Process-Tolerant Low-Power Design for the Nano-meter Regime

Kaushik Roy
Electrical & Computer Engineering
Purdue University
Exponential Increase in Leakage

Silicon Micro-electronics

\[ \frac{I_{ON}}{I_{OFF}} = 10^6 \]

Subthreshold Leakage

Gate Leakage

Source
\( n^+ \)

Drain
\( n^+ \)

Junction leakage

Bulk

Silicon Nano-electronics

\[ \frac{I_{ON}}{I_{OFF}} = 10^3 \]

Non-Silicon Technology

\[ \frac{I_{ON}}{I_{OFF}} \approx 10^{2-6} \]

Leakage Power (% of Total)

Must stop at 50%

A. Grove, IEDM 2002
Technology Trend

2003

Bulk-CMOS

FD/SOI

PD/SOI

Single gate device

2009

FinFET

Trigate

Multi-gate devices

2020

Nano devices
Carbon nanotube
III-V devices
nano-wires
Spintronics

Design methods to exploit the advantages of technology innovations
Variation in Process Parameters

Device parameters are no longer deterministic

Normalize Leakage

Delay and Leakage Spread

Source: Intel

Random dopant fluctuation

Device parameters are no longer deterministic
Reliability

Temporal degradation of performance -- NBTI

Defects

Tech. generation

Failure probability

Life time degradation

Time
Pessimistic Design Hurts Performance

(150nm CMOS Measurements, 110°C)

- Substantial variation in leakage across dies
- 4X variation between nominal and worst-case leakage
- Performance determined at nominal leakage
- Robustness determined at worst-case leakage
Global and Local Variations

Random Dopant Fluctuation

\[ \delta V_t = \Delta V_{t-GLOBAL} + \delta V_{t-LOCAL} \]
Process Tolerance: Memories

S. Mukhopadhaya, Mahmoodi, Roy
VLSI Circuit Symposium 2006, JSSC 2006, TCAD
Parametric Failures: Read Failure

Read failure => Flipping of Cell Data while Reading

\[ P_{RF} = P(V_{\text{READ}} > V_{\text{TRIPRD}}) \]

Read failure => Flipping of Cell Data while Reading
Parametric Failures in SRAM

Parametric failures can degrade SRAM yield

- Read Failures
- Write Failures
- Access Failures
- Hold Failures

Test & Repair using Redundancy

Faulty chips

Working chips
Process Variations in On-chip SRAM

\[ \sigma_{Vt} \approx 30 \text{mv}, \text{using BPTM 45nm technology} \]

Simulation of an 64KB Cache

A. Agarwal, et. al, JSSC, 05

Parametric failures $\rightarrow$ Yield degradation
Inter-die Variation & Cell Failures

Low–Vt Corners
- Read failure ↑
- Hold failure ↑

High–Vt Corners
- Access failure ↑
- Write failure ↑

inter-die Vt shift ($\Delta V_{th\text{-GLOBAL}}$)

$\sigma_{GLOBAL}$
Memory failure probabilities are high when inter-die shift in process is high.
Self-Repairing SRAM Array

Reduce the dominant failures at different inter-die corners to increase width of low failure region.
Apply correction to the global variation to reduce number of failures due to local variations.
Self-Repairing SRAM Array

Reduce the dominant failures at different inter-die corners to increase width of low failure region
How to identify the inter-die Vt corner under a large intra-die variation?

Monitor circuit parameters, e.g. leakage current

Effect of inter-die variation can be masked by intra-die variation
Array Leakage Monitoring

Adding a large number of random variables reduces the effect of intra-die variation.

Leakage of entire SRAM array is a reliable indicator of the inter-die Vt corner.
Self-Repair using Leakage Monitoring

Entire array leakage is monitored to detect inter-die corner and proper body-bias is selected.
Yield Enhancement using Self-Repair

Self-Repairing SRAM using body-bias can significantly improve design yield.
Test-Chip of Self-Repairing SRAM

Technology: IBM 0.13 \( \mu \)m
128KB SRAM
Dual-Vt Triple-well tech.
Number of Trans: \(~7\) million
Die size: 16mm2
VLSI CKT Symp. 2006, ITC 2005

Simulation results for 1MB array designed in IBM 0.13\( \mu \)m
Quantized (3 Level: FBB, ZBB, RBB) body bias scheme is a cost effective solution with good yield enhancement possibility.
Process Tolerance: Register Files
Process Compensating Dynamic Circuit Technology

Conventional Static Keeper

- Keeper upsizing degrades average performance
Process Compensating Dynamic Circuit Technology

3-bit programmable keeper

b[2:0] clk

RS0 RS1 RS7
D0 D1 D7

Opportunistic speedup via keeper downsizing

C. Kim et al., VLSI Circuits Symp. ‘03
Robustness Squeeze

- 5X reduction in robustness failing dies

![Graph showing normalized DC robustness and number of dies saved](image-url)

- Noise floor
- Conventional
- This work

Number of dies

Normalized DC robustness

- 0.7
- 0.8
- 0.9
- 1.0
- 1.1
- 1.2

- Saved dies
Delay Squeeze

- 10% opportunistic speedup

**Bar Chart**
- **Normalized delay**
- **Number of dies**
- **PCD μ = 0.90**
- **Conv. μ = 1.00**
- **μ : avg. delay**

**Legend**
- **Conventional**
- **This work**
On-Die Leakage Sensor For Measuring Process Variation

C. Kim et al., VLSI Circuits Symp. ‘04

- High leakage sensing gain – 90nm dual-Vt, Vdd=1.2V, 7 level resolution, 0.66 mW @80C°
Leakage Binning Results

Output codes from leakage sensor
Self-Contained Process Compensation

Fab → Wafer test

Process detection
- Leakage measurement
- On-die leakage sensor

→ Program PCD using fuses

Customer → Package test

→ Burn in

→ Assembly
Self-Repair: Architecture Level

Agarawal, Roy TVLSI 2005
Fault-Tolerant Cache Architecture

- BIST detects the faulty blocks
- Config Storage stores the fault information

Idea is to resize the cache to avoid faulty blocks during regular operation
Mapping Issue

More than one INDEX are mapped to same block

Include column address bits into TAG bits

Tag does not match, cache miss

Resizing is transparent to processor → same memory address
- Proposed architecture can handle more number of faulty cells than ECC, as high as 890 faulty cells
- Saves more number of chips than ECC for a given $N_{\text{Faulty-Cells}}$
- Increase in miss rate due to downsizing of cache
- Average CPU performance loss over all SPEC 2000 benchmarks for a cache with 890 faulty cells is ~ 2%
Logic: Process Tolerance
Logic: A New Paradigm for Low-Voltage, Variation Tolerant Circuit Synthesis Using Critical Path Isolation (CRISTA)

Ghosh, Bhunia, Roy -- ICCAD 2006
Razor Approach

- Post-Silicon technique for dynamic supply scaling and timing error detection/correction
- Error correction overhead is 1% for a 10% error rate

RAZOR: Dan Ernst et. al., MICRO 2003.
Vdd Scaling and Process Tolerance: Conventional Solutions

- **Low power:**
  - Reduce the supply voltage
    - Error rate increases
  - Dual-Vt/dual-VDD assignment
    - Number of critical paths increases

- **Robustness:**
  - Increase supply voltage
    - Power dissipation increases
  - Upsize the gates
    - Switching capacitance increases

*Low power and robustness: conflicting requirements*
CRISTA: Basic Idea

• Important points:
  – Scale down the supply while making delay failures predictable
  – Avoid the failures by adaptive clock stretching
  – Ensure that critical paths are activated rarely
Design Considerations for CRISTA

- Few predictable critical paths
- Low activation probability of critical paths
- Slack between critical and non-critical paths under variations
Case Study: Adder

- Interesting features:
  - Single critical path (activated by $P_0 P_1 P_2 P_3 = 1$ & $C_{i,0} = 1$)
  - Low activation probability of critical path

$V_{DD} = 1V$, $T_{CLK} = 260ps$
$P = 13\mu W$ (1-cycle)

$V_{DD} = 0.8V$, $T_{CLK} = 260ps$
$P = 7.4\mu W$ (rare 2-cycles, decoder)

44% power saving by reducing voltage and, operating critical path at 2-cycle and other paths at 1-cycle

Can we apply same technique to any random logic?
Carry Select Adder

- ~20% power saving with ~6% area overhead
Carry Save Multiplier

Vector Merging Adder

Critical Path

Longest off-critical path

- 25% power saving with ~5% area overhead
Wallace Tree Multiplier

- Partial Products
- Full Adders
- Half Adder
- Vector Merging Adder

Stage 1
Stage 2
Stage 3

Vector Merging Adder
Final Product

- 29% power saving with ~4% area overhead
Simulation Results

Ripple Carry Adder

- ISO Yield = 92%

- Power savings: 20, 25, 30, 35, 40, 45
- Area overhead: 5, 6, 7, 8, 9, 10
- Bit sizes: 12 bits, 16 bits, 32 bits

Carry Save Multiplier

- ISO Yield = 90%

- Power savings: 0, 5, 10, 15, 20, 25
- Area overhead: 6, 7, 8, 9, 10
- Bit sizes: 12 bits, 16 bits, 32 bits

Wallace Tree Multiplier

- ISO Yield = 96%

- Power savings: 26, 27, 28, 29
- Area overhead: 2.5, 3.5, 4.5, 5, 5.5, 6.5, 7.5
- Bit sizes: 8 bits, 12 bits, 16 bits

Performance penalty (WTM)

- Throughput penalty: 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4
- Area Overhead: 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4
- Bit sizes: 6 bits, 8 bits, 10 bits, 12 bits, 20 bits
Random Logic: Shannon’s Expansion

\[ f(x_1, \ldots, x_i, \ldots, x_n) = x_i \cdot f(x_1, \ldots, x_i = 1, \ldots, x_n) + x'_i \cdot f(x_1, \ldots, x_i = 0, \ldots, x_n) \]

\[ = x_i \cdot CF_1 + x'_i \cdot CF_2 \]

\[ CF_1 = f(x_1, \ldots, x_i = 1, \ldots, x_n); \quad CF_2 = f(x_1, \ldots, x_i = 0, \ldots, x_n) \]

Activation probability of cofactors can be reduced

How to choose Control Variable?
Further Isolation and Slack Creation by Sizing

- Slack creation strategy
  - Lagrangian Relaxation based sizing (*B.C. Paul et. al., DAC 2004*) is used
  - Non-critical paths are selectively made faster
  - Critical paths are slightly slowed down
Simulation Results

MCNC benchmarks, 70nm Process

- Average power saving = ~50%
- Average area overhead = 18%
- Avg performance penalty = 5.9% (with 4 control variables) for signal prob = 0.5
Two-Stage Pipeline with Test Logic

Low Power Robust Pipeline

- Stalling Logic
- Pre-decoder
- SFFs
- Carry-Lookahead Adder
- Comparator
- SFFs
- Outputs

Conventional Pipeline

- Clock generator
- Regular pipeline

Power measurement of proposed pipeline

~40% power saving with ~13% performance

20% reduction in conventional pipeline
40% extra reduction using CRISTA

<table>
<thead>
<tr>
<th>Avg. Power [mW]</th>
<th>Voltage [V]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>
VDD Scaling, Process Variation, and Quality Trade-off: DCT

Banerjee, Karakonstantis, Roy
Design Automation and Test in Europe (DATE) 2007
Basic Idea

• All computations are “not equally important” for determining outputs

• Identify important and unimportant computations based on output “sensitivity”

• Compute important computations with “higher priority”

• Delay errors due to variations/ Vdd scaling “affect only” non-important computations

• “Gradual degradation” in output with voltage scaling and process variations
DCT Based Image Compression Process

- **DCT is used in current international image/video coding standards**
  - JPEG, MPEG, H.261, H.263
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>6</th>
<th>7</th>
<th>15</th>
<th>16</th>
<th>28</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td>8</td>
<td>14</td>
<td>17</td>
<td>27</td>
<td>30</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>13</td>
<td>18</td>
<td>26</td>
<td>31</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>19</td>
<td>25</td>
<td>32</td>
<td>41</td>
<td>45</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>24</td>
<td>33</td>
<td>40</td>
<td>46</td>
<td>53</td>
<td>55</td>
</tr>
<tr>
<td>21</td>
<td>23</td>
<td>34</td>
<td>39</td>
<td>47</td>
<td>52</td>
<td>56</td>
<td>61</td>
</tr>
<tr>
<td>22</td>
<td>35</td>
<td>38</td>
<td>48</td>
<td>51</td>
<td>57</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>36</td>
<td>37</td>
<td>49</td>
<td>50</td>
<td>58</td>
<td>59</td>
<td>63</td>
<td>64</td>
</tr>
</tbody>
</table>

High energy components (important outputs 75% energy)

Low energy components (less important outputs)

Can important components be computed with higher priority?
Design Methodology

(a) Input Block

(b) 1D- intermediate DCT outputs

(c) Transpose Memory

(d) Final DCT outputs

T.x^t

1D-DCT

Faster Computation

Slower Computation

Faster Computation

Slower Computation
Path Delays for 1D-DCT outputs
Proposed DCT under Vdd scaling

Proposed Design with high/low delay paths

Scaled Vdd: Longer paths under Vdd scaling

Important Computations

Longer Delays

Paths Not Computed

Only DC component

Extreme Scaled Vdd: Shorter paths affected

Vdd3 < Vdd2 < Vdd1 (nominal)
1D-DCT Path Delay Comparisons

Computation Paths

<table>
<thead>
<tr>
<th>Path</th>
<th>Delay(ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path1 (w0)</td>
<td>2.0</td>
</tr>
<tr>
<td>Path2 (w1)</td>
<td>2.5</td>
</tr>
<tr>
<td>Path3 (w2)</td>
<td>3.0</td>
</tr>
<tr>
<td>Path4 (w3)</td>
<td>3.5</td>
</tr>
<tr>
<td>Path5 (w4)</td>
<td>4.0</td>
</tr>
<tr>
<td>Path6 (w5)</td>
<td>3.5</td>
</tr>
<tr>
<td>Path7 (w6)</td>
<td>3.0</td>
</tr>
<tr>
<td>Path8 (w7)</td>
<td>2.5</td>
</tr>
</tbody>
</table>

- **Conventional DCT**
- **Proposed DCT**
### Effect of Vdd Scaling

#### Different Architectures at Nominal Voltage

<table>
<thead>
<tr>
<th>Voltage</th>
<th>Conventional WTM DCT</th>
<th>CSHM DCT (2 alphabet)</th>
<th>Proposed DCT</th>
<th>1.0V</th>
<th>CSHM DCT (2 alphabets)</th>
<th>DCT with WTM</th>
<th>Proposed DCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9 V</td>
<td>FAILS</td>
<td>FAILS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8 V</td>
<td>FAILS</td>
<td>FAILS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Proposed Architecture at Reduced Voltage

<table>
<thead>
<tr>
<th>Voltage</th>
<th>Proposed DCT Vdd=0.9V</th>
<th>Proposed DCT Vdd=0.8V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power (mW)</td>
<td>17.53(41.2%)</td>
<td>11.09(62.8%)</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>29</td>
<td>23.41</td>
</tr>
</tbody>
</table>

- Graceful degradation of proposed DCT architecture under Vdd scaling (Vdd can be scaled to 0.75V)
- Conventional architectures **fails**
Temporal Degradation: NBTI

Kang, Roy, et. al. – TCAD, DAC-07
Temporal Reliability Issues in CMOS Technology

- *HCI* – Hot Carrier Injection
- *NBTI* – Negative Bias Temperature Instability
  - Increase in $V_T$ of PMOS with time
  - The **dominant reliability factors** in scaled tech.

- *TDDB*, etc.
NBTI: Negative Bias Temperature Instability

Interface trap generation due to Si-H bond breaking

- Interface trap ($N_{IT}$) generation at the channel interface due to the Si-H bond breaking, when negative gate bias is applied
- With time, $V_T$ increases, subthreshold slope ($S$) increases, mobility degrades,
- Drive current ($I_{DS}$) reduces and affect the PMOS speed
- Overall reduces the lifetime of PMOS
PMOS $V_T$ degrades as a power of time due to NBTI
- **Fixed exponent of 1/6** matches the simulation data

* V. Huard and M. Denais, IRPS 2004
Power-law $V_T$ degradation Model

Distance into oxide

$N_H^{(0)}$

$N_H$

$\sqrt{D_H \cdot t}$

$0$

$y$

Reaction rate:

$$\frac{dN_{IT}}{dt} = k_F[N_0 - N_{IT}] - k_R N_{IT} N_H^{(0)} \approx 0$$

Conservation of hydrogen:

$$\frac{k_F}{k_R} \cdot [N_0 - N_{IT}] = N_{IT} \cdot N_H^{(0)}$$

$$N_{IT}(t) = \int_0^{\sqrt{D_H \cdot t}} N_H^{(0)}(y, t) \cdot dy$$

$$N_{IT} = \frac{1}{2} N_H^{(0)} \cdot \sqrt{D_H t}$$

$\Delta V_T = \frac{q \cdot \Delta N_{IT}}{C_{OX}}$

$N_{IT}(t) = \sqrt{\frac{k_F N_0}{2k_R} (D_H t)^{\frac{1}{4}}}$

NBTI degrades in time of exponent 1/6
Mobility degradation factor

- Mobility degradation due to NBTI is expressed in an additional $V_T$ shift, noted as $m$

- Overall temporal $V_T$ shift model is expressed as,

$$
\Delta V_T = (1 + m) \frac{q\chi}{C_{OX}} \sqrt{E_{ox} e^{\left(\frac{E_{ox}}{E_0}\right)}} \cdot t^{0.25}
$$
Impact of NBTI on circuit performance
• Performance (delay) degradation also follows the power trend with same 0.17 exponent
• In CMOS logic, only the rising (L2H) delay’s are affected
Circuit Performance degradation cont.

- Delay degradation in ISCAS c432
- Activity factor (switching activity) does not affect much on the delay degradation
  - In reality, activity factor’s are balanced in the normal operations
Design method considering the NBTI degradation
NBTI-aware design method

- **Over-design** is required to guarantee a lifetime stability of the circuit
- **LR sizing** is used to optimize the circuit
  - Size the circuit considering the worst-case $V_T$ degradation over the lifetime
LR Sizing considering NBTI

1. Delay Constraint ($D_{\text{MAX}}$)
2. Required Lifetime ($T_{\text{Life}}$)

Calibrate switching activity's ($S_i$) in each node

Compute $V_T$ shift in each node considering $S_i$

LR sizing with delay constraint $D_{\text{MAX}}$

NBTI-aware Design

Lifetime Constraint
A new design constraint

Power-law $V_T$ model

Optimal sizing from Lagrangian Relaxation*

Guarantee a lifetime stability under NBTI degradations

* C. Chen et. al., TCAD 1999
Simulation results

1. Delay degradation in ISCAS85 benchmark circuits after 10 years

<table>
<thead>
<tr>
<th>Circuit</th>
<th>No. of Trans.</th>
<th>Nominal delay (ps)</th>
<th>% delay degrad. (10 yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_i = 1$</td>
</tr>
<tr>
<td>c432</td>
<td>590</td>
<td>525</td>
<td>8.90</td>
</tr>
<tr>
<td>c499</td>
<td>1816</td>
<td>368</td>
<td>9.20</td>
</tr>
<tr>
<td>c1908</td>
<td>1582</td>
<td>513.5</td>
<td>9.18</td>
</tr>
<tr>
<td>c3540</td>
<td>3638</td>
<td>597.3</td>
<td>9.00</td>
</tr>
<tr>
<td>c74181</td>
<td>372</td>
<td>194.6</td>
<td>9.89</td>
</tr>
<tr>
<td>c74182</td>
<td>92</td>
<td>77.2</td>
<td>10.35</td>
</tr>
<tr>
<td>c74283</td>
<td>188</td>
<td>131.9</td>
<td>7.90</td>
</tr>
<tr>
<td>c74L85</td>
<td>148</td>
<td>115.1</td>
<td>9.50</td>
</tr>
</tbody>
</table>

* All benchmarks are synthesized in BPTM 70nm technology
Simulation results cont.

2. Area overhead in NBTI-aware sizing

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Nominal delay (ps)</th>
<th>Nominal area (um)</th>
<th>% Area overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$S_i = 1$</td>
</tr>
<tr>
<td>c432</td>
<td>385</td>
<td>196.7</td>
<td>14.8</td>
</tr>
<tr>
<td>c499</td>
<td>340</td>
<td>581.47</td>
<td>7.82</td>
</tr>
<tr>
<td>c1908</td>
<td>470</td>
<td>489.67</td>
<td>7.13</td>
</tr>
<tr>
<td>c3540</td>
<td>500</td>
<td>1146.5</td>
<td>3.44</td>
</tr>
<tr>
<td>c74181</td>
<td>180</td>
<td>111.1</td>
<td>9.45</td>
</tr>
<tr>
<td>c74182</td>
<td>80</td>
<td>31.1</td>
<td>11.3</td>
</tr>
<tr>
<td>c74283</td>
<td>125</td>
<td>66.71</td>
<td>10.0</td>
</tr>
<tr>
<td>c74L85</td>
<td>120</td>
<td>42.59</td>
<td>5.85</td>
</tr>
</tbody>
</table>

* All benchmarks are synthesized in BPTM 70nm technology*
Negative Bias Temperature Instability

- PMOS specific Aging Effect
- Generation of (+) traps
- Reaction-Diffusion (RD) model
- Time exponent ~ 1/6

\[ N_{IT}(t) = \sqrt{\frac{k_F N_0}{2k_R}} (D_H t)^{1/6} \]

\[ \Delta V_T = \frac{q \cdot \Delta N_{IT}}{C_{OX}} \]

*M. A. Alam, IEDM’03*
NBTI in Digital Circuits

**Logic Circuits**
- $f_{\text{MAX}}$ decreases $\downarrow$
- Timing failure with time

**Memory Circuits**
- Static Noise Margin (SNM) $\downarrow$
- Read & Write Stability $\downarrow$
- Parametric Yield $\downarrow$

Temporal $V_{\text{Th}}$ increase in PMOS affects critical performance factors of digital VLSI circuits.
### NBTI: Random Logic Circuits

#### Delay Degrad. STD cells

<table>
<thead>
<tr>
<th>Logic Cell</th>
<th>fanin</th>
<th>Delay (ps) t=0</th>
<th>Delay (ps) 3 years</th>
<th>Δ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV</td>
<td>1</td>
<td>13.77</td>
<td>16.77</td>
<td>21.8</td>
</tr>
<tr>
<td>NAND</td>
<td>2</td>
<td>16.86</td>
<td>19.88</td>
<td>17.9</td>
</tr>
<tr>
<td>NAND</td>
<td>3</td>
<td>19.57</td>
<td>22.45</td>
<td>14.8</td>
</tr>
<tr>
<td>NOR</td>
<td>2</td>
<td>17.26</td>
<td>21.89</td>
<td>26.8</td>
</tr>
<tr>
<td>NOR</td>
<td>3</td>
<td>23.80</td>
<td>30.19</td>
<td>26.9</td>
</tr>
</tbody>
</table>

#### Graph

- 8% $f_{\text{MAX}}$ decrease
- $n \sim 1/6$
- 3 years Lifetime

#### Summary

- ISCAS’85 Benchmark Circuits, PTM 65nm
- Gate delay: analytical delay model considering NBTI
- Circuit delay: NBTI-aware Static Timing Analysis (STA)
- Circuit $f_{\text{MAX}} \rightarrow$ time exponent $n \sim 1/6$
SNM degrades by more than 10% in 3 years

% SNM Degradation $\rightarrow$ time exponent $n \sim 1/6$

WM improves with time under NBTI
Design for Reliability under NBTI

- Gate Sizing applied to guarantee lifetime functionality of design
- **11.7% overhead for Cell-based sizing**
- **6.13% overhead for TR-based sizing**
  - 45% improvement in area overhead
  - Runtime complexity for TR-based sizing is identical to that of Cell-based sizing

**Simulation Setup**
- Synthesized in PTM 65nm
- 1/6 $V_{Th}$ degradation model
- 125°C Stress temperature
- 50% Signal Probability at PI’s
**I\textsubscript{DDQ} based NBTI Characterization**

**Microphotograph**

**Inverter Chain**

**Layout**

- Test Circuit Fabricated
- 1000 stage INV chain
- DC Stress signal @V\textsubscript{in}
- I\textsubscript{DDQ} measurement @GND

<table>
<thead>
<tr>
<th>Technology</th>
<th>CMOS 130nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Die Size</td>
<td>20 (mm\textsuperscript{2})</td>
</tr>
<tr>
<td>I/O Pin</td>
<td>209</td>
</tr>
<tr>
<td>T\textsubscript{ox}</td>
<td>1.6 (nm)</td>
</tr>
<tr>
<td>V\textsubscript{DD}</td>
<td>1.2 (V)</td>
</tr>
</tbody>
</table>
Correlation between $I_{DDQ}$ & $f_{MAX}$

- $D_M < 3\, \text{ms}, \ \text{Temp}=125\, \text{°C}, \ V_{stress}=1.7\, \text{V}$
- $I_{DDQ}$ degradation $\rightarrow n \sim 1/6$ during
- Clear signature of NBTI
- Correlation between $I_{DDQ}$ and $f_{MAX}$ can be used to predict circuit performance degradation under NBTI

% $I_{DDQ}$ decrease
% $f_{MAX}$ increase

$\Delta I_{leak}(t) \propto t^{1/6} \propto \Delta f_{MAX}(t)$

$R_{freq} = K \times R_{leak}$ ($K$ : constant)
**I\textsubscript{DDQ} based Characterization Technique**

- **Initial Characterization**
  - Compute $R_{\text{leak}}$, $R_{\text{fMAX}}$
  - Compute $K_{\text{fMAX}} = R_{\text{leak}} / R_{\text{fMAX}}$

- **Reliability Extrapolation**
  - For each I\textsubscript{DDQ} measurement sample, $R_{\text{leak}}$ is computed
  - $R_{\text{fMAX}} = K_{\text{fMAX}} \times R_{\text{leak}}$
  - Estimate $f_{\text{MAX}}$ degradation

- **Lifetime Projection**
  - Project I\textsubscript{DDQ} using $K_{\text{fMAX}}$

- **Temp. Dependency**

- **NBTI Characterization Report**

**Design phase**
- Circuit-level NBTI Reliability Characterization
- I\textsubscript{DDQ} test is used
- Expensive $f_{\text{MAX}}$ testing is avoided (or minimized)
- Accurate circuit level performance degradation can be predicted
- IC specific burn-in to qualify the target produce
- Efficient way of field monitoring: dynamic local signature of produce usage
- Possible usage in other reliability sources; HCI

**Post-silicon phase**
Conclusions

- Process Variation and Process Tolerance is becoming important
- There is a need to optimize designs considering power/performance/yield