Energy Centric Model of SRAM Write Operation for Improved Energy and Error Rates

Swaroop Ghosh
Computer Science and Engineering, University of South Florida, Tampa, FL-33647, sghosh@cse.usf.edu

Abstract—We propose an energy centric model of SRAM write operation. The model provides useful insights about energy and write error rates. We introduce the concept of intrinsic energy margin induced errors. The proposed model is employed for evaluating various write assist mechanisms and their potential in reducing the intrinsic memory error rates. We also demonstrate that this model can be used for optimizing energy of memory arrays.

I. INTRODUCTION AND MOTIVATION

Cache size is growing over the technology generations in order to boost the performance and maintain a constant miss rate in the era of multi-core design. In today’s microprocessor (both server and desktop) the amount of on-chip cache (typically 6T SRAM) is in the order of 50MB [1]. Naturally cache power is becoming an important component of total power consumption. Lowering minimum operating voltage (also known as “active Vmin”) is an effective technique to reduce the power consumption. Under this operating model, read and write assist techniques are employed to lower the Vmin beyond the limits set by process variations. Few examples of read and write assist techniques are wordline underdrive, bitcell voltage modulation (for read assist), negative bitline, WL boosting and supply voltage collapse (for write assist). The assist techniques suppress the read/write errors under process variation to extend Vmin scaling. Although effective, these techniques don’t consider the impact of energy on errors. This may result into cases where energy spent on assist mechanism is more than the energy benefit obtained from Vmin reduction.

In this paper, we provide a novel energy centric perspective of the memory operation. Specifically, we propose an analytical framework to model intrinsic energy of memory during write operation in presence of noise sources (e.g., process variations). The model is validated with detailed Hspice simulation in 22nm PTM [9]. We introduce the concept of intrinsic energy margin induced errors. We demonstrate that energy centric perspective can provide us crucial insights for maximizing the energy efficiency of memories and reducing error rates. The proposed model answers several non-trivial optimization questions such as (a) how energy and error rates are related? (b) what is the best assist mechanism to make right tradeoff between energy and error rates? (c) how to improve the error rate under a given energy budget or how to improve the energy for a given error rate? To the best of our knowledge this is the first treatment of SRAM optimization that incorporates energy perspective of the bitcell.

The paper is organized as follows. The modeling methodology is explained in Section 2. The intrinsic energy model is proposed in Section 3. Simulation results along with comparison of existing write assist techniques are also described in this section. Energy optimization methodology under a target write error rate is presented in Section 4. The conclusions are drawn in Section 5.

2. MODELING METHODOLOGY

In this section, first we present the intrinsic energy model of SRAM for write operation. Next we discuss the mathematical background for modeling of intrinsic energy and write speed.

2.1 Intrinsic Energy Model:

Fig. 1 Energy centric model of SRAM (a) intrinsic energy barrier in absence of noise and, (b) intrinsic energy barrier in presence of noise (e.g., PVT, aging and thermal noise). Intrinsic energy barrier increases due to noise making the bitcell hard to write.

In the intrinsic energy model of SRAM, the memory or storage element contains two states separated by an intrinsic energy barrier (E_{int}) as shown in Fig. 1(a). The barrier appears as a result of internal node capacitance and feedback mechanism (that is inherent in SRAM bitcell). Note that E_{int} is the barrier in absence of external noise. Write operation is performed by supplying sufficient energy to allow the bit to overcome its intrinsic barrier and jump to the other state. Read is performed by supplying sufficient energy to the bitcell in order to develop differential at the senseamp input that can be resolved correctly. The presence of noise sources (e.g., process variations, aging, thermal noise, voltage and temperature variations) impacts the memory operations due to fluctuations in energy barrier between memory states. Fig. 1(b) shows the worst case effective intrinsic energy barrier in presence of noise. Due to elevated energy barrier, some bits would require more energy to complete successful write operation. In other words, insufficient amount of write energy would result into write failures. In section 3, we propose analytical model for intrinsic energy barrier in presence of variations. In this analysis, we have only considered noise due to process variations.

2.2 Key Mathematical Background:

Let us consider a function y=f(x_{1},...,x_{n}) where x_{1},...,x_{n} are independent Gaussian random variables with mean \mu_{1},...\mu_{n} and standard deviation (STD) \sigma_{1},...\sigma_{n}. The mean (\mu_{y}) and the STD (\sigma_{y}) of the random variable y can be estimated by using multivariate Taylor-series expansion [12] as follows:

\begin{align}
\mu_{y} &= \sum_{i=1}^{n} \mu_{i} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial f(x_{i},...,x_{n})}{\partial x_{i}} \mu_{j} \\
\sigma_{y}^{2} &= \sum_{i=1}^{n} \left[ \frac{\partial f(x_{i},...,x_{n})}{\partial x_{i}} \right]^{2} \mu_{i}^{2} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial f(x_{i},...,x_{n})}{\partial x_{i}} \mu_{j} \sigma_{i}^{2} + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\partial f(x_{i},...,x_{n})}{\partial x_{i}} \sigma_{i}^{2} \left[ \frac{\partial f(x_{i},...,x_{n})}{\partial x_{j}} \right]^{2} \mu_{j}^{2}
\end{align}

The probability of (y>Y_{0}) for a Gaussian probability distribution function (PDF) \[N(y;\mu_{y},\sigma_{y})\] is given by

\[P(y > Y_{0}) = \frac{1}{\sqrt{2\pi} \sigma_{y}} \int_{Y_{0}}^{\infty} N(y;\mu_{y},\sigma_{y}) dy = 1 - \frac{1}{\sqrt{2\pi} \sigma_{y}} N_{y}(y)dy\]

The mean and STD of N Gaussian random variables with mean \mu_{y} and STD \sigma_{y}, is given by

\[\mu = N \mu_{y}; \sigma = \sigma_{y} \sqrt{N}\]

3. WRITE OPERATION MODEL

In this section, we employ the mathematical background presented above for modeling of intrinsic energy, write time and intrinsic energy induced error rate under process variations.
Write bit error rate
Write Energy (fJ)
Error Rate

VDD=1.1
VDD=1.0
VDD=0.9
VDD=0.85

(a) Schematic of SRAM during write operation (b) timing diagram showing two step write process

Intrinsic energy and write time (Vcc Collapse)
Energy (fJ)
Voltage (V)

Write time (pS)
Voltage (V)
Intrinsic energy
Write time

3.1 Alpha Power Law Current Model:

We employ alpha power law model [10] to find the drain current through individual transistors of the SRAM cell. According to this law, the drain current $I_D$ is given by

$$I_D = \begin{cases} 
0 & (V_{GS} \leq V_{TH}: \text{cutoff region}) \\
\beta_D I_D \left( \frac{V_{GS}}{V_{DD} - V_{TH}} \right)^\alpha & (V_{GS} < V_{DD} - V_{TH}: \text{triode region}) \\
\beta_D I_D \left( \frac{V_{TH}}{V_{DD}} \right)^\alpha & (V_{GS} \geq V_{DD} - V_{TH}: \text{pentode region}) 
\end{cases}$$

where

$$I_D = I_D^0 \left( \frac{V_{GS}}{V_{DD} - V_{TH}} \right)^\alpha \quad \text{and} \quad V_D = V_D \left( \frac{V_{GS} - V_{TH}}{V_{DD} - V_{TH}} \right)^{\alpha/2}$$

The values of $\alpha$, $V_{DD}$, $I_D^0$, $V_{TH}$ for pmos and nmos are extracted from 22nm PTM [9] devices of size 1u/0.03u.

3.2 Intrinsic Energy Barrier Model during Write:

Intrinsic energy (during write) is the amount of energy that is needed to charge one side of the cell all the way from 0 to $V_{DD}$ and energy supplied when the other side is discharged from $V_{DD}$ to 0. Alternatively, it can also be thought of as the energy that is supplied to discharge one side of the cell to the trip point of the inverter ($V_{trip}$) that is driving the other side (Fig. 2(a)). Beyond this point, the feedback mechanism kicks-in and cell changes state. From Fig. 2(a), it can be observed that there are two contention mechanisms associated in this operation: (i) contention from P0 and (ii) contention from N1. Finally the internal node needs to be charged beyond $V_{trip}$ to complete the switching. Note that the intrinsic energy is infinite if node n0 fails to reach the trip point under DC condition. The write time can be divided into two steps: (a) $T_{trip}$ which is time required by node n0 to reach the trip point and (b) $T_{VDD}$ which is the time required by node n1 to charge from trip point to approximately $V_{DD}$. Fig. 2(b) illustrates these two components. If node capacitance is $C_n$, then $T_{trip}$ is given by

$$T_{trip} = \int_{0}^{V_{trip}} \frac{C_n dV}{I_{D0} - I_{D0} - I_{D0}}$$

In the above expression $V_{trip}$ is obtained by solving following equation

$$I_{P1}(V_G = V_D = V_{trip}, V_S = V_{DD}) + I_{X1}(V_G = V_D = V_{DD}, V_S = V_{trip}) = I_{N0}(V_G = V_D = V_{trip}, V_S = V_{DD})$$

(8)

The time required for P1-N1 to charge from $V_{trip}$ to $V_{DD}$ is given by

$$T_{trip, VDD} = \int_{V_{trip}}^{V_{DD}} \frac{C_n dV}{I_{P1} + I_{X1} - I_{N1}}$$

(9)

The energies supplied at node n0 and n1 are given by

$$E_{cnt,n0} = V_{DD} \int_{0}^{T_{trip}} I_{D0} dt + V_{DD} T_{VDD, trip} I_{D0} dt$$

(10)

The total energy ($E_{int}$, Fig. 1(a)) and write time ($T_{WR}$, Fig. 2(b)) is given by

$$E_{int} = E_{cnt,n0} + E_{cnt,n1} \text{ if } V(n0) > V_{trip}$$

$$T_{WR} = T_{trip} + T_{VDD, trip} \text{ if } V(n0) > V_{trip}$$

(11)

(12)

The value of $V(n0)$ in (11) and (12) is determined by solving KCL at node n0 and n1 in self consistent manner. In this work, we only consider the case when $V(n0) > V_{trip}$ under DC condition which is a reasonable assumption otherwise there will be hard functional failures. The success of write operation depends on providing sufficient amount of energy to overcome $E_{int}$ within the allocated write time $T_{WR}$.

The framework for computing write intrinsic energy and write time is implemented in MATLAB and the results are compared against Hspice simulation. Fig. 3(a)-(b) compares the mean of write intrinsic energy and write time obtained from model and simulation. The error is found to be approximately 10%. We believe that the source of error is the mismatch in node n0/n1 capacitance that can be tuned further to improve the model accuracy.
It is interesting to observe from Fig. 3(a) that the intrinsic energy barrier reduces with global supply voltage—sug- gesting that the cell becomes easier to write at lower voltages. This is a simple yet intuitive result that provides insight on write assist mechanisms (supply voltage scaling in this case) from energy perspective. The write time increases with reduction in \( V_{DD} \) indicating that although the energy barrier has reduced, overcoming that barrier takes additional time due to less flow of electrons.

### 3.3 Intrinsic and Error Rates under Process Variation:

The mean and STD of write intrinsic energy is estimated by using (1) and (2) in presence of process variation. The effect of process variation is modeled as \( V_{TH} \) variation. Variation in channel length, width, oxide thickness is lumped in the \( V_{TH} \) variation model. The mean and STD of \( V_{TH} \) variation for each of the transistors in 6T SRAM is assumed to be (0, 30mV). The mean and STD is computed by sweeping \( V_{TH} \) of one transistor at a time and using (1) and (2). There are two sources of error during write operation (a) insufficient energy to overcome the intrinsic energy barrier. We refer to this type of error as “intrinsic energy error” and, (b) insufficient time to complete the write operation. We refer to this error as “write time error”. These two error types are illustrated in Fig. 4 where \( T_{write} \) is maximum time allocated by operating frequency to complete the write and \( E_{write} \) is minimum energy needed to overcome the intrinsic energy barrier. In this paper, these two quantities have been assumed to be independent of each other for the sake of simplicity. However, it is possible to determine correlation between them for a more accurate modeling of write error rates. The intrinsic energy and write time error rates are computed by using (3). Fig. 5(a) shows the intrinsic energy error rate wrt write energy for a single bit at different supply voltages. It can be observed that write energy rate can be reduced by providing more energy to the bitcell. The mean and STD of array energy write error rate is computed by using (4) and the results for 32 Byte memory write access is illustrated in Fig. 5(b).

### 3.4 Evaluation of Write Assist Mechanisms:

In order to gain further insight on the efficacy of assist mechanisms in reducing the intrinsic write barrier, we evaluated two popular write assist techniques namely column based supply voltage collapse (Vcccol) [11] and negative bitline (NBL) [4-5].

Fig. 6 shows the results for Vcccol collapse assist technique. It can be observed from Fig. 6(a) that intrinsic energy decreases at lower cell voltages with increase in write time. The magnitude of intrinsic energy reduction between \( V_{DD} \) scaling and Vcccol is identical however, the write time is better with Vcccol (Fig. 6(b)) making it a better candidate for write assist in frequency constrained systems.

Fig. 7(a) shows the evaluation of NBL assist. It is interesting to observe that this technique is not so effective in lowering the energy barrier however, it improves the write speed with higher amount of assist. This is in contrast with Vcccol that worsens the write speed with higher assist. Therefore, the speed dominated write failures get corrected after application of NBL assist both due to lower energy barrier and improved speed. Fig. 7(b) compares effectiveness of NBL and Vcccol in lowering intrinsic energy barrier. It can be noted that Vcccol can outperform NBL in terms of intrinsic energy error rate. However, it is poor in fixing the speed dominated errors where NBL can be useful (Fig. 7(c)). Fig. 8 compares Vcccol and NBL in terms of write energy and write time error rates. For the sake of simplicity, graphs are plotted at maximum assist condition for both techniques (i.e., \( V_{CC}=0.85\text{V} \) for Vcccol and \( V_{BL}=-0.3\text{V} \) for NBL). As noted before, NBL is effective in suppressing the timing errors whereas Vcccol is effective in suppressing intrinsic energy induced errors. At a fixed error rate (10\(^{-7}\)), NBL can reduce write time by \( \approx 13\% \). Under the same conditions, Vcccol can reduce intrinsic energy barrier by \( \approx 16\% \).

### Table-1 Parameters used for write energy optimization

<table>
<thead>
<tr>
<th>Param</th>
<th>Value</th>
<th>Param</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array dim</td>
<td>512Rx1024C</td>
<td>C(BL)</td>
<td>75fF</td>
</tr>
<tr>
<td>Col mux</td>
<td>4:1</td>
<td>Vcccol swing</td>
<td>0.25V</td>
</tr>
<tr>
<td>C(Vcccol)</td>
<td>150fF</td>
<td>Bitcell dim(P.:N:AX)</td>
<td>0.1u:0.2u:0.2u</td>
</tr>
<tr>
<td>WR Activity Factor</td>
<td>1%-100%</td>
<td>Leakage condition</td>
<td>110C</td>
</tr>
</tbody>
</table>

This table provides the parameters used for write energy optimization, including array dimensions, column mux, capacitance, swing, bitcell dimension, write activity factor, and leakage condition. These parameters are crucial for optimizing the write energy of memory cells under various conditions.
4. WRITE ENERGY OPTIMIZATION

In the previous section, we provided intrinsic energy model and compared two assist mechanisms in terms of write error reduction. In this section, we apply the proposed model for write energy optimization of memory array. In order to study the write energy optimization, we consider two scenarios: (a) assuming energy overhead from assist mechanism to be negligible (which is the ideal case) and (b) with appropriate energy estimates of assist circuitry. In the following paragraphs, we describe both of the above conditions. Furthermore, we consider Vcc collapse as an example assist mechanism for optimizations.

4.1 Experimental Setup:

For energy estimation, we have assumed the memory array parameters described in Table-1. The target write time and write energy is chosen by fixing the bit error rate (Fig. 9(a)-(b)). Although the resulting target values (84ps, 0.8fJ) are pessimistic, it still provides vital insights on energy optimization. Note that this methodology is different than conventional methodology that is purely based on write time error rate and may result into a sub-optimal solution. For fair estimation of array level energy tradeoff, we have considered write energy as well as retention energy. Read energy is not considered because the optimization for write is assumed to be independent of real read operation (ignoring dummy read). Column based Vcc collapse [5,11] is assumed for write assist. Activity factor of 1% to 100% for write operation is considered for fair evaluation of write assist energy overhead.

4.2 Without Energy Overhead of Assist:

In absence of energy overhead from assist circuitry, the energy barrier of the bitcell could be reduced freely by scaling the supply voltage and collapsing the bitcell voltage by an extra 250mV (Table-1). Fig. 10(a) plots the write energy, retention and total energy consumption with respect to write time. It can be observed that ~3X reduction in total energy is possible for the target delay of 84ps.

4.3 With Assist Energy Overhead:

At 100% activity factor (i.e., write access every cycle), the energy overhead from assist circuitry becomes dominant component of total energy consumption. It can be noted from Fig. 10(b) that using assist with 100% activity doesn’t make sense from energy perspective. At low activity factors, both assist and write energy reduces significantly but leakage energy of idle bitcells starts dominating. Fig. 10(c) illustrates the energy optimization with respect to activity factor and Vcc collapse. It can be observed that assist provides energy savings only when write activity factor is very low (which can be a fair assumption in real operation). Fig. 11 depicts the total energy with respect to activity factor and assist. The energy consumption of SRAM without any assist mechanism is ~500fJ (from Fig. 10(a)). Therefore, in order to make assist technique attractive, the activity factor needs to be restricted below 10%. At 1% activity factor, the energy consumption at optimal point is 110fJ which is ~5X better than nominal design (i.e., without assist).

5. CONCLUSIONS AND FUTURE WORK

We proposed an intrinsic energy model of SRAM write operation. The model provides useful insight towards understanding assist mechanisms in terms of their ability to improve intrinsic error rates and energy consumption. We investigated two commonly used assist techniques (Vcc collapse and negative bitline) and compared them in terms of intrinsic energy induced error and write timing errors. We also studied array level energy optimization with the aid of proposed energy centric model and drew solid conclusions for energy optimizations. The future work will involve developing intrinsic energy models for read and retention modes of operation.

ACKNOWLEDGEMENTS

The author would like to acknowledge Dr. Arijit Raychowdhury for initial discussions and useful inputs.

REFERENCES