

Leakage Current Based Stabilization Scheme for Robust Sense-Amplifier Design for Yield Enhancement in Nano-scale SRAM

Saibal Mukhopadhyay, Arijit Raychowdhury, Hamid Mahmoodi, and Kaushik Roy
 Dept. of ECE, Purdue University, West Lafayette, IN-47906, USA
 <sm, araycho, mahmoodi, kaushik>@ecn.purdue.edu

Abstract

In this paper, we develop a method to analyze the probability of access failure in SRAM array (due to random V_t variation in transistors) by jointly considering variations in cell and sense-amplifiers. Our analysis shows that, improving robustness of sense-amplifier is extremely important for reducing memory access failure probability and improving yield. We present a process variation tolerant sense amplifier suitable for SRAM array designed in sub-100nm CMOS technologies. The proposed technique reduces the failure probability of sense amplifiers by more than 80% with negligible penalty in the sensing delay.

1. Introduction

The inter-die and intra-die variation in the process parameters (channel length (L), width (W), and transistor threshold voltage (V_t)) in nano-scale technologies can significantly degrade the yield of an SRAM design [1,2]. The V_t variation in the transistors in the SRAM cell causes in read, write, access and hold failures in the cell. On the other hand, the statistical variations in the device parameters also result in significant variation in the sensing delay (difference between the time sense-amplifier is turned "ON" and the time sense-amplifier outputs are available) of the sense amplifiers used in SRAMs and in the worst case may amplify the input signal in the wrong direction (functional failure due to incorrect sensing) [3-10]. Access failure in SRAM is defined as the reading of wrong data during the access of the memory array. The access failure in an SRAM array can occur due to random V_t variation in SRAM cell or sense-amplifier. Hence, analysis of access failure has to consider the impact of both SRAM cell and sense-amplifiers. Moreover, in nano-meter regime design of robust sense amplifiers is of extreme importance for reducing access failures in SRAM array and improving yield of SRAM design.

In this paper, we propose a method to estimate the access failure probability in memory considering both SRAM cell and sense-amplifier together (Fig. 1). Our analysis shows that, improving the robustness of sense-amplifier can significantly reduce the access failure probability in memory. To improve the process variation tolerance in sense-amplifier, we propose a feed-forward path based stabilization scheme, implemented using leakage current. The proposed technique significantly improves the robustness in sense-amplifier designed in sub-100nm CMOS technologies.

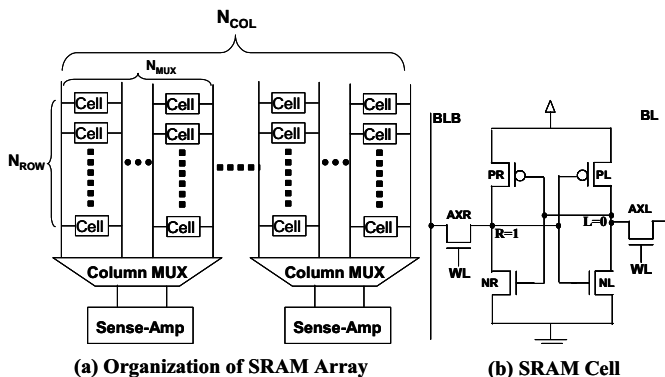


Fig. 1: Structure of SRAM array and cell.

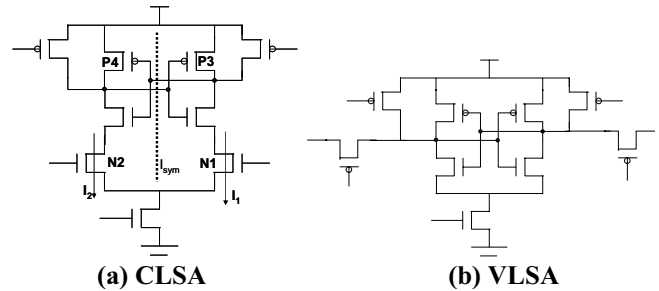


Fig. 2: Voltage mode sense amplifiers: (a) Current Latch based Sense Amp (CLSA), (b) Voltage Latch based Sense-Amp (VLSA)

In SRAM, principally two types of sense amplifiers are used: (a) voltage mode and (b) current mode [3-10]. In this work, we focus on the Voltage Mode Sense Amplifiers (VMSA). We apply the proposed leakage current based stabilization technique to two different types of VMSA, namely, Current Latch based Sense Amplifiers (CLSA) and Voltage Latch based Sense Amplifiers (VLSA) [3-10] (Fig. 2). The proposed technique reduces the probability of incorrect sensing in the sense-amplifiers. Moreover, as the feed-forward scheme is implemented using the subthreshold leakage of a transistor which increases exponentially with technology scaling, the effectiveness of the proposed technique also improves with technology scaling.

2. Modeling of Access Failure in SRAM

In this section, we propose a model to estimate access failure probability in memory considering the SRAM cell and sense-amplifiers. In this analysis, we consider the V_t variation in the transistors due to random dopant fluctuations (RDF). The V_t variation of the transistors due to RDF, are considered as independent Gaussian random variables with $\sigma_{VT} = \sigma_{VT0} \times \sqrt{(W/W_0)}$, where W_0 is width of the minimum size transistor [1, 2].

2.1. Bit-Differential produced by the SRAM Cell

Let us consider the memory cell in Fig. 1 which stores "1" at the node R and "0" at node L. While reading the cell the bit-line BL discharges through the access transistor AXL and the pull-down transistor NL, while BLB stays high. Let us also consider that, the time delay between the word-line activation and the sense amplifier firing is given by T_{SE} . The difference between two bit-lines (bit-differential) at the time of firing of sense amplifier (Δ_{BIT}) can be obtained as [2]:

$$\Delta_{BIT} = \left(I_{dsatAX-PD} - \sum_{i=1, \dots, N_{Row}} I_{subAX} \right) \frac{T_{SE}}{C_{BL}} \quad (1)$$

where, $I_{dsatAX-PD}$ is the current of through AXL-NL, I_{subAX} is the subthreshold leakage of the access transistors connected to the bit-line BLB and C_{BL} is the bit-line capacitance. The random V_t variation in the threshold voltage of AXL and NL results in a distribution of Δ_{BIT} (Fig. 4a, obtained from an SRAM cell designed in BPTM 70nm, $\sigma_{Vt} = \sigma_{Vt0} \times \sqrt{(W/W_0)}$ with $W_0 =$ width of PMOS pull-up transistor).

2.2. Failure in Sensing Operation in Sense Amplifier

The input offset voltage (V_{os}) of a sense amplifier is defined as the minimum bit-differential ($\Delta_{BIT} = V_{BLB} - V_{BL}$) required by a sense-

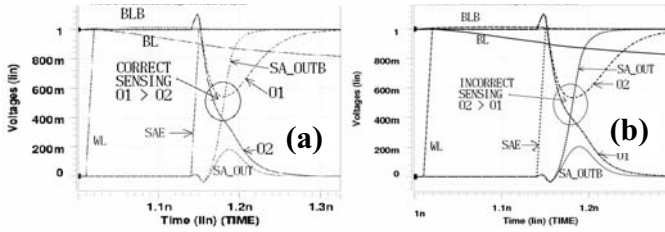


Fig. 3: Sensing operation (a) Correct and (b) Incorrect amplifier for correct sensing [4]. Let us consider the Current Latch Sense Amplifier shown in Fig. 2a to understand the origin of V_{os} . In the pre-charge cycle nodes $O1$ and $O2$ are pre-charged to high. During sensing, when the sense amplifier is turned “on” by raising SAE high, the nodes $O1$ and $O2$ discharges by currents I_1 and I_2 , respectively. However, if $V_{BL} < V_{BLB}$ strength of the transistors N2 is higher than N1 (i.e. $I_2 > I_1$). Hence, $O2$ discharges at a rate faster than $O1$ thereby developing voltage difference between $O1$ and $O2$ ($\Delta V_o = V_{O1} - V_{O2}$). As $O2$ reduces at a faster rate than $O1$, the PMOS P3 gets turned “ON” earlier (when, $O2$ reduces below $V_{DD} - |V_t(P3)|$) than the PMOS P1. Turning “ON” of P3 further increases the voltage difference between $O1$ and $O2$. After the voltage difference becomes reasonably high, the cross coupled inverter action amplifies the voltage difference and $O2$ goes to zero and $O1$ returns back to V_{DD} . Outputs $O1$ and $O2$ are inverted to produce final output SA_OUT and SA_OUTB . This is a correct sensing operation (Fig. 3a).

The CLSA is symmetric about the line ‘ I_{sym} ’ and if all the devices in the two sides of the symmetry line are identical, the sense amplifier can sense any bit-differential correctly i.e. $V_{os} = 0$. Let us now assume that, the V_t of the transistors N1 is reduced and that of N2 is increased ($V_{tN2} > V_{tN1}$). Under this condition, although ($V_{BL} < V_{BLB}$) the current I_2 can be lower than I_1 as shown below:

$$I_1 \propto (V_{BL} - V_{tN1}) \text{ and } I_2 \propto (V_{BLB} - V_{tN2}); \quad (2)$$

$$\therefore (V_{tN2} - V_{tN1}) > (V_{BLB} - V_{BL}) \Rightarrow I_1 > I_2$$

If $I_1 > I_2$, the node $O1$ discharges at a rate faster than $O2$ resulting in an incorrect sensing (Fig. 3b). A mismatch in W and/or L between $N1$ and $N2$ can also result in such incorrect operation. Hence, for proper sensing we must have:

$$\Delta_{BIT} = (V_{BLB} - V_{BL}) > (V_{tN2} - V_{tN1}) \quad (3)$$

Thus, the input offset voltage in the above case is greater than zero (i.e. greater than $V_{tN2} - V_{tN1}$). From the above discussion and the definition of the input offset voltage, the probability of failure (i.e. incorrect sensing event) in a sense amplifier is given by [4]:

$$P_F = P(V_{os} > \Delta_{BIT}) \quad (4)$$

The V_t variation in the transistors of the sense amplifier results in a distribution (Normal in nature [4]) of the offset voltage V_{os} (Fig. 4a). The mean of the distribution is zero as in the perfectly balanced case the sense amplifier can correctly sense any bit-differential [4].

2.3. Modeling of Access Failure Probability in SRAM Array

The access failure in a memory is due to a slow SRAM cell resulting in lower Δ_{BIT} [2, 4] and/or due to a sense amplifier with high V_{os} [4].

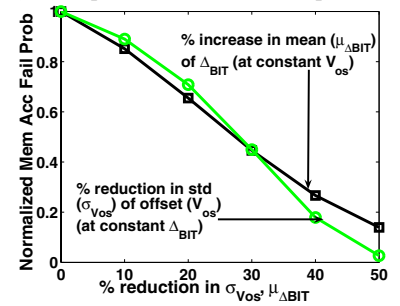
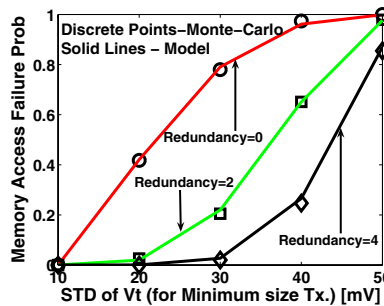
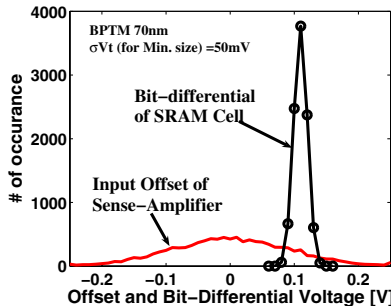


Fig. 4: Memory access failure probability: (a) variation of bit-differential (Δ_{BIT}) and offset voltage (V_{os}), (b) prob. of access failure in memory, (c) effect of V_{os} in reducing access failure (% reduction is applied on V_{os} for $\sigma V_t = 50\text{mV}$, redundancy=2)

Hence, the estimation of the access failure in memory has to consider the variation in both of them. To understand how the access failure probability in a memory can be estimated let us consider Fig. 1, which shows the basic organization of the SRAM array. There are N_{ROW} numbers of cells in a column and N_{COL} number of columns. There exists a sense amplifier for each group of N_{MUX} number of columns which are multiplexed together using a column multiplexer (hereafter referred to as *MuxGroup*). The V_t variation of the transistors in the SRAM cell results in the variation in the Δ_{BIT} value developed while accessing an cell. For *no access* failure in a *MuxGroup*, the minimum value of the bit-differential (say Δ_{BITSA}) that can develop in any of the columns needs to be higher than the offset voltage of the sense amplifier. Hence, the event of successful sensing is defined as:

$$\Omega_{CorrectSense} := V_{os} < \Delta_{BITSA} \quad (5)$$

$$\Delta_{BITSA} = \min_{i=1}^{N_{MUX}} \left\{ \Delta_{BITCOL}^{(i)} \right\} = \min_{i=1}^{N_{MUX}} \left\{ \min_{k=1}^{N_{ROW}} (\Delta_{BIT}^{(k)}) \right\}$$

where, $\Delta_{BIT}^{(k)}$ is the bit-differential produced while accessing the k -th cell in the i -th column and $\Delta_{BITCOL}^{(i)}$ is the minimum value of the bit-differential that can develop in the i -th column. Since we have considered only the random V_t variation (principally due to the random dopant fluctuation) the $\Delta_{BIT}^{(k)}$ for different values of k , in all the columns are Identical and Independent Normal Distributions (IID). Hence, the probability ($f_{BITCOL}(x)$) and the cumulative ($F_{BITCOL}(x)$) distribution function (pdf and cdf) of $\Delta_{BITCOL}^{(i)}$ (for any column) is given by [16]:

$$F_{BITCOL}(\delta_{BITCOL}) = 1 - (1 - \Phi_{BIT}(\delta_{BITCOL}))^{N_{ROW}} \quad (6)$$

$$f_{BITCOL}(\delta_{BITCOL}) = N_{ROW} (1 - \Phi_{BIT}(\delta_{BITCOL}))^{N_{ROW}-1} \phi_{BIT}(\delta_{BITCOL})$$

where, $\Phi_{CELL}(\delta_{BIT})$ and $\phi_{CELL}(\delta_{BIT})$ represents the cdf and pdf of the distribution of the Normal random variable Δ_{BIT} . Similarly, Δ_{BITCOL} for different columns are also IID variables (as the variation the bit-differential for cells in different columns is also identical and independent). Hence, the pdf (f_{BITSA}) and cdf (F_{BITSA}) of the variable Δ_{BITSA} is given by [16]:

$$F_{BITSA}(\delta_{BITSA}) = 1 - (1 - F_{BITCOL}(\delta_{BITSA}))^{N_{MUX}} \quad (7)$$

$$f_{BITSA}(\delta_{BITSA}) = N_{MUX} (1 - F_{BITCOL}(\delta_{BITSA}))^{N_{MUX}-1} f_{BITCOL}(\delta_{BITSA})$$

From (5) the probability of the correct sensing for a *MuxGroup* is given by:

$$P_{CS} = P(\Omega_{CorrectSense}) = P(V_{os} < \Delta_{BITSA}) = P(Z := (V_{os} - \Delta_{BITSA}) < 0) \quad (8)$$

Moreover, the V_t variation in the SRAM cells and the sense amplifiers are independent of each other. Hence, (8) can be estimated as [16]:

$$P_{CS} = F_Z(0) = \iint_{(v_{os} - \delta_{BITSA}) < 0} f_{BITSA}(\delta_{BITSA}) \phi_{OS}(v_{os}) dv_{os} d\delta_{BITSA}$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\delta_{BITSA}} \phi_{OS}(v_{os}) dv_{os} \right) f_{BITSA}(\delta_{BITSA}) d\delta_{BITSA} \quad (9)$$

$$= \int_{-\infty}^{\infty} \Phi_{OS}(\delta_{BITSA}) f_{BITSA}(\delta_{BITSA}) d\delta_{BITSA}$$

where, $\Phi_{OS}(v_{os})$ and $\phi_{OS}(v_{os})$ represents the cdf and pdf of the

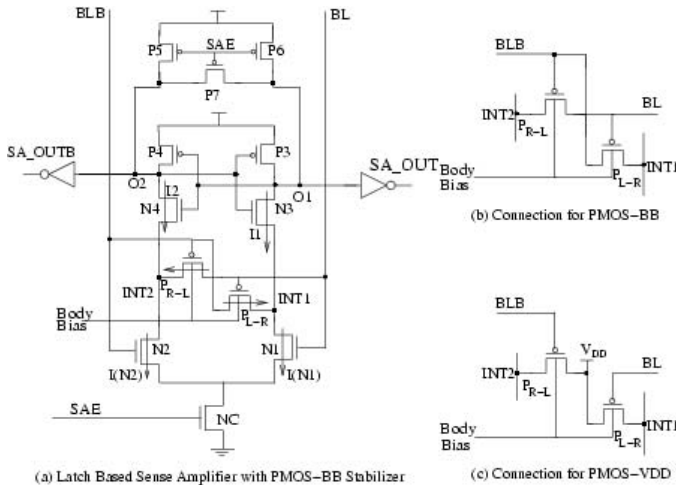


Fig. 5: Latch Based Sense Amplifier with the PMOS stabilizer.

Normal random variable V_{OS} . Using (6), (7), the above equation can be evaluated using a numerical integral over δ_{BIT} (varying from 0 to $0.3V_{DD}$ instead of $-\infty$ to ∞). The access failure in an SRAM array occurs if incorrect sensing occurs in any of the *MuxGroup* (assuming no redundancy). Hence, the access failure probability of the array (P_{MEM-AC}) is given by:

$$P_{MEM-AC} = 1 - (P_{CS})^{N_{COL}/N_{MUX}} \quad (10)$$

If there exists N_{RC} number of redundant *MuxGroups*, P_{MEM-AC} is modified to:

$$P_{MEM-AC} = \sum_{k=N_{RC}+1}^{(N_{COL}/N_{MUX})+N_{RC}} \binom{(N_{COL}/N_{MUX})+N_{RC}}{k} (1-P_{CS})^k P_{CS}^{(N_{COL}/N_{MUX})+N_{RC}-k} \quad (11)$$

The above method is general in nature. However, if correlation between the cell and the columns in a *MuxGroup* exists, (6) and (7) need to be modified [16]. An increase in the V_t variation increases the memory access failure probability (Fig. 4b, obtained using SRAM cell and CLSA designed in BPTM 70nm technology). From (10) it can be an increase in Δ_{BIT} (e.g. using transistor sizing [2]) increases the P_{CS} and reduces P_{MEM-AC} (Fig. 4c). Similarly, reducing the variability of V_{os} , also reduces the memory access failure probability (Fig. 4c). Moreover, from Fig. 4a it can be concluded that, variation in V_{os} in sense-amplifier has a stronger impact on access failure compared to variation in Δ_{BIT} produced by cell. Hence, design of robust sense-amplifier is necessary to reduce access failure in memory.

3. Design of Robust Voltage Mode Sense-Amplifier

It is observed in section-2 that, reducing failure probability of sense-amplifier (by reducing V_{os} and its variability) reduces the memory access failure probability. In this section, we propose a leakage current based stabilization scheme for Current Latch Sense-Amplifier to improve its robustness. We also apply the proposed technique to Voltage Latch Sense Amplifiers. We evaluate the failure probability of the sense-amplifier as [4]:

$$P_F = \frac{\text{\#of Incorrect Sensing (Faulty SA)}}{\text{\#of SA in a Chip}} \quad (12)$$

It has been shown in [4] that, the definitions (4) and (12) are equivalent.

3.1. Previous Research on Mismatch Compensation

The V_t (similarly, L and W) mismatch in the transistors has a considerable impact on the stability of sense amplifiers [4], [6], [7]. DRAM designers have proposed several offset compensation schemes [6], [8], [9]. These techniques try to compensate the mismatch by either pre-charging the bit-lines or the two outputs to different initial values. However, in VMsA's for SRAM, bitlines and the outputs are precharged and equalized to V_{DD} [4], [11]. Hence, these techniques can not be directly applied to VMsA's in SRAM. V_t mismatch

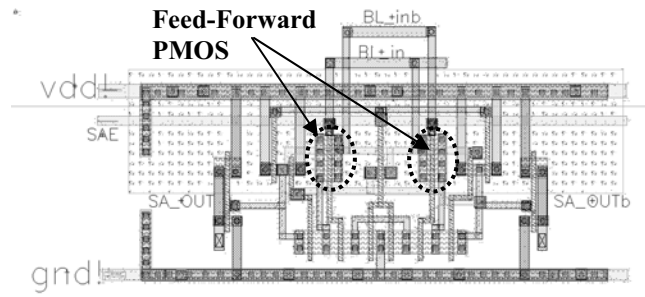


Fig. 6: Layout of the proposed CLSA.

compensation schemes described for current mode sense amplifiers [10] are also not readily applicable for voltage mode sense amplifiers. In [4] authors have discussed that, the probability of functional failures can be reduced by using a lower voltage at the gate of the transistors N1 and N2 (i.e. $V_g(N2)=V_{INC} < V_{DD}$, $V_g(N1)=V_{INC} - \Delta$). However, such a case is only possible if the sense-amplifiers are used as a second-stage of the current mode sense amplifiers [4]. In case of the voltage mode sensing, gates of N1 and N2 are connected to bit-lines (as explained in section-2) and hence, pre-charged to V_{DD} .

3.2. Leakage based Stabilization in CLSA

To enhance the robustness of the voltage mode sense amplifiers, we have introduced two PMOS transistors, P_{L-R} and P_{R-L} (collectively called the PMOS stabilizers), between the inputs (i.e. the bit-lines BL and BLB) and the drain of the driver transistors $N1$ and $N2$, as shown in Fig. 5. Let us consider that, the voltage at BL (V_{BL}) drops linearly whereas that of BLB (V_{BLB}) remains at V_{DD} . As V_{BL} reduces, V_{gs} ($= -\Delta = V_{BL} - V_{BLB}$) of the PMOS P_{L-R} becomes negative and the transistor gets weakly turned on. Hence, the drain-to-source current ($I_{DS}(P_{L-R})$) of the transistor P_{L-R} is added to the drain of $N1$ (node $INT1$). As long as $|\Delta|$ is less than $|V_t|$ of P_{L-R} the current $I_{DS}(P_{L-R})$ is the *subthreshold leakage* [12]. However, if $|\Delta| > |V_t|$, weak on-current flows into the node $INT1$. With this additional current, the KCL at node $INT1$ becomes:

$$I_{DS}(N1) = I_1 + I_{DS}(P_{L-R}) \quad (13)$$

where, I_1 is the effective discharging current for node $O1$. Hence, for a fixed size of $N1$ (i.e. for a value of $I_{DS}(N1)$) this additional current ($I_{DS}(P_{L-R})$) reduces the strength of I_1 . Thus, the added P_{L-R} helps to slow down the discharge of node $O1$. On the other hand, for P_{R-L} we have $V_{gs} = V_{BLB} - V_{BL} > 0$. Thus, the subthreshold current through P_{R-L} is negligible (as in PMOS $I_{sub} \propto \exp[-(V_{gs} - |V_t|)/n(kT/q)]$ [12]). Hence, the current through $N2$ remains same as I_2 . Assuming the current through the transistor NC to be I_0 and following the analysis presented in [4], the difference in the discharging current between $O2$ and $O1$ (i.e. $\delta I = I_2 - I_1$) in the conventional CLSA is given by:

$$I_2 - I_1 = \delta I = I_{ds}(N2) - I_{ds}(N1) = \sqrt{2\beta I_0} (V_{BLB} - V_{BL}) = \sqrt{2\beta I_0} \Delta \quad (14)$$

where, β is the trans-conductance parameter of transistors $N1$ and $N2$ (assumed to be equal). In case of the proposed CLSA, the current difference is given by:

$$\begin{aligned} I_{ds}(N2) - I_{ds}(N1) &= I_2 + I_{ds}(P_{R-L}) - I_1 - I_{ds}(P_{L-R}) \\ \Rightarrow \delta I &= I_2 - I_1 = I_{ds}(N2) - I_{ds}(N1) + I_{ds}(P_{L-R}) - I_{ds}(P_{R-L}) \\ &= \sqrt{2\beta I_0} \Delta + I_{sub0} \left[\exp\left(\frac{\Delta}{nKT/q}\right) - \exp\left(\frac{-\Delta}{nKT/q}\right) \right] \end{aligned} \quad (15)$$

where, I_{sub0} is the subthreshold current of the PMOS at $V_{gs}=0$. From (15) it can be observed that, the addition of the stabilization PMOSs produces a larger current difference ($I_1 - I_2$) for the same value of input voltage difference (Δ). Moreover, if due to V_t mismatch, the strength of $N1$ goes high (i.e. V_t of $N1 \downarrow \Rightarrow I_{DS}(N1) \uparrow$) a part of the extra current is obtained from P_{L-R} , (instead of completely from I_1 as in the case of conventional CLSA). Hence, a reduction in the V_t of $N1$ results in a smaller increase in the value of I_1 , which improves the

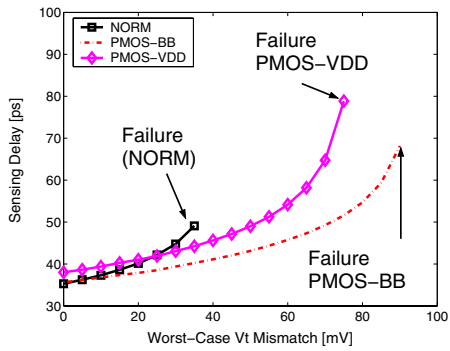


Fig. 7: Sensing delay vs worst-case V_t mismatch.

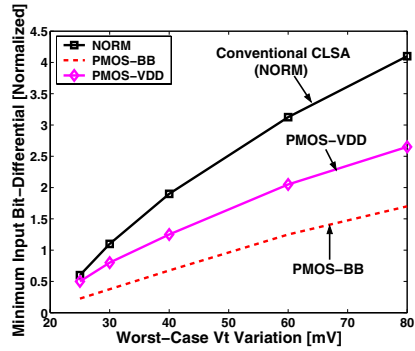


Fig. 8: Variation of input offset voltage: (a) with worst-case V_t mismatch in all transistors, (b) with random V_t variation in all transistors

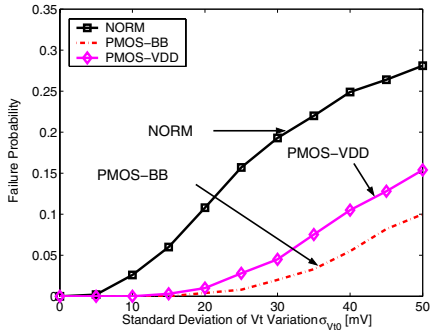
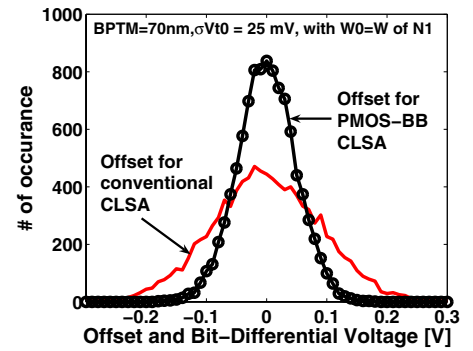


Fig. 9: Failure probability vs V_t variation.

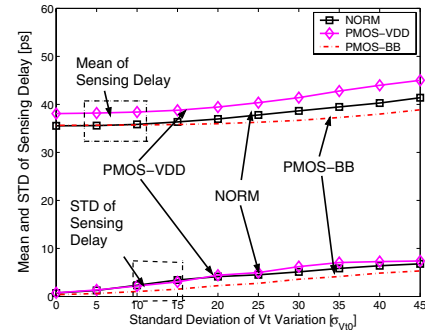


Fig. 10: Sensing delay vs worst-case V_t mismatch.

robustness. In other words, a part of the current difference $|I_1 - I_2|$ is independent of the threshold voltage of $N1$ and $N2$ and depends on the strength of the PMOS stabilizers. Hence, the overall sensitivity of the sensing delay (which is a function of $|I_1 - I_2|$) with respect to the V_t (also L and W) mismatch of $N1$ and $N2$ reduces. The proposed PMOS stabilization scheme is a feed-forward mechanism where input bit-differential produces an additional discharging current difference at the output [17]. Hence, for a correct sensing operation (i.e. to develop a certain value of $|I_1 - I_2|$) the proposed feed-forward scheme requires a lower input voltage difference between the bit-lines BL and BLB. In other words, the proposed scheme reduces the input offset voltage, thereby reducing the probability of the incorrect operation. It should be noted that, the feed-forward path needs to be active only during the time interval when both the nodes O1 and O2 are discharging. The robustness improvement is ensured even if the bit-lines are pre-charged back to V_{DD} after this time. In order to increase the effectiveness of the feed-forward path, the PMOS transistors have to be strong. The described scheme is suitable in *nano-scaled CMOS* as the *subthreshold leakage increases exponentially* with technology scaling [11, 12].

In the described stabilization scheme PMOS transistors are connected to bit-lines (hereafter referred to as PMOS-BB). Hence, as the PMOS P_{L-R} is weakly turned on, the current drawn from the bit-line BLB (which is not discharging through the accessed cell) causes V_{BLB} to drop from V_{DD} by a small amount. This drop can be eliminated by connecting the source of the PMOS P_{L-R} and P_{R-L} to the V_{DD} instead of the bit-lines (Fig. 5c). This scheme will be referred as PMOS-VDD in this paper. In the PMOS-VDD case, as before, one PMOS (P_{L-R}) will have negative V_{gs} ($=V_{BL} - V_{DD} = -\Delta$), but the other PMOS (P_{R-L}) will have $V_{gs} = V_{BLB} - V_{DD} = 0$ (instead of positive). Hence, the current added to the node $INT1$ remains same but that added to $INT2$ is larger than the PMOS-BB case. This will (a) increase the sensing delay and (b) reduces $|I_1 - I_2|$ from PMOS-BB case. Hence, it will be less effective in reducing failure probability.

3.3. Sensitivity to Parameter Variation

To estimate the sensitivity of the proposed schemes to the parametric variations we have designed the conventional CLSA (referred to as

NORM), PMOS-BB CLSA and PMOS-VDD CLSA and simulated using 70nm devices in SPICE [13]. The netlists including the parasitic capacitances, extracted from layouts (drawn in 180nm and scaled down to 70nm), were used in the simulation. Fig. 6 shows the layout of the proposed PMOS-BB scheme. The sizes of all the transistors are kept same in the three implementations. The size of PMOS stabilizers in PMOS-BB and PMOS-VDD schemes are also kept same. We first considered worst-case mismatch in the V_t of $N1$ and $N2$ (V_t of $N1 \downarrow$ and V_t of $N2 \uparrow$) for the operation described earlier (i.e. BL drops and BLB remains high). Simulation results show that, with the stabilizers, the sensing delay becomes less sensitive to V_t mismatch and the maximum V_t mismatch tolerance increases significantly (Fig. 7). Among the two stabilization schemes, PMOS-BB gives better mismatch tolerance as explained earlier. Fig. 8a shows the variation of the minimum input bit-differential required for correct sensing (i.e. Input Offset voltage) with the variation in worst-case V_t mismatch in the transistors in the sense-amplifier. The V_t mismatch is applied to the PMOS P3, P4 and NMOS N1, N2, N3 and N4 in the worst-case directions (i.e. for P3, N2, N4: $\delta V_t > 0$ and P4, N1, N3: $\delta V_t < 0$, assuming BLB stays high and BL goes low). For the PMOS-BB and PMOS-VDD schemes, along with the other transistors worst-case V_t mismatch is also applied to the feed-forward PMOS transistors (i.e. P_{L-R} : $\delta V_t > 0$ and P_{R-L} : $\delta V_t < 0$). It can be observed that, the input offset voltage reduces with the introduction of the PMOS stabilizers. Moreover, PMOS-BB scheme has a lower input offset voltage compared to the PMOS-VDD scheme. Hence, use of the PMOS-BB scheme significantly improves the robustness of the sense amplifier, which results in a lower failure probability. Fig. 8b shows the variation of the input offset voltage considering random V_t variation (σ_{Vt0}), obtained using Monte-Carlo simulation in SPICE. The variation is applied to *all* the transistors (including the stabilizers). As explained earlier, δV_t of all the transistors were assumed independent of each other (L and W mismatch in the neighbouring transistors were also considered). Standard deviation (STD) of δV_t of each transistor was chosen based on their sizes (following $\sigma_{Vt} = \sigma_{Vt0} \times \sqrt{(W/W_0)}$, $W_0 =$ width of $N1, N2$). Input offset voltage in PMOS-BB CLSA has a lower spread compared to that in the conventional CLSA.

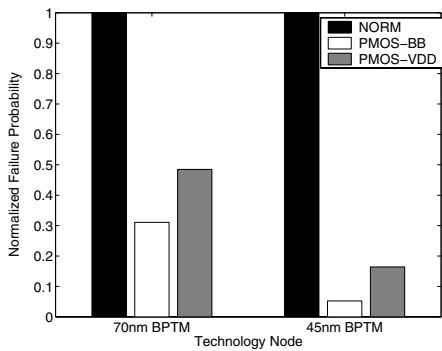


Fig. 11: Impact of technology scaling. P_F are normalized to that of the conventional CLSA of the corresponding technology. ($\sigma_{V_t}=40\text{mV}$)

shows that, the introduction of the stabilizers significantly reduces the failure probability (obtained using (12)). The PMOS-BB scheme reduces the failure probability by more than 80% for reasonable V_t variations (σ_{V_t} within 10mV-50mV) in the transistors. It can be observed that, the mean and the STD of sensing delay does not increase (Fig. 10) with the stabilizers. In fact, use of PMOS-BB reduces the mean delay for a high value of σ_{V_T} . This emphasizes the fact that, the process variation tolerance has *not* been obtained at the cost of reduced performance of the functional sense-amplifiers. Simulation with BPTM 45nm [13] devices shows that the stabilization becomes more effective with technology scaling (Fig. 11). This is due to the fact that, the subthreshold increases exponentially with technology scaling. As the gate leakage of PMOS is small [12], it does not impact the stabilization effect.

3.4 Design Considerations

The major design parameter in the proposed scheme is the strength of the PMOS stabilizers (increasing strength increases I_{sub0} in (15)). As the width of PMOS transistors compared to the NMOS $N1$ and $N2$ increases the failure probability reduces (Fig. 12). Increasing the width also reduces the V_t variation in the stabilizer itself. However, the width can not be increased too much because of (a) area constrain and (b) larger junction capacitance added to the node $INT1$ and $INT2$ will slow down the circuit. The strength can also be increased by reducing the V_t of the PMOS transistors (Fig. 13). This can be achieved either by using low- V_t transistors (multiple V_t process) or by applying separate forward body bias to the PMOS transistors (Fig. 5). However, making the V_t of the transistors too low will also increase the current of the 'off' PMOS (i.e. P_{R-L} in previous discussion), thereby reducing I_2 . That will reduce the speed of the operation.

The proposed scheme has 8-15% of layout area overhead (depends on whether sense amplifier is shared by 4 or 8 columns) in our design (assuming 180nm design rules). However, this increase in sense amplifier area results in negligible area overhead in the SRAM array.

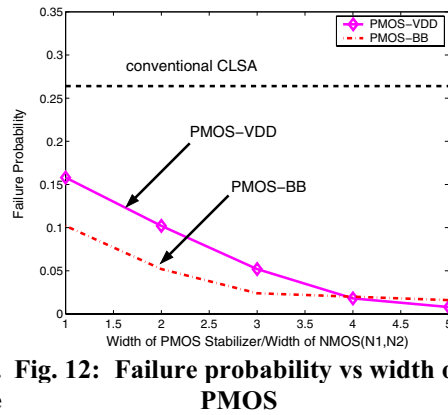


Fig. 12: Failure probability vs width of PMOS

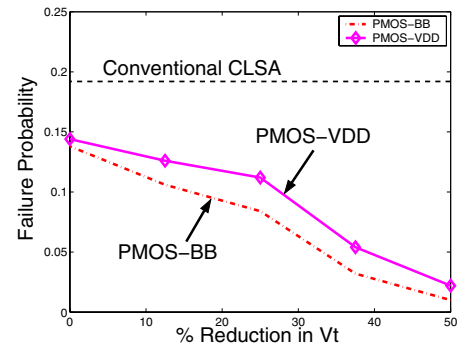


Fig. 13: Failure probability vs Low- v_t usage in PMOS stabilizers

It should be noted that, the difference between the time when the sense-amplifier is turned "ON" (i.e. SAE is raised high) and the word-lines are deactivated (and pre-charging of the bit-lines starts) is very small [14]. As soon as the bit-lines are pre-charged to back to V_{DD} the V_{gs} of the P_{L-R} reduces to "0" which turns "off" the transistor P_{L-R} . Thus, the subthreshold current through P_{L-R} flows for a very small period of time which makes the power overhead due to this additional current negligible. However, the current through the PMOS P_{L-R} flows during the time when both nodes (O1 and O2) are discharging which ensures the robustness improvement as explained in section 3.2.

3.5. Sensitivity to Temperature and Bit-differential

The operating current in the proposed stabilization scheme is the *subthreshold leakage*. Since the subthreshold leakage current increases with the temperature [12], the effectiveness of the stabilization increases (i.e. failure probability reduces) at a higher temperature (Fig. 14a). If the sense amplifier is enabled at a lower bit-differential a smaller value of V_t mismatch can result in a failure (see (3)-(4)). Hence, if the input bit-differential at the time of sense amplifier firing is reduced the failure probability increases (Fig. 14b). Fig. 14b also shows that, for a target tolerable value of the failure probability, the stabilization scheme allows an early firing of sense-amp enable signal, which helps to reduce memory access time [4-15].

3.6. Effect on Memory Access Failure Probability

To evaluate the impact of the proposed stabilization technique on SRAM yield we have estimated the memory access failure probability using the proposed PMOS-BB CLSA sense-amplifier (using method described in section 2.3). The same SRAM cell is used for both conventional and PMOS-BB CLSA and δV_t is applied to all the transistors based on the sizes (following $\sigma_{V_t} = \sigma_{V_{t0}} \sqrt{W/W_0}$, W_0 = width of PMOS in the cell). It can be observed that, proposed sense-amplifier significantly reduces the memory access failure probability (Fig. 15). Moreover, the effect is increased at a higher temperature. The bit-differential produced by SRM cell reduces at a higher temperature (due to reduction in the "ON" current of access and pull-

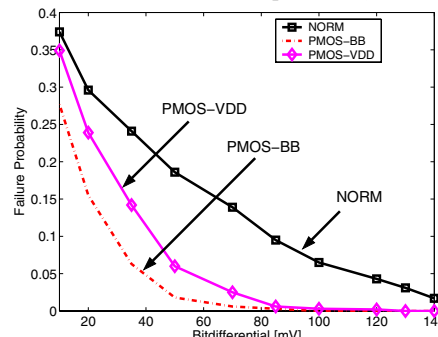
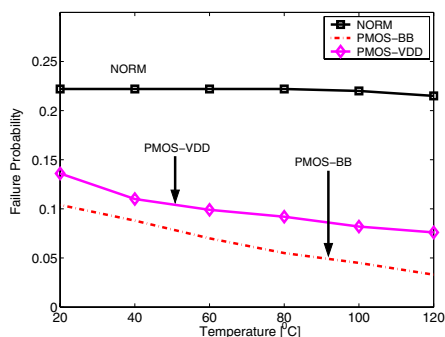


Fig. 14: Effect of (a) temperature and (b) input-bit-differential on failure probability of sense-amplifier.

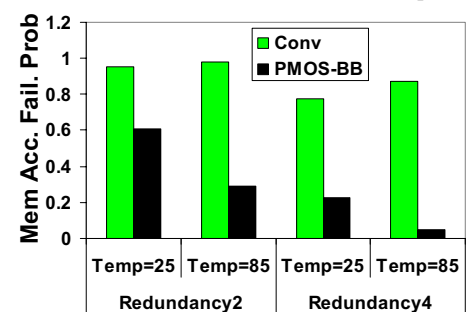


Fig. 15: Effect of stabilization technique on Memory access failure probability

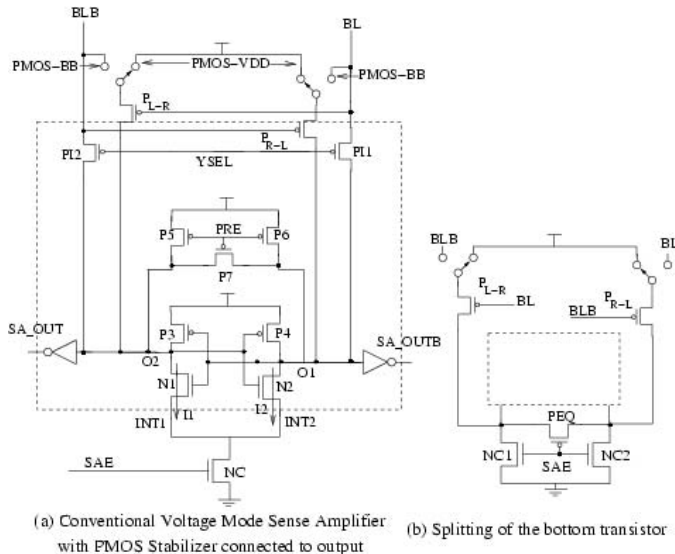


Fig. 16: Proposed PMOS Stabilization scheme applied to Conventional Voltage Mode Sense Amplifier

down NMOS transistors in the cell). In case of conventional CLSA offset increases marginally with temperature (as evident from Fig. 14a). Hence, the memory access failure increases with temperature. However, if leakage based stabilization is used, input offset reduces with temperature (as evident from Fig. 14a and eq. (15)). Hence, use of the proposed sense-amplifier counter-acts the effect of temperature on bit-differential. It was observed, the with proposed sense-amplifier access failure in memory reduces at a higher temperature. The effect of stabilization is more pronounced with higher redundancy (for an equal increase in P_{CS} , memory failure reduces more at higher redundancy, see (11)). The reduction of the memory access failure probability improves yield of an SRAM design.

3.7. Proposed Stabilization Technique for VLSA

The proposed stabilization scheme can also be applied to the voltage latch based sense amplifier as shown in Fig. 16a. In this case, as WL goes high, the $YSEL$ is turned low to turn on the input PMOS transistors $P11$ and $P12$. The pre-charge circuit is also turned off (by turning PRE high). After a sufficient bit-differential is developed (the same difference will also be developed between $O1$ and $O2$), the SAE is turned on and the sensing occurs. In the case of VLSA, the robustness can be improved by connecting the feed-forward path to the source (instead of drain as in PMOS-BB and PMOS-VDD case) of the driver transistors $N1$ and $N2$ (Fig. 16b). However, that will require splitting of the bottom transistor (NC) into two ($NC1$ and $NC2$) and addition of an equalizer (PEQ). The equalizer is controlled by SAE and it reduces the offset caused by the mismatch between $NC1$ and $NC2$. In this paper these schemes will be referred to as PMOS-BB-BOT and PMOS-VDD-BOT. If the feed-forward is applied to the output, the PMOS-VDD results in lower failure probability (Fig. 17). The transistor P_{L-R} in the PMOS-BB case is shunted by an "on"

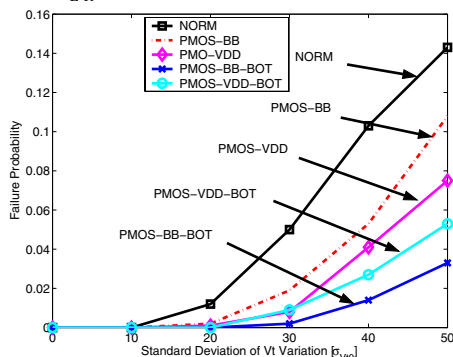


Fig. 17: Failure probability vs Vt variation in CVMSA.

PMOS ($P11$) which reduces its effectiveness (V_{DS} is very small). However, as discussed earlier PMOS-VDD increases the sensing delay. On the other hand, if PMOS-BB-BOT is used, the failure probability reduces as P_{L-R} is not shunted (Fig. 17). However, this requires $W_{NC1}=W_{NC2} > 0.5W_{NC}$ to maintain the sensing delay. Moreover, the division of NC increases the susceptibility to differential noise in the nodes $INT1$ and $INT2$. The sensitivity to the differential noise can be reduced by optimized layout which minimizes the coupling capacitances to the nodes $INT1$ and $INT2$. The extracted netlist with coupling parasitic capacitances has been simulated for differential noise tolerance and it has been observed that more than $2V_{DD}$ of differential noise is required to cause a functional failure.

4. Conclusion

In this paper, we have analyzed the access failure probability in memory considering both sense-amplifiers and SRAM cell. It is observed that, design of process variation tolerant sense-amplifier is of extreme importance to reduce access failure probability in memory and improving design yield. To improve the robustness of the sense-amplifier, we have proposed a feed-forward path based stabilization scheme for voltage mode sense amplifiers used in SRAM. In this technique the feed-forward transistors are introduced and their subthreshold leakage is utilized to reduce failures in sense amplifier due to intra-die parametric variation. As the subthreshold leakage is expected to increase exponentially with technology scaling, the effectiveness of the proposed technique improves in nano-scale technologies. The use of the proposed sense-amplifiers in SRAMs can significantly improve the reliability and yield of memories under process variations in nano-meter regime.

References:

- [1] A. Bhavnagarwala, et. al., "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE JSSC, vol. 36, no. 4, pp. 658-665, April 2001.
- [2] S. Mukhopadhyay, et. al., "Statistical design and optimization of SRAM cell for yield enhancement," ICCAD, 2004, Nov. 2004, pp. 10 - 13
- [3] T. Kobayashi, et. al., "A current controlled latch sense amplifier and a static power-saving input buffer for low-power architecture," IEEE JSSC, vol. 28, Apr., 1993, pp. 523-527.
- [4] B. Wicht, et. al., "Yield and speed optimization of a latch type voltage sense amplifier," IEEE JSSC, vol. 39, July, 2004, pp. 1148-1158.
- [5] T. Sakurai, "High-speed circuit design with scaled-down MOSFETs and low supply voltage," in Proceedings of IEEE ISCAS, May 1993, pp. 1487 - 1490.
- [6] S.A. Parke, Midwest Symp. on CAS, 1997, pp-209-212.
- [7] R. Sarpeshkar et. al., "Mismatch sensitivity of a simultaneously latched CMOS sense amplifier," IEEE JSSC, vol. 26, Oct. 1991 pp. 1413 - 1422.
- [8] T. Kawahara, et. al. "A high-speed, small-area, threshold-voltage-mismatch compensation sense amplifier for gigabit-scale DRAM arrays," IEEE JSSC, vol. 28, July 1993, pp. 816 - 823.
- [9] J.-W. Sub, et. al., "Offset-trimming bit-line sensing scheme for gigabit-scale DRAM's," IEEE JSSC, vol. 31, no. 7, July 1996, pp. 1025 - 1028.
- [10] K. Seno, et. al. "9 ns 16 Mb CMOS SRAM with offset reduced current sense amplifier," ISSCC, 1993, Feb. 1993, pp. 248 - 249, 297.
- [11] Y. Rabaey, Digital Integrated Circuits, Prentice Hall, 1996.
- [12] Y. Taur et. al, Fundamentals of Modern VLSI Devices, Cambridge Univ. Press, 1998.
- [13] BPTM: Berkeley Predictive Technology Model, www-device.eecs.berkeley.edu/~ptm/
- [14] H. Pilo, et. al. "A 0.9ns random cycle 36Mb network SRAM with 33mW standby power," Symposium of VLSI Circuits, 2004, June, 2004, pp. 284-287.
- [15] K. Itoh, VLSI Memory Chip, Springer 2001.
- [16] A. Papoulis, Probability, Random Variables and Stochastic Process, McGraw-Hill, 2001.
- [17] S. Dhong, et. al. "A 4.8GHz fully pipelined embedded SRAM in the streaming processor of a CELL processor", ISSCC, Feb, 2005, pp. 486-488.