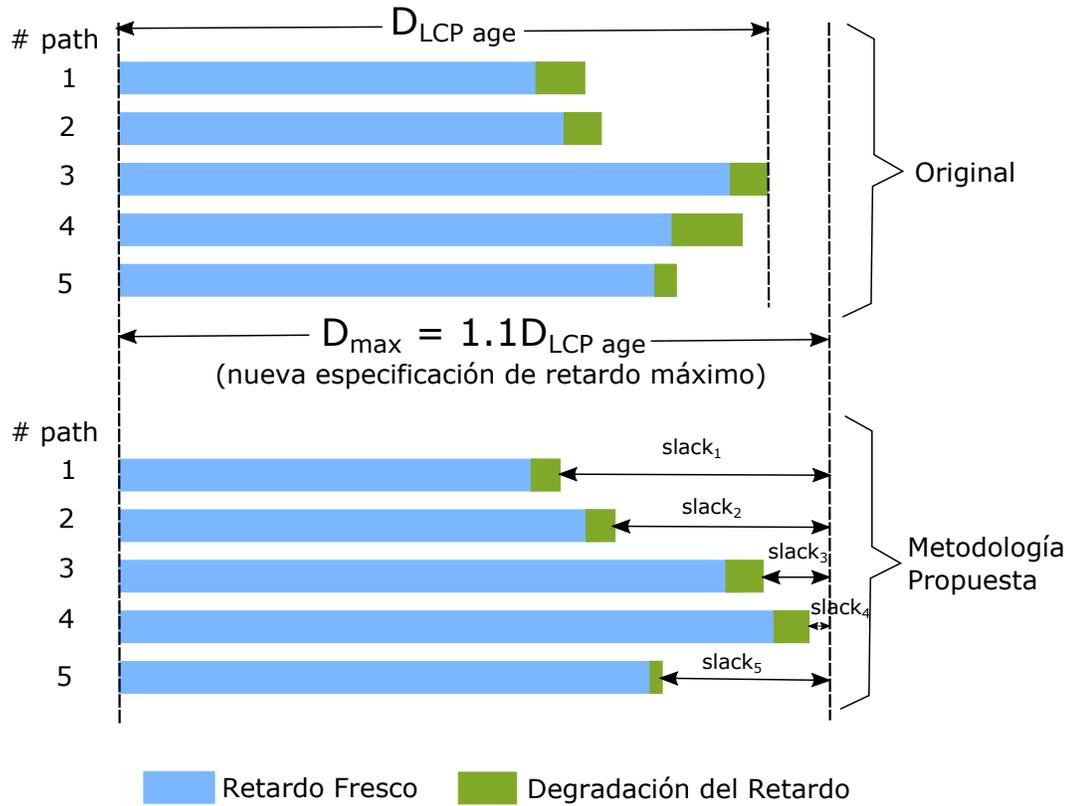


Figura 5.17: Retardo y degradación del retardo antes y después de bajar el voltaje de alimentación de una o varias compuertas

5.2.6. Validación de la Métrica

Para validar la métrica se toma un circuito compuesto por una cadena de 6 inversores con diferentes capacitancias de carga que se ilustra en la figura 5.18 el cual solo tiene un camino lógico, este circuito con todos sus inversores trabajando con un voltaje de alimentación de $1.2V$ tiene un retardo fresco $477ps$, una degra-

5.2. METODOLOGÍA DE SELECCIÓN DE COMPUERTAS

dación de retardo de $38.7ps$ y un retardo degradado igual de $515.7ps$, su consumo de potencia es de $23\mu W$. para este caso la nueva especificación de retardo es de 20 % más del retardo degradado es decir $1.2(515.7)ps = 618.9ps$.

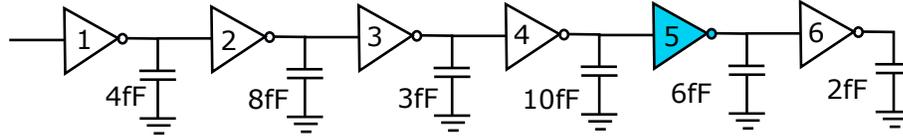


Figura 5.18: Circuito para evaluación de la métrica

El siguiente paso ha sido bajar el voltaje de alimentación de cada una de las compuertas, para así obtener los cambios de potencia, degradación de retardo y valores de los tiempos de holgura, para así calcular la métrica para cada compuerta. Este circuito al tener un solo camino lógico el valor de $\Delta(\Delta D_{aged})$ corresponde al porcentaje de disminución de la banda de guarda.

Los valores obtenidos se muestran en la tabla 5.1 de manera descendente, la métrica tiene mayor valor para el inversor 5, debido que al bajar el voltaje de esta compuerta la banda de guarda se reduce en un 47 % es el mas alto de todas los inversores, ademas el consumo de potencia disminuye en 10.8 %, y el tiempo de holgura es de $73ps$, el inversor que sigue es el número 1, a pesar de que ahorra un poco mas de potencia, el tiempo de holgura es la mitad comparado con el inversor 5. El siguiente valor de métrica es para el inversor 3, el cual tiene una disminución de consumo de potencia igual al anterior pero su tiempo de holgura es 3 veces menor. El inversor 2 sigue con el cuarto valor de métrica, esto se debe a su baja recuperación de banda de guarda y un tiempo de holgura de la mitad. El siguiente es el inversor 4, a pesar de tener la mayor disminución de potencia el tiempo de holgura es el mas pequeño estos dos cambios se deben a su alta capacitancia de carga, lo cual indica que al bajar el voltaje de este inversor el camino lógico se vuelve mas lento, pero el consumo de potencia dinámica baja mas que en otros inversores. Por último esta el inversor 6 que a pesar de tener el mayor tiempo de holgura su recuperación de banda de guarda y disminución del consumo de

potencia son casi nulos.

Inversor	$\Delta(\Delta D_{aged})$	ΔP	slack	Métrica
5	-47 %	-10.8 %	73 ps	2.7e-8
1	-29.7 %	-14.9 %	36 ps	1.6e-8
3	-32 %	-15 %	27 ps	1.33e-8
2	-19.4 %	-18.1 %	36 ps	1.26e-8
4	-24 %	-22.5 %	14 ps	8.25e-9
6	-6.4 %	-3.29 %	95 ps	2.1e-9

Tabla 5.1: Datos de disminución de banda de guarda y consumo de potencia, tiempo de holgura y métrica para cada uno de los 6 inversores

El funcionamiento de la métrica se explica ya que al escoger el inversor 5 (azul) se disminuye el consumo de potencia y la degradación del retardo a un costo bajo de aumento de retardo total, lo que no pasa con los inversores 1,3, y 4, que a pesar de tener valores altos de disminución tanto de consumo de potencia como de degradación de retardo, su penalización al retardo total es mas grande y esto hace que tiempo de holgura sea pequeño y por lo tanto sus métricas tienen un valor mas bajo.

5.2.7. Obtención del conjunto de compuertas con la métrica más alta

Como se explicó anteriormente se escoge la compuerta con mayor valor en la métrica, a esta compuerta se le baja su voltaje de alimentación y se mantiene en este valor; luego se procede a bajar el voltaje de las compuertas restantes re-calculando nuevamente retardos, degradación de retardos y potencias, esto se realiza para calcular nuevamente la métrica de las demás compuertas y así seleccionar una nueva compuerta la cual trabajará con el voltaje de alimentación bajo. Es posible que al bajar el voltaje de alimentación de una determinada compuerta,

uno o varios caminos lógicos a los que pertenece violen la especificación de retardo máxima, para evitar esto automáticamente a esa compuerta se le asigna un valor de métrica $M = 0$, siendo descartada del procedimiento y manteniendo su voltaje de alimentación igual al nominal. El anterior procedimiento se realiza iterativamente hasta que se llegue a un valor máximo de la métrica es decir cuando la disminución de las degradaciones de retardo y de potencia sean las máximas posibles.

5.3. Resultados en circuitos ISCAS

A continuación se resumen los resultados obtenidos del trabajo de tesis. En la tabla 5.2 se muestran la cantidad de caminos lógicos topológicos y compuertas, antes y después de la selección, obtenidos mediante el criterio mencionado. El circuito S298 utilizó un criterio de selección de 40 %, esto se debe a que tiene un número de compuertas menor a 300, como se mencionó anteriormente.

También se muestra en la última columna la cantidad de compuertas que fueron seleccionadas por medio de la métrica para trabajar a un voltaje de alimentación menor, como se puede observar esta cantidad de compuertas es aproximadamente la mitad del número de compuertas seleccionadas en todos los circuitos de prueba.

Circuito	caminos lógicos topológicos	Compuertas	caminos lógicos Seleccionados	Compuertas Seleccionadas	Compuertas con voltaje bajo
S298	231	166	89	47	22
S838	1714	279	329	116	61
C880	13519	335	1681	97	53
S5378	11728	1021	1297	156	83

Tabla 5.2: Caminos lógicos y Compuertas. Antes y después de la selección

5.3.1. Degradación de Retardo

La tabla 5.3 muestra la degradación de retardo total, es decir el máximo retardo que presenta el circuito envejecido menos el máximo retardo que presenta el circuito al principio de su operación y el promedio de degradación de retardo que es el promedio de la degradación del retardo de los caminos lógicos seleccionados

Circuito	camino lógico Critico (LCP)		Caminos lógicos Seleccionados	
	Degradación Retardo V_{DD} Nominal (ps)	Degradación Retardo Dual VDD (ps)	Promedio de la Degradación del Retardo V_{DD} Nominal (ps)	Promedio de la Degradación del Retardo Dual VDD (ps)
S268	23.04	20.36 (-11.63 %)	21.27	12.39 (-41.77 %)
S838	35.53	31.29 (-11.92 %)	41.39	31.11 (-24.83 %)
C880	60.85	37.87 (-37.77 %)	45.88	33.86 (-24.14 %)
S5378	189.62	140.69 (-25.80 %)	134.12	101.75 (-26.21 %)

Tabla 5.3: Degradación del retardo Total y Promedio del los circuitos Original y Dual-VDD

El comportamiento del circuito *S268* el cual presenta menor disminución de la degradación total del retardo y alta reducción de la degradación de retardo promedio con -41.77 % se debe a que en el circuito *S268* presenta una gran cantidad de caminos lógicos con alta degradación cuando trabajo con el voltaje nominal lo que implica que tiene una degradación de retardo promedio alta, pero después de aplicar la metodología se reduce significativamente la cantidad de caminos lógicos con degradación de retardo alta. El circuito *S5378* presenta un comportamiento intermedio, la degradación del LCP disminuye de manera similar que la de los demás caminos lógicos. El circuito *C880* presenta mayor disminución de la degradación de retardo total y su banda de guarda se disminuye en un 37.77 %, se debe a que la mayor cantidad de caminos lógicos su degradación de retardo se encuentran en una región en particular por lo cual la degradación del camino

lógico crítico es más probable que este en esta región.

En la figura 5.19 se observa cómo la degradación de los retardo de los caminos lógicos seleccionados se desplazan hacia valores menores, también se observa una tendencia en todos los circuitos de acumular mayor número de caminos lógicos a la izquierda de los histogramas, esto quiere decir que el numero de caminos lógicos con degradación mas baja aumenta, mostrando la efectividad del Dual-VDD para bajar el impacto del NBTI

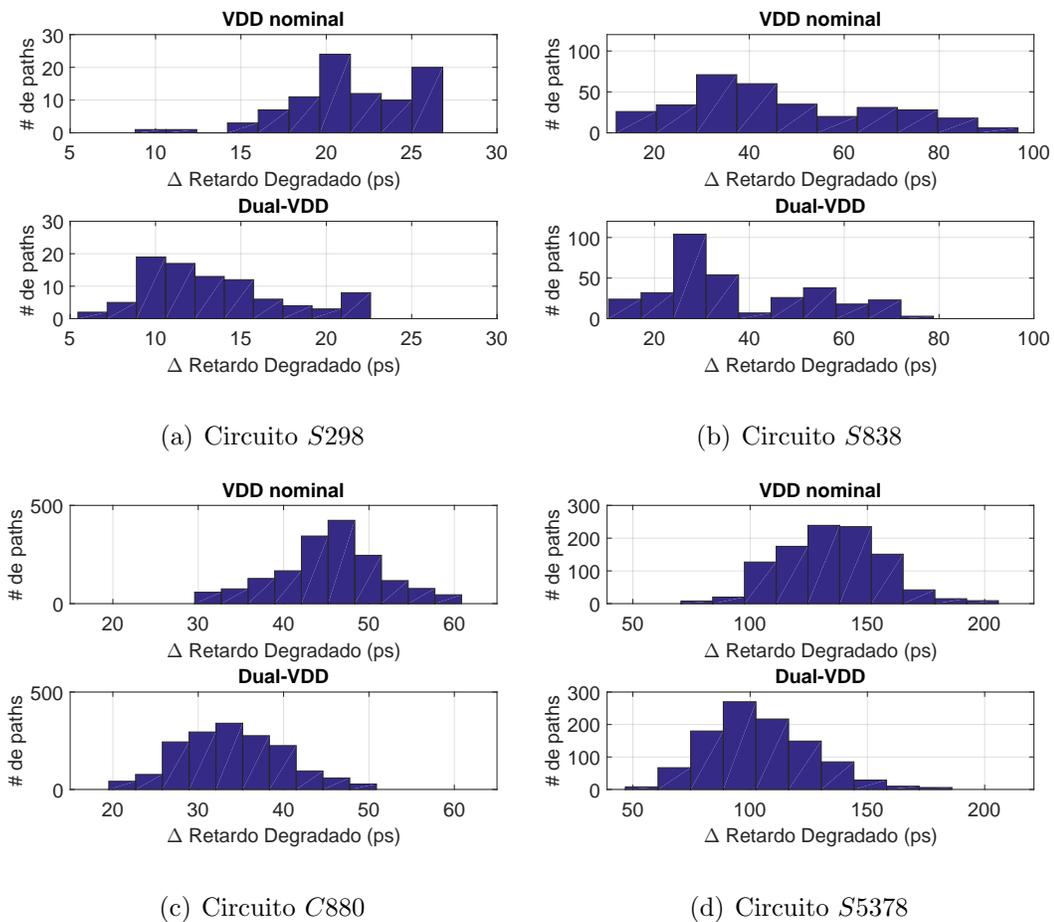


Figura 5.19: Histogramas de degradación de retardo de los caminos lógicos seleccionados para el circuito VDD nominal y Dual-VDD

5.3.2. Consumo de Potencia

Aunque reducir el voltaje de las entradas a una compuerta aumenta el consumo de potencia estática por el aumento de las corrientes de subumbral, esto se compensa con la disminución de la Potencia Dinámica, en la tabla 5.4 se muestran los consumos de potencia del circuito original y el circuito con Dual-VDD.

Circuito	Circuito V_{DD} Nominal			Circuito Dual- V_{DD}		
	Potencia Dinámica (μW)	Potencia Subumbral (nW)	Potencia Total (μW)	Potencia Dinámica (μW)	Potencia Subumbral (nW)	Potencia Total (μW)
S268	139.81	112.36	139.92	117.74 (-15.79 %)	177.60 (+58.07 %)	117.92 (-15.72 %)
S838	31.85	166.52	32.02	26.60 (-16.48 %)	131.47 (-21.06 %)	26.73 (-16.52 %)
C880	96.69	163.55	96.86	82.54 (-14.63 %)	117.33 (-28.26 %)	82.66 (-14.66 %)
S5378	126.67	1200.72	127.87	111.98 (-11.60 %)	1083.3 (-9.78 %)	113.06 (-11.58 %)

Tabla 5.4: Consumo de Potencia circuitos V_{DD} Nominal y Dual-VDD

El circuito que presenta un aumento en el consumo de potencia estática de subumbral es el *S268*, esto se debe a que tiene varias compuertas NOR3 manejadas por voltajes bajos en todas sus entradas lo que aumenta significativamente la potencia estática. Para los demás circuitos el consumo de potencia estático es menor con Dual-VDD, debido a que dominan los efectos de que las temperaturas sean mas bajas que las del circuito original.

De igual manera se observa que el efecto del incremento o decremento de la potencia estática por corrientes de subumbral, no impacta demasiado en el consumo de potencia total, esto se debe a que la potencia dinámica tiene de 2 a 3 ordenes de magnitud mas. Debido a lo anterior en el consumo de potencia total se puede observar que para los cuatro circuitos de prueba el consumo de potencia disminuye con Dual-VDD, lo que significa que al bajar el voltaje de alimentación de varias compuertas se afecta mas la disminución de potencia dinámica y por ende baja las temperaturas de operación, y esta a su vez baja el valor de las corrientes de subumbral.

Al aplicar la técnica Dual-VDD usando la métrica de selección propuesta se logra reducir el impacto de NBTI en el retardo de los caminos logicos críticos,

también se logra reducir la banda de guarda necesaria al disminuir la degradación del retardo del camino lógico mas lento de los circuitos y por ultimo consumo de potencia total también se redujo. La degradación del retardo promedio se redujo significativamente para todos los circuitos con un incremento en el retardo total no mayor al 10 %. En el caso del consumo de la potencia estática debido a corrientes de subumbral también hubo una reducción en la mayoría de circuitos en los que se probó sin insertar convertidores de voltaje.

Capítulo 6

Conclusiones

El impacto del envejecimiento o desgaste debido a NBTI en el rendimiento de los circuitos y sistemas electrónicos ha aumentado debido al escalado de las tecnologías de semiconductores. Como consecuencia, el tiempo de vida útil de un circuito se reduce. Esto es de fundamental importancia en un amplio rango de aplicaciones electrónicas. En aplicaciones de la industria automovilística, espacial y médica su importancia es aún mayor. Por lo tanto, los diseñadores deben tomar medidas para garantizar el correcto funcionamiento de los circuitos.

En esta tesis se ha propuesto una metodología para aumentar la robustez de los circuitos a envejecimiento de NBTI utilizando voltaje dual de alimentación. Esta metodología está orientada para aplicaciones donde sea de interés reducir el consumo de potencia del circuito sin utilizar convertidores de alimentación, y a la vez garantizar la robustez del circuito a envejecimiento de NBTI. La metodología propuesta usa un voltaje de alimentación menor al nominal para un conjunto de compuertas especialmente seleccionadas en los caminos críticos del circuito. Se propone también una métrica de selección de compuertas. Se ha desarrollado una herramienta que implementa esta metodología usando C++, y se ha aplicado en circuitos ISCAS combinacionales y secuenciales.

El consumo de potencia se reduce principalmente debido a que el voltaje de

alimentación de varias compuertas se reduce. Además, la disminución del retardo nominal al utilizar un voltaje de alimentación menor al nominal lleva a una reducción adicional del consumo de potencia. La reducción del voltaje de alimentación permite disminuir la degradación del retardo del circuito. Esto debido a que la degradación del dispositivo es menor como consecuencia de utilizar un voltaje de alimentación menor, y a que la temperatura del circuito también disminuye como consecuencia del voltaje de alimentación menor y una menor frecuencia de operación. Debe notarse, que la sensibilidad del retardo a variaciones en el voltaje umbral también aumenta lo que llevaría a una mayor degradación, sin embargo, en este trabajo se ha encontrado que el efecto de la disminución de la degradación del dispositivo es más significativo que el aumento de sensibilidad del retardo a variaciones en el voltaje umbral.

En este trabajo se ha propuesto una métrica para la selección de las compuertas más adecuadas a usar un voltaje de alimentación menor al nominal. Esta métrica persigue seleccionar aquellas compuertas más adecuadas para tener una menor penalización del retardo nominal del circuito, una mayor reducción de la degradación debido a NBTI y la mayor reducción del consumo de potencia.

En la metodología propuesta se han estimado los perfiles de temperatura del circuito de acuerdo a su operación. Esto permite obtener una metodología realista debido a la importancia del impacto de la temperatura en el envejecimiento debido a NBTI. Se realiza una estimación de la temperatura de trabajo del circuito a partir de la potencia que consume.

Lista de Figuras

1.1. Tasa de Fallas representada en la Curva de tina de baño. [3]	4
1.2. Degradación del rendimiento e incumplimiento del tiempo de vida requerido. [5]	5
1.3. Tendencias en la curva de tina de baño	6
1.4. Mecanismos de Hot Carriers	7
1.5. TDDDBD	8
1.6. Estructura lógica de un circuito digital	9
1.7. Margen de Seguridad para garantizar la correcta operación del cir- cuito	9
2.1. Condición de polarización donde el efecto de NBTI se presenta . . .	13
2.2. Interface del silicio y óxido de compuerta	14
2.3. Tuneleo de huecos y rompimiento de enlaces $S_i - H$ en la interface (Reacción).	15
2.4. Difusión de los átomos de hidrógeno en el óxido (Difusión).	15
2.5. Pasivación de enlaces con átomos de hidrógeno que se recuperan (recuperación)	16
2.6. Dependencia del NBTI con el campo eléctrico y temperatura. [7] . .	17
2.7. Desviación del voltaje umbral como consecuencia del NBTI. [6] . .	18
2.8. Dependencia del NBTI con la frecuencia de operación. [9]	19
2.9. LTDM como borde superior de la degradación del V_{th} [14]	20
2.10. Curvas obtenidas al aplicar el LTDM	21

2.11. Variación del Voltaje de umbral debido a NBTI para diferentes voltajes de operación	22
2.12. Variación del Voltaje de umbral debido a NBTI para diferentes temperaturas de operación	23
2.13. Variación del Voltaje de umbral debido a NBTI para diferentes temperaturas y tensiones de operación	24
2.14. Variación del retardo de subida de una compuerta inversora	25
2.15. Degradación del retardo para diferentes conjuntos de probabilidad en las entradas principales de un circuito digital. [14]	26
2.16. Degradación del retardo diferente para dos caminos del circuito C17. [14]	27
2.17. Error de operación debido al envejecimiento	28
3.1. Curvas de movilidad vs campo eléctrico en silicio forzado [19]	31
3.2. Conjunto de pareto para disminuir efecto de NBTI y consumo de potencia simultáneamente durante el período de stanby [20]	32
3.3. Técnica de reemplazo de compuertas para controlar el estado de algunos nodos internos	33
3.4. Sobre dimensionamiento de las compuertas	35
3.5. Condiciones de estrés en una compuerta inversora	36
3.6. Condiciones de estrés en una compuerta NOR de dos entradas	36
3.7. Ordenamiento de los nodos de entrada para disminuir el envejecimiento en una compuerta NOR3	37
3.8. Tipos de sensores	40
3.9. Sensor de predicción de error	41
3.10. Funcionamiento del sensor de predicción de error	42
3.11. Sensor de detección de error	42
3.12. Funcionamiento del sensor de detección de error	43
3.13. Disminución del efecto NBTI al bajar el voltaje de alimentación	44

4.1. Caracterización del retardo de propagación de una compuerta digital	48
4.2. Retardo de un inversor en función de multiples variables	49
4.3. Sensibilidad del retardo a cambio de voltaje de umbral de un in- versor en función de multiples variables	50
4.4. Diagrama de Potencias Disipadas por un Circuito Digital	51
4.5. Corrientes y Voltajes de un inversor con carga CL	51
4.6. Diagrama de la corriente de corto circuito en un Inversor	54
4.7. Pulsos Espurios en una compuerta AND2	57
4.8. Conducción de calor a través de un solido	58
4.9. Circuito Equivalente de la conducción de calor a través de un solido	58
4.10. representación de los materiales que componen un CI	59
4.11. Circuito equivalente para calculo de temperatura de un CI	61
4.12. Interacciones presentes en un circuito digital Dual-VDD	61
4.13. Algoritmo iterativo para cálculo de retardo, potencia y temperatura	63
5.1. Ejemplo de Escalamiento de voltaje agrupado (CVS)	65
5.2. Ejemplo de Escalamiento de voltaje agrupado Extendido (ECVS)	65
5.3. Comportamiento del retardo fresco vs el voltaje de alimentación .	67
5.4. Cambio de las componentes de los factores de ΔD_{aged}	68
5.5. Layout transistores PMOS Dual-VDD	69
5.6. Interconexiones para PMOS para Layout Dual VDD	70
5.7. V_{th} para diferentes voltajes de VDD y V_{bulk}	71
5.8. Impacto en la degradación del V_{th} debido al efecto cuerpo	72
5.9. Retardo Normalizado para diferentes voltajes de V_{DD} y V_{bulk}	72
5.10. Corriente Estática que se presenta cuando una compuerta con V_{DDL} maneja una compuerta compuerta con V_{DDH}	73
5.11. Conversor de nivel asíncrono [43]	73
5.12. Corriente de subumbral de una compuerta NAND4	74
5.13. Potencia de subumbral de una compuerta NAND4	75

5.14. Diagrama de flujo de la metodología propuesta	76
5.15. Identificación de caminos lógicos en un circuito prueba	77
5.16. Selección de caminos lógicos	80
5.17. Retardo y degradación del retardo antes y después de bajar el voltaje de alimentación de una o varias compuertas	82
5.18. Circuito para evaluación de la métrica	83
5.19. Histogramas de degradación de retardo de los caminos lógicos seleccionados para el circuito VDD nominal y Dual-VDD	87

Lista de Tablas

4.1. Factores de Actividad de Compuertas	53
4.2. Conductividad Térmica y Resistencia Térmica respectiva	60
5.1. Datos de disminución de banda de guarda y consumo de potencia, tiempo de holgura y métrica para cada uno de los 6 inversores . .	84
5.2. Caminos lógicos y Compuertas. Antes y después de la selección . .	85
5.3. Degradación del retardo Total y Promedio del los circuitos Original y Dual-VDD	86
5.4. Consumo de Potencia circuitos V_{DD} Nominal y Dual-VDD	88

Bibliografía

- [1] Moore, G.E., "Progress in digital integrated electronics [Technical literature, Copyright 1975 IEEE. Reprinted, with permission. Technical Digest. International Electron Devices Meeting, IEEE, 1975, pp. 11-13.], in Solid-State Circuits Society Newsletter, IEEE , vol.20, no.3, pp.36-37, Sept. 2006
- [2] Manfred Dietrich, Joachim Haase, "Process Variations and Probabilistic Integrated Circuit Design", ISBN 978-1-4419-6620-9, Springer New York Dordrecht Heidelberg London
- [3] Alvin W. Strong, Ernest Y. Wu, Rolf-Peter Vollertsen, Jordi Suñé, Giuseppe La Rosa, Stewart E. Rauch III, Timothy D. Sullivan. "Reliability wearout mechanisms in advanced CMOS technologies". IEEE Press Series on Microelectronic Systems. 2009.
- [4] Seyab, Nor Zaidi Haron, Said Hamdioui, "CMOS scaling impacts on reliability, What do we understand?" Delft University of Technology, Computer Engineering Laboratory.
- [5] M. A. Alam, "A broad overview of reliability of semiconductor MOSFET" Purdue University, W. Lafayette, IN. disponible digitalmente en: <https://nanohub.org/resources/5863/download/2008.11.11-ece612-l20.pdf>
- [6] Alam, M. A., Kufluoglu, H., Varghese, D. & Mahapatra, S. (2007). "A comprehensive model for pmos nbtj degradation: Recent progress", Microelectronics Reliability Vol. 47(No.6): 853-862.

- [7] Mahapatra, S.; Bharath Kumar, P.; Alam, M.A., “Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs,” *Electron Devices, IEEE Transactions on* , vol.51, no.9, pp.1371,1379, Sept. 2004
- [8] S. Rangan, N. Mielke, E. C. C. Yeh, “Universal recovery behavior of negative bias temperature instability,” *IEDM*, pp. 341-344, 2003
- [9] Chen, G.; Chuah, K. Y.; Li, M. -F; Chan, D. S H; Ang, C.H.; Zheng, J.Z.; Jin, Y.; Kwong, D.L., “Dynamic NBTI of PMOS transistors and its impact on device lifetime,” *Reliability Physics Symposium Proceedings, 2003. 41st Annual. 2003 IEEE International* , vol., no., pp.196,202, 30 March-4 April 2003
- [10] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy. “Impact of NBTI on the temporal performance degradation of digital circuit” , *IEEE Electronic Device Letters*, 26(8):560-562, August 2005.
- [11] Robert Entner, “Modeling and Simulation of Negative Bias Temperature Instability”, Dissertation, Wien, im April 2007.
- [12] Bhardwaj, S.; Wenping Wang; Vattikonda, R.; Yu Cao; Vrudhula, S., “Predictive Modeling of the NBTI Effect for Reliable Design” *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE* , vol., no., pp.189,192, 10-13 Sept. 2006
- [13] “Solid state circuit technologies”, Edited by Jacobus W. Swart, ISBN 978-953-307-045-2, 2010.
- [14] Wenping Wang; Shengqi Yang; Bhardwaj, S.; Vrudhula, S.; Liu, T.; Yu Cao, “The Impact of NBTI Effect on Combinational Circuit: Modeling, Simulation, and Analysis,” *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.18, no.2, pp.173,183, Feb. 2010

- [15] Wenping Wang; Reddy, V.; Yang, B.; Balakrishnan, V.; Krishnan, Srikanth; Yu Cao, “Statistical prediction of circuit aging under process variations,” Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE , vol., no., pp.13,16, 21-24 Sept. 2008
- [16] Song Jin; Yinhe Han; Huawei Li; Xiaowei Li, “ P^2 CLRAF: An Pre- and Post-Silicon Cooperated Circuit Lifetime Reliability Analysis Framework,” Test Symposium (ATS), 2010 19th IEEE Asian , vol., no., pp.117,120, 1-4 Dec. 2010
- [17] Lorenz, D.; Georgakos, G.; Schlichtmann, U., “Aging analysis of circuit timing considering NBTI and HCI” On-Line Testing Symposium, 2009. IOLTS 2009. 15th IEEE International , vol., no., pp.3,8, 24-26 June 2009
- [18] M.S. Khan and S. Hamdioui, “NBTI-Aware Nanoscaled Circuit Delay Assessment and Mitigation”, Proc. 3rd HiPEAC Workshop on Design for Reliability, Heraklion, Greece, January, 2011.
- [19] A.E. Islam, et al. “On the Possibility of Degradation Free Field Effect Transist”, Applied Physics Letter, 2008
- [20] Yu Wang; Xiaoming Chen; Wenping Wang; Balakrishnan, V.; Yu Cao; Yuan Xie; Huazhong Yang, “On the efficacy of input Vector Control to mitigate NBTI effects and leakage power,” Quality of Electronic Design, 2009. ISQED 2009. Quality Electronic Design , vol., no., pp.19,26, 16-18 March 2009
- [21] Kumar, S.V.; Kim, C.H.; Sapatnekar, S.S., “NBTI-Aware Synthesis of Digital Circuits,” Design Automation Conference, 2007. DAC '07. 44th ACM/IEEE , vol., no., pp.370,375, 4-8 June 2007
- [22] Wenping Wang; Zile Wei; Shengqi Yang; Yu Cao, “An efficient method to identify critical gates under circuit aging,” Computer-Aided Design, 2007. IC-

- CAD 2007. IEEE/ACM International Conference on , vol., no., pp.735,740, 4-8 Nov. 2007
- [23] Yu Wang; Xiaoming Chen; Wenping Wang; Yu Cao; Yuan Xie; Huazhong Yang, “Gate replacement techniques for simultaneous leakage and aging optimization,” Design, Automation & Test in Europe Conference & Exhibition, 2009. DATE '09. , vol., no., pp.328,333, 20-24 April 2009
- [24] Yu Wang; Xiaoming Chen; Wenping Wang; Yu Cao; Yuan Xie; Huazhong Yang, “Leakage Power and Circuit Aging Cooptimization by Gate Replacement Techniques,” Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.19, no.4, pp.615,628, April 2011
- [25] S. Kiamehr, F. Firouzi, and M. Tahoori, “Input and transistor reordering for nbti and hci reduction in complex cmos gates,” in Proceedings 01 the great lakes symposium on VLSI. ACM, 2012, pp. 201-206.
- [26] H. Luo, Y. Wang, K. He, R. Luo, H. Yang, and Y. Xie, “A novel gate-level nbti delay degradation model with stacking effect,” Integrated Circuit and System Design. Power and Timing Modeling, Optimization and Simulation, pp. 160-170, 2007.
- [27] Kai-Chiang Wu; Marculescu, D., “Joint logic restructuring and pin reordering against NBTI-induced performance degradation,” Design, Automation & Test in Europe Conference & Exhibition, 2009. DATE '09. , vol., no., pp.75,80, 20-24 April 2009
- [28] Kiamehr, Saman; Firouzi, Farshad; Tahoori, Mehdi.B, “Aging-aware timing analysis considering combined effects of NBTI and PBTI,” Quality Electronic Design (ISQED), 2013 14th International Symposium on , vol., no., pp.53,59, 4-6 March 2013

- [29] Hashimoto, M., “Run-time adaptive performance compensation using on-chip sensors,” Design Automation Conference (ASP-DAC), 2011 16th Asia and South Pacific , vol., no., pp.285,290, 25-28 Jan. 2011
- [30] Bowman, K.; Tschanz, J.; Wilkerson, C.; Shih-Lien Lu; Karnik, T.; De, V.; Borkar, S., “Circuit techniques for dynamic variation tolerance,” Design Automation Conference, 2009. DAC '09. 46th ACM/IEEE , vol., no., pp.4,7, 26-31 July 2009
- [31] Das, S.; Blaauw, D, “Adaptive design for nanometer technology,” Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on , vol., no., pp.77,80, 24-27 May 2009 doi: 10.1109/ISCAS.2009.5117689
- [32] Agarwal, M.; Paul, B.C.; Ming Zhang; Mitra, S, “Circuit Failure Prediction and Its Application to Transistor Aging,” VLSI Test Symposium, 2007. 25th IEEE, vol., no., pp.277,286, 6-10 May 2007
- [33] Ernst, D.; Nam Sung Kim; Das, S.; Pant, S.; Rao, R.; Pham, T.; Ziesler, C.; Blaauw, D; Austin, T.; Flautner, K.; Mudge, T., “Razor: a low-power pipeline based on circuit-level timing speculation” Microarchitecture, 2003. MICRO-36. Proceedings. 36th Annual IEEE/ACM International Symposium on , vol., no., pp.7,18, 3-5 Dec. 2003
- [34] Das, S.; Roberts, D.; Seokwoo Lee; Pant, S.; Blaauw, D; Austin, T.; Flautner, K.; Mudge, T., “A self-tuning DVS processor using delay-error detection and correction,” Solid-State Circuits, IEEE Journal of , vol.41, no.4, pp.792,804, April 2006
- [35] Vazquez, J.C., Champac, V., Semião, J., Teixeira, I.C., Santos, M.B., Teixeira, J.P., ”Process Variations-Aware Statistical Analysis Framework for Aging Sensors Insertion”, Journal of Electronic Testing, 2013.

- [36] Ghosh, S.; Bhunia, S.; Roy, K., “CRISTA: A New Paradigm for Low-Power, Variation-Tolerant, and Adaptive Circuit Synthesis Using Critical Path Isolation” *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on , vol.26, no.11, pp.1947,1956, Nov. 2007
- [37] Jinjun Xiong; Zolotov, V.; Lei He, “Robust Extraction of Spatial Correlation,” *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on , vol.26, no.4, pp.619,631, April 2007.
- [38] Lizheng Zhang; Yuheng Hu; Chen, C.C., “Statistical timing analysis with path reconvergence and spatial correlations,” *Design, Automation and Test in Europe*, 2006. DATE '06. Proceedings , vol.1, no., pp.5 pp.,, 6-10 March 2006
- [39] Blaauw, D; Chopra, K.; Srivastava, A.; Scheffer, L., “Statistical Timing Analysis: From Basic Principles to State of the Art,” *Computer-Aided Design of Integrated Circuits and Systems*, IEEE Transactions on , vol.27, no.4, pp.589,607, April 2008
- [40] Kyung-Ki Kim; Wei Wang; Ken Choi, “On-Chip Aging Sensor Circuits for Reliable Nanometer MOSFET Digital Circuits”, *Circuits and Systems II: Express Briefs*, IEEE Transactions on , vol.57, no.10, pp.798,802, Oct. 2010
- [41] Usami, K., and Horowitz, M., “Clustered voltage scaling technique for low-power design,” *Proc. Int. Symp. Low-Power Electronics Design*, 1995, pp. 3-8.
- [42] Igarashi, M.; Usami, K.; Nogami, K.; Minami, F.; Kawasaki, Y.; Aoki, T.; Takano, M.; Sonoda, S.; Ichida, M.; Hatanaka, N., “A low-power design method using multiple supply voltages,” in *Low Power Electronics and Design*, 1997. Proceedings., 1997 International Symposium on , vol., no., pp.36-41, 18-20 Aug. 1997
- [43] Kulkarni, S., and Sylvester, D., “High performance level conversion for dual VDD design,” *IEEE Trans. VLSI Syst.*, Sep. 2004, pp. 926-936.

- [44] Roy, K.; Mukhopadhyay, S.; Mahmoodi-Meimand, H., "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits, in Proceedings of the IEEE , vol.91, no.2, pp.305-327, Feb 2003
- [45] Hongliang Chang; Sapatnekar, S.S., "Full-chip analysis of leakage power under process variations, including spatial correlations, in Design Automation Conference, 2005. Proceedings. 42nd , vol., no., pp.523-528, 13-17 June 2005
- [46] Pedram, M.; Nazarian, S., "Thermal Modeling, Analysis, and Management in VLSI Circuits: Principles and Methods, in Proceedings of the IEEE , vol.94, no.8, pp.1487-1501, Aug. 2006
- [47] Diril, A.U.; Dhillon, Y.S.; Chatterjee, A.; Singh, A.D., "Level-shifter free design of low power dual supply voltage CMOS circuits using dual threshold voltages, in Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.13, no.9, pp.1103-1107, Sept. 2005