Modeling and Estimation of Failure Probability due to Parameter Variations in Nano-scale SRAMs for Yield Enhancement

Saibal Mukhopadhyay, Hamid Mahmoodi-Meimand, and Kaushik Roy
Dept. of Electrical and Computer Engineering, Purdue University, West Lafayette, IN-47907, USA
<smukhop, mhammoodi, kaushik@ecn.purdue.edu>

ABSTRACT

In this paper we have analyzed and modeled the failure probabilities (access time failure, read/write stability failure, and hold stability failure in the standby mode) of SRAM cells due to process parameter variations. A method to predict the yield of a memory chip designed with a cell is proposed based on the cell failure probability. The developed method can be used in the early stage of a design cycle to optimize the design for yield enhancement.

1. INTRODUCTION

Increasing inter-die statistical variations in the process parameters (channel length ($L$), width ($W$), and transistor threshold voltage ($V_{th}$)) has emerged as a serious problem in the nano-scaled circuit design [1]. The inter-die parameter variations, coupled with the intrinsic on-die variation in $V_{th}$ due to random dopant fluctuation, can result in failure of SRAM cells [1]. A cell failure can occur due to: (a) an increase in the cell access time (access time failure), (b) unstable read/write operations (read/write stability failure), or (c) failure in the data holding capability of the cell at a lower supply voltage (hold stability failure in the standby mode). A failure in any of the cells in a column (or row) of the memory will make that column (or row) faulty. If the number of faulty columns (or rows) in a memory chip is larger than the number of redundant columns (or rows), then the chip is considered to be faulty. Hence, the failure probability of a cell is directly related to the yield of a memory chip. Consequently, estimation of the failure probability for a cell is necessary in the design phase to ensure a good yield. In this paper, we have developed a method to predict the yield of a memory chip under inter-die variation of $L$, $W$, and $V_{th}$ by estimating the failure probability of the cell considering intra-die parameter variation. The method is developed considering intra-die $V_{th}$ variation (principally due to random dopant fluctuation) and can be extended to include on-die random variation in $L$ and $W$. In particular, we have

1. Modeled the access time, the read stability, the write stability, and the hold stability failures of a cell due to process variation.

2. Developed a method to estimate the failure probability of a memory and to predict the yield of a memory chip.

3. Presented a statistical analysis of the impact of circuit (transistors sizing) and architecture ($k$ of rows and columns) on the cell failure probability and memory yield.

2. DEVICE CHARACTERISTICS

In our SRAM cell (Fig.1) we have used transistors of 50nm gate length ($L_{eff}=25nm$) designed using MEDIC [2], [3]. In our analysis, we have used the short channel MOSFET theory to model the currents and threshold voltage considering the device geometry and doping profile [3], [4]. Fig. 2 shows the Id-Vg characteristics of the designed transistors.

3. ESTIMATION OF CELL FAILURE PROBABILITY ($P_f$) and YIELD PREDICTION

A failure in an SRAM cell can occur due to: (a) an increase in the access time of the cell resulting in a violation of delay requirement, defined as access time failure, (b) destructive read (read failure) and/or unsuccessful write (write failure), resulting in a dynamic stability failure (Fig. 3) and, (c) the destruction of the cell content in the standby mode with the application of a pre-specified (designed) lower supply voltage ($V_{th,0}$), known as hold-stability failure (Fig. 3). In a die, failures are principally caused by the mismatch in the device parameters ($L$, $W$, $V_{th}$) of different transistors (intra-die) in the cell. Such device mismatch modifies the strength of the different transistors resulting in different failure events.

3.1. Distribution of the Intrinsic $V_{th}$ Variation ($\sigma_{V_{th}}$)

To estimate the failure probability, the threshold voltage ($V_{th}$) of the cell transistors are considered as six independent random variables (RV) [1]. The probability distribution function (pdf) of $V_{th}$ fluctuation ($\delta V_{th}$) of each transistor is assumed to be Gaussian (mean=0). The standard deviation ($\sigma_{V_{th}}$) depends on the manufacturing process, doping profile, and the transistor size. In the proposed method, $\sigma_{V_{th}}$ for a minimum sized transistor ($\sigma_{V_{th}}$) is an input parameter and the dependence of $\sigma_{V_{th}}$ on the transistor size is given by [4]:

$$\sigma_{V_{th}} = \sigma_{V_{th}}(L_{min}, L_{max}/W)$$

(1)

3.2. Access Time Failure ($P_{AF}$)

The cell access time ($T_{ACCESS}$) is defined as the time required to produce a pre-specified ($A_{MIN}, V_{th,0}$) voltage difference between two bit-lines (bit-differential). If due to $V_{th}$ variation, the access time of the cell is higher than the maximum tolerable limit ($T_{MAX}$), an access time failure is said to have occurred. The probability of access time failure ($P_{AF}$) of a cell is given by:

$$P_{AF} = P(T_{ACCESS} > T_{MAX})$$

(2)

While reading the cell storing $V_r = '1$ and $V_w = '0$ (Fig. 1, Fig. 3), bit-line BR will discharge through $A_1$ and $B_2$ (by the current $I_{BR}$). Simultaneously, BL will discharge by the gate leakage, the subthreshold leakage, and the junction leakage of $A_1$ of all the cells $I_{BL}$ connected to BL. The discharging currents $I_{BR}$ and $I_{BL}$ are given by:

$$I_{BR} = \sum_{i=1}^{W} I_{safe} + \sum_{i=1}^{N} I_{safe}(i) + I_{safe}(i)$$

(3a)

$$I_{BL} = \sum_{i=1}^{W} I_{safe} + I_{safe}(i) + I_{safe}(i)$$

(3b)

Where, $N$ is the #of cells attached to a bit-line (or column). Hence,
The access time given by (5) closely follows the MEDICI simulation result (Fig. 4a). The access time $T_{ACCESS}$ is constant (small $\Delta V_{BR}$), and $I_{\text{sub}}$ and $I_{d}$ are linear functions of time (Fig. 3a). We further assume that $C_{BR} = C_{\text{BL}}$, $C_{\text{AXL}} = C_{\text{AXL}}$, and $I_{\text{AXL}} = I_{\text{AXL}}$ (since they are not a strong function of $V_{T}$), and the variance obtained using (14) closely matches the MEDICI simulation result

$$T_{ACCESS} = \frac{C_{BR}C_{\text{AXL}}}{C_{BL}I_{d}}(V_{W} - V_{R})$$

$$\sigma_{T_{ACCESS}}^{2} = \frac{\left(\frac{\partial T_{ACCESS}}{\partial T_{AXL}}\right)^{2}}{2} \sigma_{T_{AXL}}^{2}$$

where $\sigma_{T_{AXL}}$ is the standard-deviation and $V_{R}$ is the mean of the $\Delta V_{BR}$ distribution of the access transistors $A_{X_{L}}$. The derivatives can be estimated numerically. The distribution using (6) closely matches the exact distribution (Fig. 4b). Using the derived pdf ($P_{\text{ACCESS}}(T_{ACCESS})$), $P_{W}$ can be estimated as:

$$P_{W} = \int_{T_{ACCESS}}^{T_{MAX}} f_{\text{ACCESS}}(t) dt = 1 - F_{\text{ACCESS}}(T_{MAX})$$

where, $F_{\text{ACCESS}}$ is the cumulative distribution function (cdf) of a Gaussian pdf. $P_{W}$ of a cell using the model closely matches the one obtained using Monte-Carlo simulation (Table-1).

### 3.3. Read Stability Failure ($P_{R}$)

While reading the cell shown in Fig. 1 ($V_{W} = V_{R} = V_{DD}$), $V_{R}$ increases due to the voltage divider action of $A_{X_{L}}$ and $A_{H}$ to a positive value $V_{READ}$. If $V_{READ}$ is higher than the trip point of the inverter $P_{R} = P_{T_{RIP}}(V_{R})$ then the cell flips after reading the cell (Fig. 3(a)) [7]. This represents a read failure ($R_{F}$) event. Hence, the read-failure probability ($P_{R}$) is given by:

$$P_{R} = P\left[V_{READ} > V_{T_{RIP}}\right]$$

$V_{READ}$ can be obtained by solving KCL simultaneously at node $R$ and $L$ as given by:

$$I_{R} = I_{\text{AXL}} + I_{\text{AXH}} + I_{\text{AXL}} + I_{\text{AXH}} + I_{\text{AXL}} + I_{\text{AXH}}$$

Neglecting the leakage currents to node $R$, we have [7]:

$$P_{R} = P\left[V_{READ} > V_{T_{RIP}}\right]$$

Similarly, $V_{T_{RIP}}$ can be obtained by solving [7]:

$$I_{\text{AXL}}(V_{T_{RIP}}) = I_{\text{AXH}}(V_{T_{RIP}})$$

Fig. 5 shows that $V_{T_{RIP}}$ (obtained using (11) and MEDICI simulation) is a linear function of independent RV's $\Delta V_{BR}$ and $\Delta V_{BL}$. Similarly, $P_{R}$ (obtained using (10) and MEDICI simulation) is a linear function of independent RV's $\Delta V_{BR}$ and $\Delta V_{BL}$. Hence, the pdf of $V_{T_{RIP}}$ and $P_{R}$ can be approximated as Gaussian distributions (since $\Delta V$s are Gaussian) with the means the variances obtained using (6) (Fig. 6a, 6b). $P_{R}$ is given by:

$$P_{R} = P\left[Z = (V_{T_{RIP}} - V_{T_{RIP}}) > 0\right]$$

where, $\mu_{Z} = \sigma_{\Delta V_{BR}} = \sigma_{\Delta V_{BL}}$ and $\sigma_{Z} = \sigma_{\Delta V_{BR}}^2 + \sigma_{\Delta V_{BL}}^2$.$P_{R}$ closely follows the values obtained using Monte-Carlo simulations (Table-1).
Figure 7: Variation and distribution of $T_{\text{WGE}}$ with variation in $\sigma_{\text{VT}}$ distribution function. Hence, to improve the accuracy of the model

Figure 8: Variation and distribution of $V_{\text{DDH}}$ in (a) $\sigma_{\text{VT}}$ applied in the directions: $(\delta V_{\text{t}}, \delta V_{\text{t}}<0, \delta V_{\text{t}}<0, \delta V_{\text{t}}>0)$ at the tail region, we can use a non-central $F$ distribution [6]. Using the pdf (Gaussian/non-central $F$) of $T_{\text{WGE}}$, the $P_{\text{HF}}$ is given by:

$$P_{\text{HF}} = \int_{T_{\text{WGE}}} (t_{\text{HF}}) d(t_{\text{HF}}) = 1 - F_{\text{HF}}(T_{\text{WGE}})$$

(15)

$F_{\text{HF}}(t_{\text{HF}})$ represents the cdf of the probability distribution (Gaussian/non-central $F$) [6]. $P_{\text{HF}}$ obtained using (15) closely matches the result using Monte-Carlo simulations (Table-I).

3.5. Hold Stability Failure ($H_s$)

In the standby mode, the $V_{\text{DDH}}$ of the cell is reduced to lower the leakage power consumption. However, if the lowering of $V_{\text{DDH}}$ causes the data stored in the cell to be destroyed, then cell is said to have failed in the hold-mode [8] (Fig. 3c). Hence, for a hold-failure event, the minimum supply voltage that can be applied to the cell in the hold-mode ($V_{\text{DDHmin}}$), without destroying the data, is higher than the designed stand-by mode supply voltage ($V_{\text{DDH}}$). Thus, the probability of hold-stability failure ($P_{\text{HF}}$) is given by:

$$P_{\text{HF}} = P[V_{\text{DDHmin}} > V_{\text{DDH}}]$$

(16)

Lowering the $V_{\text{DDH}}$ of the cell (say $V_{\text{DDH}}$ represents the cell $V_{\text{DDH}}$ at the hold mode) reduces the voltage at the node storing '1' ($V_{\text{L}}$) in Fig. 1). Due to leakage of $N_{\text{c}}$, $V_{\text{L}}$ will be less than $V_{\text{DDH}}$ for low $V_{\text{DDH}}$. The hold-failure occurs if $V_{\text{L}} < V_{\text{DDH}}$. Hence, $V_{\text{DDHmin}}$ can be obtained by solving:

$$V_{\text{DDHmin}} = V_{\text{DDH}}(V_{\text{DDHmin}}, V_{\text{DDHmin}}, V_{\text{DDHmin}}) = V_{\text{DDH}}(V_{\text{DDHmin}}, V_{\text{DDHmin}}, V_{\text{DDHmin}})$$

(17)

The estimated value of $V_{\text{DDHmin}}$ closely follows the values obtained from MEDICI simulation (Fig. 8a). From (17), it is evident that $V_{\text{DDHmin}}$ is a function of RVs $\delta V_{\text{tD}}, \delta V_{\text{tD}}, \delta V_{\text{pD}}$, and $\delta V_{\text{pD}}$. The distribution of $V_{\text{DDHmin}}$ ($V_{\text{DDHmin}}(V_{\text{DDHmin}}, V_{\text{DDHmin}}, V_{\text{DDHmin}})$) can be approximated as a Gaussian one with mean and variance obtained using the procedure described in (6) (a non-central $\chi^2$ distribution improves the accuracy for $V_{\text{DDHmin}}$ close to 0) (Fig. 8b). Hence, we can estimate $P_{\text{HF}}$ as:

$$P_{\text{HF}} = \int_{V_{\text{DDHmin}}} (V_{\text{DDHmin}}) d(V_{\text{DDHmin}}) = 1 - F_{\text{DDHmin}}(V_{\text{DDHmin}})$$

(18)

The $P_{\text{HF}}$ obtained using (18) closely matches the result using Monte-Carlo simulations (Table-I).

3.6. Estimation of Overall Failure Probability (PF)

The overall failure probability is given by:

$$P_{\text{F}} = P[F_{\text{F}}] = P[F_{\text{R}} + F_{\text{W}} + F_{\text{H}}] = P_{\text{RF}} + P_{\text{WF}} + P_{\text{HF}}$$

(19)

To simplify the estimations of the probabilities of the joint events, let us consider Table-II. It shows the probability of each different transistor to have the same failure in $\delta V_{\text{t}}$ and $\delta V_{\text{t}}$ or $\delta V_{\text{D}}$ or $\delta V_{\text{p}}$. For example, the joint event $(F_{\text{R}}F_{\text{W}})$ can be computed using similar arguments. We have also assumed that probabilities of simultaneous occurrence of more than two events are negligible $= 0$. The estimated probabilities match the Monte-Carlo results very closely (Table-II). All of the different failure probabilities increase significantly with an increase in the sigma of $\delta V_{\text{t}}$ variation (Fig. 10).

### Table-I: Failure Probability Estimations for Different Cells (MonteCarlo / Estimation)

<table>
<thead>
<tr>
<th>Cell</th>
<th>$P_{\text{F}}$</th>
<th>$P_{\text{RF}}$</th>
<th>$P_{\text{WF}}$</th>
<th>$P_{\text{HF}}$</th>
<th>$P_{\text{RF}}$</th>
<th>$P_{\text{WF}}$</th>
<th>$P_{\text{HF}}$</th>
<th>$P_{\text{RF}}$</th>
<th>$P_{\text{WF}}$</th>
<th>$P_{\text{HF}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.0019 / 0.159</td>
<td>0.005 / 0.005</td>
<td>0.007 / 0.006</td>
<td>0.008 / 0.009</td>
<td>0.0029 / 0.027</td>
<td>0.0029 / 0.027</td>
<td>0.006 / 0.003</td>
<td>0.006 / 0.010</td>
<td>0.006 / 0.008</td>
<td>0.006 / 0.008</td>
</tr>
<tr>
<td>C2</td>
<td>0.009 / 0.159</td>
<td>0.0028 / 0.0028</td>
<td>0.0029 / 0.0029</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
<td>0.0002 / 0.0002</td>
</tr>
</tbody>
</table>

### Table-II: Estimation of Probabilities of Joint Events

<table>
<thead>
<tr>
<th>Event</th>
<th>$P_{\text{RF}}(\sigma_{\text{VT}})$</th>
<th>$P_{\text{WF}}(\sigma_{\text{VT}})$</th>
<th>$P_{\text{HF}}(\sigma_{\text{VT}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta V_{\text{t}}&lt;0$</td>
<td>$\sigma_{\text{VT}}$</td>
<td>$\sigma_{\text{VT}}$</td>
<td>$\sigma_{\text{VT}}$</td>
</tr>
<tr>
<td>$\delta V_{\text{t}}&gt;0$</td>
<td>$\sigma_{\text{VT}}$</td>
<td>$\sigma_{\text{VT}}$</td>
<td>$\sigma_{\text{VT}}$</td>
</tr>
</tbody>
</table>

### 3.7. Estimation of Yield

To estimate the yield of memory, we define the failure probability of column ($P_{\text{COI}}$) (or row ($P_{\text{ROW}}$)) as the probability that any of the cells (out of $N$ cells) in that column (or row) fails. Assuming a column redundancy, the probability of failure of a memory chip ($P_{\text{CHIP}}$) designed with $N_{\text{COI}}$ number of columns and $N_{\text{NC}}$ number of redundant columns, is defined as the probability that more than $N_{\text{NC}}$ (i.e. at least $N_{\text{NC}}+1$) columns fail (similar definition is applicable for row redundancy). Hence, $P_{\text{COI}}$ and $P_{\text{ROW}}$ can be given by:

$$P_{\text{COI}} = 1 - (1 - P_{\text{RF}})^{N_{\text{COI}}}$$

(20)

To estimate the yield, we have used Monte-Carlo simulations for interdie distributions of $L, W, and V_{\text{t}}$ (assumed to be Gaussian). For each inter-die values of the parameters (say $L_{\text{INTER}}$, $W_{\text{INTER}}$, and $V_{\text{INTER}}$) we estimate $P_{\text{RF}}$, $P_{\text{W}}$, and $P_{\text{H}}$ considering the intra-die distribution of $\delta V_{\text{t}}$. Finally, the yield is defined as:

$$\text{Yield} = 1 - \sum_{i=1}^{N_{\text{INTER}}} P_{\text{RF}}(L_{\text{INTER}}, W_{\text{INTER}}, V_{\text{INTER}})$$

(21)

where, $N_{\text{INTER}}$ is the total number of inter-die Monte-Carlo simulations (i.e. total number of chips). The yield decreases significantly with increasing $\sigma_{\text{VT}}$ (Fig. 10). In the estimation, we have assumed a standard deviation of 1% for inter-die distribution of $L, W$, and $V_{\text{t}}$. $N_{\text{COI}}=512, N_{\text{NC}}=24, \sigma_{\text{VT}}=30 \text{mV}$. The estimated probabilities match the Monte-Carlo results very closely (Table-II). All of the different failure probabilities increase significantly with an increase in the sigma of $\delta V_{\text{t}}$ variation (Fig. 10).
4. DISCUSSIONS

4.1. Impact of Vdd and Temperature Variation

A drop in the supply voltage increases the cell failure probabilities. (Fig. 11a). The impact of supply voltage drop is most significant for the access time failure. The derived model can also be extended to include Vdd of a cell as an independent variable. Fig. 11b shows that the failure probabilities increase with an increase in the temperature. The impacts of temperature increase are more severe on the access time and write failure because of (a) increase in the junction capacitances.

4.2. Transistor Size and Cell Failure Probability

The length and width of different transistors of the cell (i.e., \(W_{\text{NOM}}, W_{\text{PUP}}, V_{\text{THP}}, V_{\text{TRP}}\) ) impact the cell failure probability principally by modifying: (a) the nominal values of the PMOS pull-down transistors (i.e., \(W_{\text{NOM}}\) ), (b) the rate of change of these parameters with Vt variation thereby changing the mean and the variance of these parameters, and (c) the variance of the Vt parameter (see (1)).

For example, Fig. 12 shows that, along with the nominal value, the mean and the standard deviation of the failure probability significantly vary with \(W_{\text{PUP}}\) and \(W_{\text{NOM}}\). In this section, we study the impact of variation of strength of different transistors (only width is used to vary the strength) on the cell failure probability. Variation of the strength using channel length follows the same trend.

Fig. 13 shows that, a weak access transistor (small \(W_{\text{NOM}}\) ) reduces \(P_{\text{AF}}\) (Fig. 12) but increases \(P_{\text{AF}}\) (Fig. 13) and has very small impact on \(P_{\text{AF}}\). Decreasing \(W_{\text{PUP}}\) decreases \(P_{\text{AF}}\) (increases \(P_{\text{AF}}\) ), increases \(P_{\text{AF}}\) (lowers \(V_{\text{THP}}\) ), and \(P_{\text{AF}}\) does not depend strongly on PMOS strength (Fig. 13). However, \(P_{\text{AF}}\) improves with an increase in \(W_{\text{PUP}}\) as the node L is more strongly coupled to the supply voltage \((V_{\text{DD}})\). Increasing \(W_{\text{PUP}}\) increases the strength of pull-down NMOS (\(N_{\text{P}}\) and \(N_{\text{B}}\)). This reduces \(P_{\text{AF}}\) (Fig. 12) and \(P_{\text{AF}}\) by increasing the strength of \(N_{\text{P}}\) (Fig. 13). Increase in width of \(N_{\text{B}}\) has little impact on \(P_{\text{AF}}\). Although, it slightly increases the nominal value of \(V_{\text{THP}}\), the reduction of \(\sigma_{\text{Vth}}\) of \(N_{\text{P}}\) (see (1)) tends to reduce \(\sigma_{\text{THP}}\) and hence \(P_{\text{AF}}\) remains almost constant. An increase in the \(V_{\text{THP}}\) of \(P_{\text{AF}}\) initially reduces \(P_{\text{AF}}\) with the increase in \(W_{\text{PUP}}\). However, a higher width of \(N_{\text{P}}\) reduces \(V_{\text{THP}}\) (from the applied \(V_{\text{DD}}\)) due to an increase in the leakage of \(N_{\text{P}}\). Consequently, a very high \(W_{\text{PUP}}\) increases \(P_{\text{AF}}\). Due to the variation in the failure probability, the choice of the transistor sizes has a strong impact on the yield (Fig. 14a). Hence, it can be concluded that, a statistical approach to the design of the transistor sizes is necessary to maximize the yield. The derived failure probability models can be effectively used for such statistical optimizations.

4.3. Memory Architecture and Yield

Increasing the number of cells in a column (column length or \# of rows) increases the cell failure probability (particularly \(P_{\text{AF}}\) as \(C_{\text{NL}}\) and \(I_{\text{BL}}\) increases in (5) resulting in higher \(T_{\text{ACCESS}}\)). Moreover, \(P_{\text{COL}}\) and hence \(P_{\text{MEAN}}\) increases significantly with the column length (see (20)). However, for a constant memory size, increasing column length reduces the number of columns, which tends to reduce \(P_{\text{MEAN}}\) (assuming a constant redundancy). Consequently, to maximize the yield, the design of the memory organization has to consider its impact on the failure probabilities. Fig. 14b shows the variation column failure probability, the memory failure probability and yield of a 2KB cache with the column length (\# of redundant column kept constant). It can be observed that, yield strongly depends on the column length.

5. CONCLUSION

In this work, we have proposed a semi-analytical method to estimate the failure probability of a SRAM cell due to parameter variation. The derived models have been used to predict the yield of memory at an early stage of design. The proposed models provide a statistical approach for optimizing the memory design, which is necessary for maximizing yield in nano-meter regime.

Acknowledgement: We thankfully acknowledge many helpful discussions with Dr. Dinesh Somasekhar of Intel Corp.

Reference:

Fig. 11: Impact of (a) supply voltage drop and (b) temperature increase on the failure probability

Fig. 12: The impact of transistor size on the distributions of \(T_{\text{ACCESS}}\)

Fig. 13: Variation of Cell Failure Probabilities with Cell structure

Fig. 14: Impact of (a) circuit (transistor size) and (b) architecture (\# of rows (column length) and \# of columns) on yield. In (b) transistor sizes were chosen to maximize yield following (a). \(\sigma_{\text{Vth}}=20 \text{mV}\)