Value-dependence of SRAM leakage in deca-nanometer technologies

Maziar Goudarzi\(^a\) and Tohru Ishihara

System LSI Research Center, Kyushu University
3–8–33 Momochihama, Sawara-ku, Fukuoka 814–0001, Japan
a) goudarzi@slrc.kyushu-u.ac.jp

Abstract: Within-die process variation increases with technology scaling in nanometer era. Due to uncorrelated random variations in the threshold voltage (\(V_{\text{th}}\)), neighboring transistors in a 6-T SRAM have different \(V_{\text{th}}\) and dissipate different subthreshold leakages. Since 3 transistors leak when the cell stores a 1 and the other 3 leak when it stores a 0, total cell leakage depends on its stored value. Using Monte Carlo simulations, we show that this difference averages 46\% at a variation of 58\% in \(V_{\text{th}}\). This phenomenon can be used to reduce leakage of SRAM-based memories by value control.

Keywords: process variation, within-die variation, SRAM, leakage, nanometer technology

Classification: Science and engineering for electronics

References

1 Introduction

Process variations are increasing with technology scaling. Variations in transistor parameters such as threshold voltage ($V_{th}$) are seen even within the same die, and their magnitude is increasing with every technology generation. A main cause behind such variations is the random dopant fluctuation (RDF) which differently affects even adjacent transistors and changes their $V_{th}$. Since subthreshold leakage ($I_{off}$, which is the dominant source of leakage in SRAM memories in nanometer technologies [1]) exponentially depends on $V_{th}$, much bigger variations are observed in $I_{off}$ following the variations in $V_{th}$. In the conventional 6-transistor SRAM cells, three of the transistors contribute to subthreshold leakage when storing a 1, and the other three contribute to leakage when storing a 0. Consequently, different $I_{off}$ leakages of the transistors results in different cell leakages when storing 0 compared to 1. This suggests a new opportunity in reducing SRAM leakage: SRAM leakage reduction by value-control. This can be done by (i) storing 0 (1) most of the time in a cell that leaks less when storing 0 (1) (e.g. by reordering code blocks in a scratchpad memory), and/or (ii) by invalidating and refilling highly leaky SRAM cells with the less leaky value when going to standby mode. Since share of leakage in total power consumption of SRAM-based memories is increasing with technology scaling [1, 2] and SRAM-based memories comprise the largest component of processor-based embedded systems (e.g. 70% of StrongARM [3]) and are among main sources of power consumption there [1], finding new opportunities for leakage reduction is important. Especially note that SRAM value-control can be done at system and software level without needing expensive changes at the circuit-level or manufacturing process.

Since RDF is a physical effect caused by the fact that transistor channel sizes are approaching atomic sizes, it cannot be resolved by external control of the manufacturing process, and hence, $V_{th}$ variations are expected to increase with technology scaling [4]. We show that the amount of potential reduction in SRAM leakage by value-control increases at higher $V_{th}$ variation, and consequently, the significance of value-control for reducing SRAM leakage increases with technology scaling.

To the best of our knowledge, no previous work has observed this leakage reduction opportunity. Monte Carlo simulations in Section 4 show that in a 16 KB memory leakage can be theoretically reduced by 56% (comparing full match to the full mismatch between the cell value and its less-leaky state) in a technology node with 58% within-die $V_{th}$ variation (which is estimated to happen in process technologies in 2012 [4]) and this continues to grow.

2 Within-die variation in threshold voltage

Ideally, all transistors that are designed to be similar should have the same $V_{th}$ after fabrication. However, variations in the manufacturing process make their $V_{th}$ different from one another. These variations can be decomposed into die-to-die and within-die variations. The die-to-die element results in
different $V_{th}$ changes across dies, but equally affects all transistors in the same die. The within-die element, however, results in different $V_{th}$ values for different transistors within the same die and even within the same SRAM cell.

Variations can also be categorized to systematic vs. random ones. Unlike random variation, the systematic one can be analytically determined based on factors such as the location of the transistor on the die. Finally, random variations can be spatially correlated or totally uncorrelated. We focus on random dopant fluctuations, RDF, which is an uncorrelated random effect and results in uncorrelated random $V_{th}$ variation. With technology scaling, the channel area of minimum-sized transistors quadratically decreases which in turn reduces the number of dopant atoms residing in the channel area. When only hundreds or tens of dopant atoms exist in the channel area (which is the case in nanometer technologies [5]), a small change in their number can substantially affect the threshold voltage. This effect has been long known [6] but has only recently become significant when approaching deca-nanometer technologies. The $V_{th}$ variability due to RDF is modeled by Gaussian distribution whose standard deviation is given by Eq. (1) [7]:

$$\sigma_{V_{th-wid}} = \frac{q}{C_{ox}} \sqrt{\frac{N_a \cdot W_{dm}}{3 \cdot L \cdot W}}$$

where $\sigma_{V_{th-wid}}$ is the standard deviation of within-die $V_{th}$ variation, $q$ is the electron charge, $C_{ox}$ is the oxide capacitance, $N_a$ is the substrate doping concentration, $W_{dm}$ is the maximum depletion width, and $L$ and $W$ are the effective channel length and width of the transistor respectively.

Latest ITRS estimations [4] (Table I) show that RDF is now the dominant source of variability in $V_{th}$ and will become the only main factor by 2012. It also predicts that $V_{th}$ variability will be over 100% by 2018.

**Table I.** ITRS estimations of $V_{th}$ variability [4].

<table>
<thead>
<tr>
<th>DRAM 1/2 Pitch (nm) (contacted)</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>% $V_{th}$ variability</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td>57</td>
<td>50</td>
</tr>
<tr>
<td>Doping Variability impact on $V_{th}$</td>
<td>24%</td>
<td>29%</td>
<td>31%</td>
<td>35%</td>
<td>40%</td>
</tr>
<tr>
<td>% $V_{th}$ variability</td>
<td>26%</td>
<td>29%</td>
<td>33%</td>
<td>37%</td>
<td>42%</td>
</tr>
<tr>
<td>Includes all sources</td>
<td>26%</td>
<td>29%</td>
<td>33%</td>
<td>37%</td>
<td>42%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>32</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>40%</td>
<td>40%</td>
<td>58%</td>
<td>58%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>112%</td>
<td>112%</td>
<td>112%</td>
</tr>
<tr>
<td>42%</td>
<td>42%</td>
<td>58%</td>
<td>58%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>81%</td>
<td>112%</td>
<td>112%</td>
<td>112%</td>
</tr>
</tbody>
</table>
3 Value-dependence of SRAM leakage

We focus on $I_{\text{off}}$ as the primary contributor to leakage in nanometer caches [1]. Figure 1 shows a 6-transistor SRAM cell storing a 1 logic value. Clearly, only M1, M2, and M5 transistors can leak in this state while the other three may leak only when the cell stores a 0 (note that bit-lines are precharged to supply voltage, $V_{\text{DD}}$). Process variation, especially in such minimum-geometry devices, causes each transistor to have a different $V_{\text{th}}$ and consequently different $I_{\text{off}}$ value, finally resulting in different subthreshold leakage currents when storing 1 and 0.

![Subthreshold Leakage Path](image)

**Fig. 1.** A 6-transistor SRAM cell with transistor sizes and leakage paths.

4 Amount of leakage difference

To quantify this effect, we used Monte Carlo simulation to simulate several similar memories under process variation. For each memory instance, maximum leakage difference was averaged once per cell and once per entire memory. We use the following notations to report potential leakage differences:

- $\text{leak}_0$: leakage power of the cell when storing 0.
- $\text{leak}_1$: leakage power of the cell when storing 1.
- $\text{low} = \min(\text{leak}_0, \text{leak}_1)$
- $\text{high} = \max(\text{leak}_0, \text{leak}_1)$

\[
\text{per-cell difference} = \frac{|\text{leak}_0 - \text{leak}_1|}{\text{high}} \quad (2)
\]

\[
\text{Per-memory difference} = \frac{\sum \text{high} - \sum \text{low}}{\sum \text{high}} \quad (3)
\]

Equation (2) gives leakage difference between less-leaky and more-leaky states of a single cell, while Eq. (3) gives, in the entire memory, the difference between the worst case (all cells storing more-leaky values) and the best case (all cells storing less-leaky values). $\text{Leak}_0$ and $\text{Leak}_1$ are computed as follows:
We considered both die-to-die and within-die variations in our experiments. Each memory is assumed to reside in a separate die. The $V_{th}$ of each transistor $j$ of die $i$ is calculated from Eq. (4) where $V_{th-nominal}$, $\Delta V_{th-d2d}$ and $\Delta V_{th-wid}$ represent nominal $V_{th}$, die-to-die $V_{th}$ variation and within-die $V_{th}$ variation respectively:

$$V_{th}(i,j) = V_{th-nominal} + \Delta V_{th-d2d}(i) + \Delta V_{th-wid}(i,j)$$  \hspace{1cm} (4)$$

In our experiments, the $V_{th-nominal}$ was assumed to be 320 mv (corresponding to a commercial 90 nm process). $\Delta V_{th-d2d}$ and $\Delta V_{th-wid}$ values are independent Gaussian-distributed random variables with 0 mean and $\sigma_{V_{th}}$ standard deviations respectively. Only one $\Delta V_{th-d2d}$ is used for all transistors of each die while separate $\Delta V_{th-wid}$ values are used for each transistor of each die. Furthermore, $\sigma_{V_{th-wid}}$ of each transistor depends on its $W$ and $L$ (see Eq. (1)). For example, at 58% within-die variation (i.e. $3 \times \sigma_{V_{th-wid}}/V_{th-nominal} = 0.58$), $\sigma_{V_{th-wid}}$ of minimum-sized transistor ($W = L = L_{min}$) is 61.87 mv which using Eq. (1) gives 30.94 mv, 41.25 mv, and 50.52 mv respectively for $\sigma_{V_{th-wid}}$ of “M3,M5,” “M2,M4,” and “M1,M6” in Fig. 1. The $\sigma_{V_{th-d2d}}$ was assumed 20 mv. After determining $V_{th}$ of each transistor (Eq. (4)), BSIM3 equations were used to compute individual transistor leakages which were appropriately summed up to obtain $Leak0$ and $Leak1$.

We Monte Carlo simulated 1000 16 KB memories and obtained the average leakage difference per-cell and per-memory given in Fig. 2 for within-die variation ranging from 20% to over 100%. Data points correspond to ITRS estimations in Table I.

![Fig. 2. Average leakage saving opportunity increases with $V_{th}$-variation.](image_url)
V_{th}), but the percentage of leakage differences (Eqs. (2) and (3) averaged over 1000 memories) remain invariant for a given \( \sigma_{V_{th-wid}} \). This makes sense since this leakage difference is enabled by the \( V_{th} \) variation, not the \( V_{th} \) average value.

Our experiments on other memory sizes (1, 2, 4, 8 KB) also give the same results when averaged over 1000 memory instances since this is a random effect which results in the same average given a large enough number of instances.

5 Conclusion

We showed that when approaching atomic sizes in deca-nanometer technologies, uncorrelated random \( V_{th} \) variations result in asymmetry in leakage of the conventional 6-T SRAM when storing 0 and 1. We quantified this leakage difference using Monte Carlo simulation and showed that increasingly more difference in leakage exists in SRAM-based memories based on the values stored in them. This value-dependence of SRAM leakage is important since it enables new techniques for SRAM leakage reduction by value-control.

Acknowledgments

This work is supported by VLSI Design and Education Center (VDEC), The University of Tokyo with the collaboration of STARC, Panasonic, NEC Electronics, Renesas Technology, and Toshiba. This work is also supported by Core Research for Evolutational Science and Technology (CREST) project of Japan Science and Technology Corporation (JST). We are grateful for their support.