IEEE TRANSACTIONS ON
CIRCUITS AND SYSTEMS
I: REGULAR PAPERS

Modeling and Mitigation of Static Noise Margin Variation in Subthreshold SRAM Cells

Nan Zheng, *Student Member, IEEE*, and Pinaki Mazumder, *Fellow, IEEE*

***Abstract*— In energy-constrained applications, SRAM systems operating in the subthreshold region are often deployed to reduce power consumption. Subthreshold SRAM designs, however, confront numerous challenges such as susceptibility to process variation and reduced on-off current ratio. Statistical modeling of the variation in cell stability is critical in SRAM design, especially for designs operating in the subthreshold region where the process and temperature variations are the most pronounced. In this paper, statistical models for estimating static noise margins (SNM) of SRAM cells are built from the perspective of a shifted voltage transfer characteristic. Read (Hold) SNM of a subthreshold 8T cell is analyzed. It is shown that the distribution of a single-sided Read SNM is a weighted sum of several normal distributions instead of a regular Gaussian distribution. The proposed statistical model is verified with simulation results in 65 nm technology. Furthermore, to mitigate performance and yield degradation, an adaptive body biasing circuit is developed. It is demonstrated through simulation that, with a negligible area and power overhead, the proposed circuit achieves a 15% improvement in the worst-case Read SNM.**

***Index Terms*— Subthreshold memory, SRAM modeling, static noise margin, variation, adaptive body biasing.**

I. Introduction

In portable devices and implantable devices, memory operating in the subthreshold region is often used to prolong battery life. The dynamic power scales with the square of the supply voltage; the leakage power is also reduced when lowering the supply voltage. Despite the success in reducing power consumption by operating circuits in the subthreshold region, reliability becomes the most serious issue. Reduced supply voltage shrinks both the write and read (henceforth, capital letter will not be used for Read and Write operations) noise margin significantly. A more problematic issue is the serious degradation in the worst-case read stability and writability because of global and local process variations. To counteract the serious reliability degradation in low supply voltage, an 8T SRAM cell was proposed in [1] where a separate read port was employed to achieve a disturb-free read. Since then, various low-voltage SRAM cells with a dedicated read port have been proposed, including the leakage reduced 10T cell in [2], and 9T [3] and 10T cells [4] that support a bit-interleaving array with the aim of reducing the impact of soft errors.

Characterizing variation in cell stability is necessary, considering the large number of cells involved in an SRAM system. This is even more critical for SRAM systems operating in the subthreshold region because of the larger variations in static noise margin (SNM). Monte Carlo (MC) simulations are often carried out to capture the worst-case read stability. However, MC simulations for a large array are often computationally prohibitive. Importance sampling [5, 6] was proposed as a powerful technique to characterize the variation of SNM. By transforming the sampling density function, MC simulations based on importance sampling is much more efficient than traditional full MC simulations. Nevertheless, many samples are still needed. Therefore, an accurate statistical model in an analytical or a semi-analytical form is able to help designers estimate the performance of the SRAM under process variation in the early design phase without resorting to time-consuming MC simulations repeatedly.

Static noise margin for superthreshold SRAM cells is modeled in [7-9] and the subthreshold model is derived in [10, 11]. Statistical analysis of superthreshold SRAM under process variation is discussed in [7, 12, 13]. For a superthreshold SRAM cell, variation of SNM is often estimated by a weighted sum of the variation in threshold voltage of each transistor. Employing this approach in a subthreshold SRAM, however, is questioned in [10], as the weight of the threshold voltage variation of one transistor is dependent on the threshold voltages of other transistors.

To counteract performance and yield degradation caused by variation, body biasing is introduced as an effective and efficient technique. In [14], an operational amplifier is employed to adjust the body bias. This method, however, has at least three disadvantages. First, an op-amp with a wide input common mode range and a rail-to-rail output is hard to build under subthreshold supply voltages. Second, a current source, a block that is often absent in SRAM, is needed. Third, offset of the op-amp might introduce significant error. In [15], two inverters are employed as error amplifiers and feedback is used to correct the body bias. However, stability of the feedback system is not discussed and the circuit only supports a balanced P-N ratio. In an SRAM cell, however, a stronger NMOS is often preferred for a more robust write operation. In [16], an open-loop system is employed to generate a proper bias. However, considering process, voltage, and temperature variation, a feedback system is more suitable.

In this paper, statistical modeling of SNM for subthreshold SRAM cells is presented. Threshold voltage variation in a cell inverter is lumped into one term called transition voltage variation, which is more directly related to SNM variation. A non-linear mapping from transition voltage to SNM variation is studied in Section II. The mapping function is obtained from a simple circuit simulation and is approximated with a piecewise linear fitting. The final statistical model of the SNM variation is derived in closed-form. An 8T cell is used for demonstration, even though the obtained results also apply

Manuscript received. This work is supported by CCF1421467.
The authors are with University of Michigan, Ann Arbor, MI 48105 USA (e-mail:zhengn@umich.edu; mazum@umich.edu)

Digital Object Identifier: 10.1109/TCSI.2017.2700818

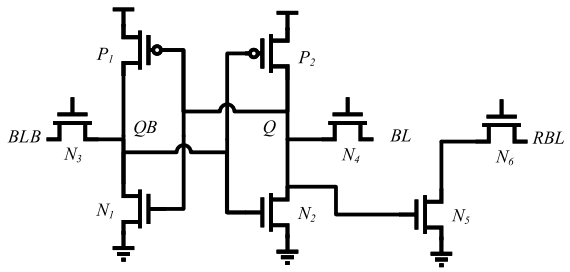


Fig. 1. Schematic of an 8T SRAM cell.

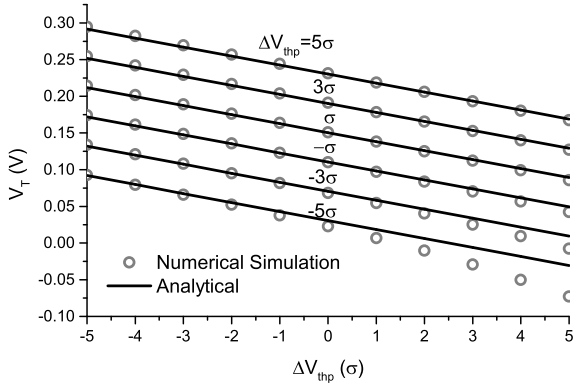


Fig. 2. Transition voltages of an SRAM cell inverter as the threshold voltage of NMOS and PMOS transistor varies. Solid line is obtained with (3), whereas circles are obtained from the circuit simulation tool.

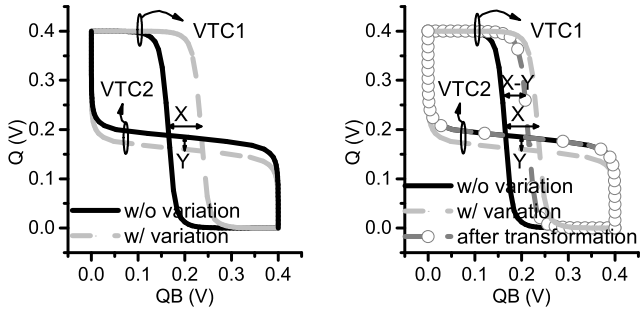


Fig. 3. Illustration of (a) two VTCs in the nominal case and two VTCs under process variation, and (b) VTCs under variation that are shifted by the same amount such that VTC2 is moved back to its original location.

to other 8T derivatives, which are the most popular cells used in subthreshold designs. The proposed model provides a convenient way to quickly estimate the distribution of SNMs and the data retention voltage (DRV). Furthermore, to mitigate the performance and yield degradation caused by the variation in SNM, an adaptive body biasing circuit is also proposed to compensate for the global variation. Simple circuit structure is used to achieve a negligible area and power consumption penalty. A stabilization technique inspired by the ring amplifier is introduced to stabilize the feedback loop. With the proposed adaptive body biasing circuit, it is demonstrated that a 15% improvement in the worst-case SNM can be achieved.

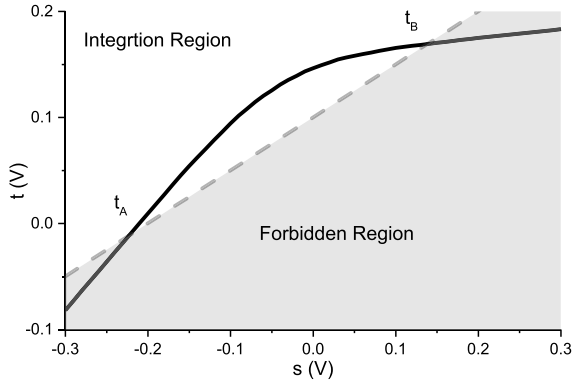


Fig. 4. Plot of a mapping function. t_A and t_B are the lower and upper limit for the integral in (23)

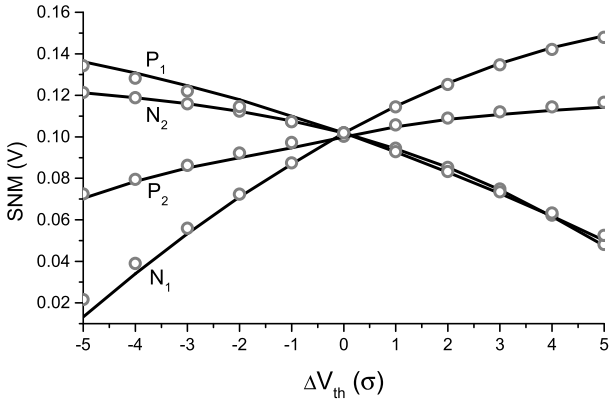


Fig. 5. Comparison of sensitivities obtained from $g(s)$ and the circuit simulation tool. Two sets of results match well, demonstrating that $g(s)$ can be employed for estimating the SNM.

II. Modeling of Variations in SNM

Fig. 1 shows the typical configuration of an 8T SRAM cell. In the figure, the two cell inverters consist of N_1, P_1 and N_2, P_2 are of interest. N_3 - N_6 can be omitted for the purpose of analyzing read (hold) SNM except that one has to be cautious about the systematic mismatch introduced by N_5 and N_6.

The voltage transfer characteristic (VTC) of an inverter is an important tool to analyze the stability of an SRAM cell. Most voltage-based SNMs can be calculated if the VTCs of cell inverters are known. The drain current of an NMOS transistor operating in the subthreshold region can be modeled as shown in (1) [17].

$$I = I_0 \frac{W}{L} e^{\frac{V_{GS}-V_{th}}{nU_T}} \left(1 - e^{-\frac{V_{DS}}{U_T}}\right) \tag{1}$$

where V_{th} is the threshold voltage, U_T is the thermal voltage, n is the subthreshold factor, W and L are width and length of the transistor, and I_0 is a technology-dependent fitting parameter.

For a CMOS inverter, let us define a quantity, called transition voltage V_T, as the input voltage that drives output from the inverter to $V_{DD}/2$. Mathematically, this can be written as $VTC(V_T) = V_{DD}/2$ where $VTC(\cdot)$ is the voltage transfer characteristic of the inverter. We can obtain the expression of V_T as shown in (2) by equating the current on the PMOS

transistor and the current on the NMOS transistor in the inverter.

$$V_T = \frac{n_p V_{thn} + n_n V_{thp} + n_n V_{DD} + n_n n_p U_T \ln\left(\frac{I_{0p} W_p L_n}{I_{0n} W_n L_p}\right)}{n_n + n_p} \tag{2}$$

It is well known that variation of threshold voltage, which is mainly caused by the random dopant fluctuation (RDF), is the dominant source of variation of cell stability in a subthreshold memory system [10, 18]. This is because the exponential dependence of current on threshold voltage dilutes impacts brought by other sources. Therefore, only the variation in the threshold voltage is modeled in this paper. Threshold voltages are typically modeled as independent, normally distributed random variables with standard deviations of the form $A_{V_{th}} \Big/ \sqrt{WL}$ [19], where $A_{V_{th}}$ is a technology-dependent constant. Therefore, with the help of (2), it can be shown that transition voltage of a cell inverter obeys normal distribution with a standard deviation of

$$\sigma\left(V_T\right) = \sqrt{\left(\frac{n_p}{n_n + n_p}\right)^2 \frac{A_{V_{thn}}^2}{W_n L_n} + \left(\frac{n_n}{n_n + n_p}\right)^2 \frac{A_{V_{thp}}^2}{W_p L_p}} \tag{3}$$

The transition voltage obtained from (3) and the one from the simulation tool are compared in Fig. 2. In this paper, 65 nm technology is employed in the simulation for demonstration purposes and all simulation results are obtained with the Cadence Spectre. The threshold voltages of both NMOS and PMOS transistors vary from -5σ to 5σ, corresponding to a worst variation of $5\sqrt{2}\sigma$. The analytical results matched well with the numerical results except for the region where V_T is negative or close to zero. This discrepancy, however, is forgivable as a near- or sub-zero V_T implies an unreliable cell that cannot hold the data. In practice, most SRAM designs have the objective of controlling the worst-case SNM so it is larger than zero with some margin.

In the noise margin calculation, two single-sided SNMs can be found, which correspond to two lobes of the butterfly curve. The double-sided SNM of an SRAM cell, which characterizes the worst-case noise margin, is defined as the minimum of these two single-sided SNMs. For analysis purposes, results for a single-sided SNM called SNM_s are developed first. The results will then be used to find the double-sided SNM called SNM_d. In order to find the single-sided SNM variation, a mapping function $g\left(x, y\right)$ is needed such that

$$M = g\left(X, Y\right) \tag{4}$$

where X and Y are random variables representing variations in transition voltages of two VTCs used to calculate SNM, and M is the single-sided SNM we are interested in.

It is difficult to get an analytical expression for $g\left(\cdot\right)$ accurately. One alternative way is to obtain $g\left(\cdot\right)$ numerically by running through different x and y. Careful examination of the definition of SNM, however, reveals that a function $g\left(\cdot\right)$ with one argument is sufficient for the mapping. Suppose the variation in transistors pushes VTCs of the inverters away from the original places by the amount of X and Y, as shown in Fig. 3 (a). Then, the change in V_T is also X and Y. By shifting both VTC1 and VTC2 by $-Y$, VTC2 returns to its nominal place as shown in Fig. 3(b). Therefore, the single-sided SNM can be obtained as

$$M = g\left(S\right) + Y \tag{5}$$

where

$$S = X - Y. \tag{6}$$

Therefore, only a one-argument function $g\left(s\right)$ is needed. Function $g\left(s\right)$ can be obtained by keeping the VTC of one inverter fixed and shifting the VTC of another inverter for different s. The obtained single-sided SNM is the value of $g\left(s\right)$.

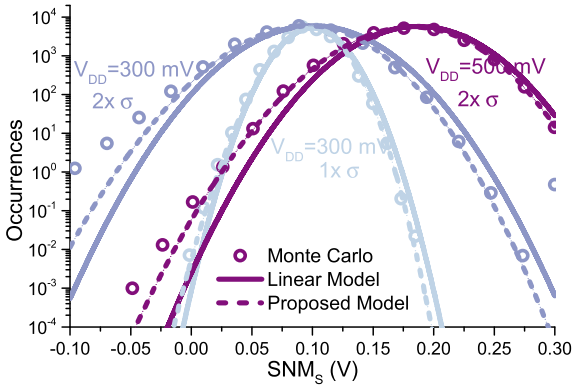


Fig. 6. Comparison of distributions of single-sided SNMs obtained from the conventional linear superposition model, proposed model and MC simulations. Results obtained with exaggerated (2x) variations are also plotted to show the trend as variations grow in advanced technologies. Results obtained from the proposed model matches better with the results obtained from MC simulations. Estimation error exists in the region where SNM is negative. The discrepancy between the proposed method and the conventional method increases as the level of variations grows.

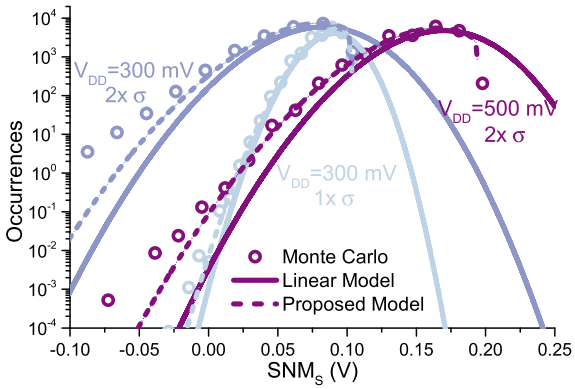


Fig. 7. Comparison of distributions of double-sided SNMs obtained from the conventional model, proposed model and MC simulations. Results obtained with exaggerated (2x) variations are also plotted to show the trend as variations grow in advanced technologies. Results obtained from the proposed model match better with the results obtained from MC simulations. Estimation error exists in the region where SNM is negative. The discrepancy between the proposed method and conventional method increases as the level of variations grow. The proposed model matches the entire distribution very well, whereas the conventional mode is only able to match the tail.

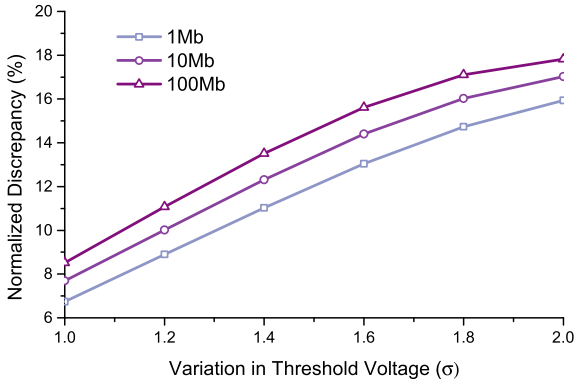


Fig. 8. Discrepancies between the proposed model and the conventional model under different levels of variations. Three different sizes of memory array are considered.

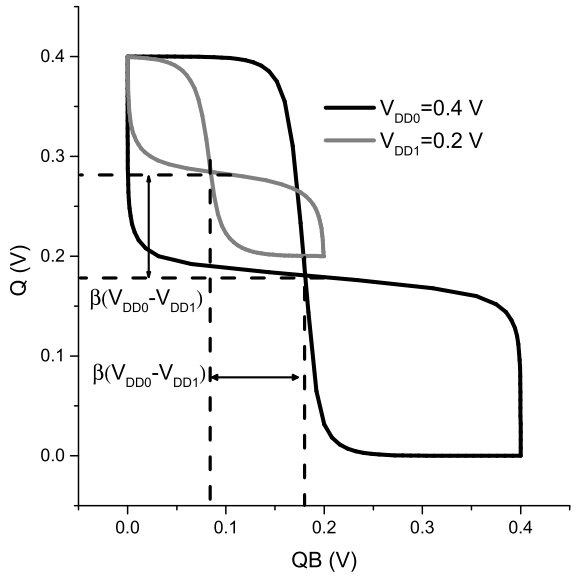


Fig. 9. Butterfly curves for an SRAM cell that is in read (hold) state under different supply voltages. Note that VTCs for the lower supply voltage are shifted up for illustration purpose. Change of supply voltage can be treated as shift in VTCs.

To obtain an analytical expression for the distribution of a single-sided SNM, another random variable is defined as

$$T = g\left(X - Y\right). \tag{7}$$

Without loss of generality, let us focus on the single-sided noise margin corresponding to the upper-left lobe of the butterfly curve. $g\left(s\right)$ is a strictly increasing function in this case. Through the change of variable [20], the probability density function (PDF) of T is given by

$$f_T\left(t\right) = \frac{f_S\left(g^{-1}\left(t\right)\right)}{g^{'}\left(g^{-1}\left(t\right)\right)}. \tag{8}$$

Therefore, the PDF of M can be obtained as

$$f_M\left(m\right) = \int_{-\infty}^{+\infty} \frac{f_X\left(g^{-1}\left(t\right)+m-t\right) f_Y\left(m-t\right)}{g^{'}\left(g^{-1}\left(t\right)\right)} dt. \tag{9}$$

A typical $g\left(s\right)$ is plotted in Fig. 4. There are three regions in $g\left(s\right)$. When s is small, SNM_s is proportional to s with

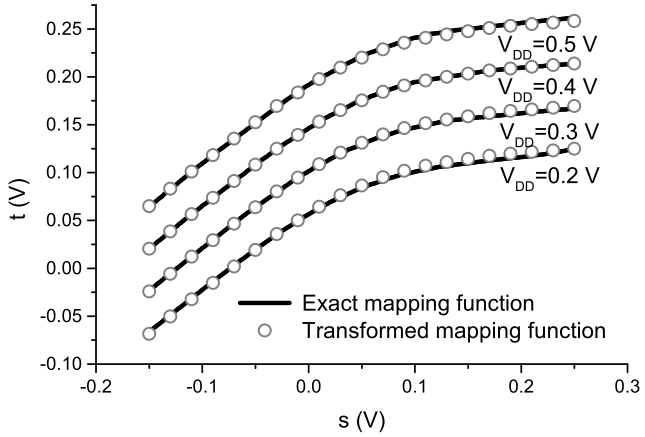


Fig. 10. Comparison between mapping functions for different supply voltages that are adapted from the one obtained with a 0.4 V supply voltage and mapping functions directly obtained from the circuit simulation tool.

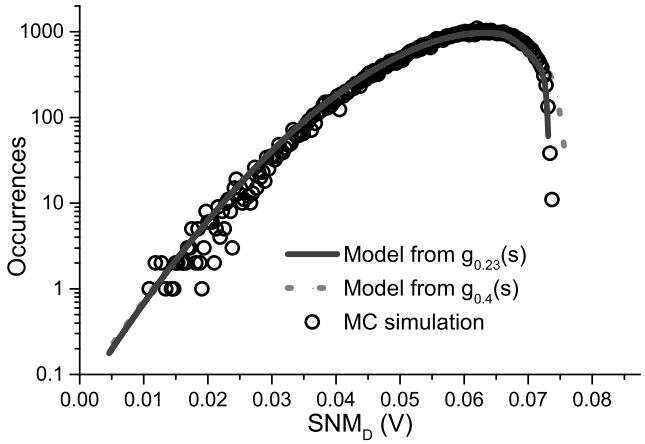


Fig. 11. Histogram of the hold SNM for an SRAM cell. Results obtained from both the adapted and non-adapted mapping functions are shown. Good agreements with 70000-run MC simulations are achieved.

a slope approximately equal to 1, because this is the region where VTC1 constrains SNM_s. When s is large, SNM_s is proportional to s with a slope approximately equal to 0 because this is the region in which VTC2 constrains SNM_s. It is shown in the Appendix that the slopes for VTC1-constrained region and VTC2-constrained region are

$$k_{VTC1-constrained} = \frac{G}{G+1} \tag{10}$$

and

$$k_{VTC2-constrained} = \frac{1}{G+1} \tag{11}$$

where G is the absolute value of slope of the VTC in the transition area.

Between the VTC1-constrained region and the VTC2-constrained region, both VTCs have significant impacts on the single-sided SNM. $g^{'}\left(s\right)$, in this case, gradually changes from $k_{VTC1-constrained}$ to $k_{VTC2-constrained}$. Based on the nature of this curve, it is convenient to do a piecewise linear fitting on $g\left(s\right)$ such that

$$g\left(s\right) \approx k_i s + a_i \tag{12}$$

(a) (b)

Fig. 12. (a) Conceptual diagram of the proposed adaptive body biasing circuit. (b) Simplified equivalent circuit of (a)

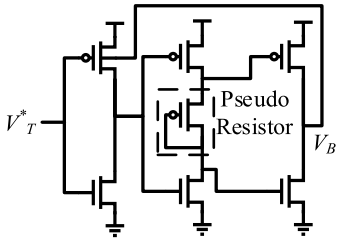


Fig. 13. One example of generating body bias to mitigate the read SNM degradation when global variation is present. V_T^* is the desired transition voltage.

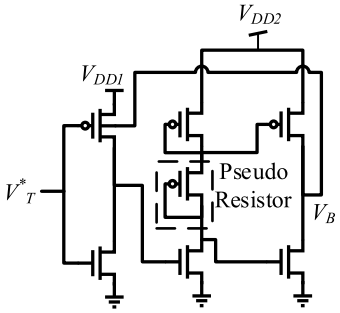


Fig. 14. Another example of generating body bias to mitigate read SNM degradation when the global variation is present. In the figure, $V_{DD2} > V_{DD1}$. Compared to the circuit in Fig. 13, the tuning range of the body bias is extended.

for

$$s \in (s_i, s_{i+1}), \quad i = 1, 2, \cdots, N \tag{13}$$

where (s_i, t_i) are knots for the piecewise linear fitting.

Then, (9) can be approximately written as

$$f_M(m) \approx \sum_{i=1}^{N} \int_{t_i}^{t_{i+1}} \frac{1}{k_i} f_X\left(\frac{t-a_i}{k_i} + m - t\right) f_Y(m-t)\, dt. \tag{14}$$

As X and Y are normally distributed random variables with a mean of zero and standard deviations shown in (3), integral in (14) can be analytically derived and the final expression is

$$f_M(m) = \sum_{i=1}^{N} \left(C_i(m, t_{i+1}) - C_i(m, t_i)\right) \phi\left(\frac{m-a_i}{\overline{\sigma}_i}\right) \tag{15}$$

where

$$C_i(m,t) = \Phi\left(\frac{(t-m)\,\bar{\sigma}_i^2 + (a_i - m)(k_i - 1)\,\sigma_Y^2}{k_i \sigma_X \sigma_Y \overline{\sigma}_i}\right) \tag{16}$$

$$\overline{\sigma}_i = \sqrt{k_i^2 \sigma_X^2 + \left(k_i^2 - 1\right)^2 \sigma_Y^2} \tag{17}$$

In the equations, $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and the cumulative density function (CDF) of the normal distribution.

It is noted that all information needed for calculating the SNM is contained in $g(s)$, given (3) is held. To illustrate this, SNM calculated based on (3) and (5) are compared with results obtained directly from the simulation tool in Fig. 5. Two sets of results match well. Compared to the conventional linear model, the proposed method requires similar or even less amounts of circuit simulations to model the SNM variations. For example, 41 circuit simulations are necessary to generate 41 VTCs for points in Fig. 5 in order to utilize the linear model, whereas only one circuit simulation is needed for the proposed algorithm to generate the VTC in the nominal case. Obtaining SNMs in both cases are conducted in a high-level programming tool. The remaining part of the proposed algorithm is conducted analytically, which can be done efficiently and conveniently once the algorithm is programmed in a script. For SRAM cells operating in the superthreshold region, it is generally assumed that the total SNM variation is a linear combination of SNM variations caused by each transistor. Therefore, the slope of each curve at $\Delta V_{th} = 0$ in Fig. 5 is employed as the sensitivity of the SNM with respect to the threshold voltage of that transistor. Correctness of this approach, however, is challenged in [10], as it is found that sensitivity curves shown in Fig. 5 are dependent on each other. One observation from (15) is that the PDF of a single-sided SNM is the sum of several modulated Gaussian PDF, meaning that the distribution of SNM_s is not necessarily normal. In the case where the variation of transition voltage, which is ultimately determined by the variation in threshold voltages, is not large, variation of s is kept local. $g(s)$, in this case, can be represented by one straight line. Consequently, (15) can be simplified to

$$f_M(m) = \phi\left(\frac{m - m_0}{\overline{\sigma}}\right) \tag{18}$$

where

$$\overline{\sigma} = \sqrt{k^2 \sigma_X^2 + \left(k^2 - 1\right)^2 \sigma_Y^2} \tag{19}$$

and m_0 is the SNM in the nominal case and k is the localized slope of $g(\cdot)$.

Equation (18) shows that the distribution of a single-sided SNM is normal when the variation is not too large, which is the assumption used in most literature. Consequently, the conventional way of representing the total SNM variation as

a linear combination of SNM variations caused by transistors is an acceptable approximation. However, when the variation is large, the distribution deviates from a Gaussian one. This observation is particularly useful, as the variation of the threshold voltage is growing as the technology advances if we consider the fact that transistors used for SRAM cells are usually with minimum sizes.

Another interesting observation made from (19) is that the variation of single-sided SNMs achieves its minimum when $k = 0.5$, considering $\sigma_X = \sigma_Y$. This means that the variation of a single-sided SNM is smaller when the butterfly curve is in neither the VTC1-constrained region nor the VTC2-constrained region. Intuitively, if the butterfly curve is biased in the region where both VTCs have significant effects on the SNM, then the reduced single-sided SNM, which is caused by the shift of one VTC, has a higher chance of being compensated by the shift of another VTC. Therefore, $V_T = V_{DD}/2$ not only provides the maximum nominal SNM but also features the smallest SNM variation.

Simulation results of single-sided SNMs obtained from (15) are compared with 20000-run MC simulations results obtained from the circuit simulation tool in Fig. 6. To increase the efficiency of the MC simulations, an importance sampling with 1.5 times larger standard deviation is employed. One nominal case circuit simulation is needed to help build the mapping function. Shifting of the VTC and calculation of SNM_s are conducted in a programming tool. (3) is employed to obtain the values of σ_X and σ_Y. For comparison purposes, SNM distribution obtained with the conventional superposition method [7, 12, 13] are also plotted. To demonstrate the trend as the variation in threshold voltage increases, results obtained with doubled variations are also plotted. In Fig. 6, distributions obtained from MC simulation match well with our proposed model for the region where SNM is positive. When the SNM is negative, distribution obtained from the MC simulations start deviating from the proposed model, as a consequence of the discrepancy between the estimated V_T and measured V_T shown in Fig. 2. However, this estimation error when SNM is negative (or close to zero), again, is forgivable as one rarely cares about SNM that is close to zero or even negative. To illustrate this point, the same SRAM array with a supply voltage of 0.5 V is also simulated. This is a more realistic supply voltage for an SRAM array with such a large threshold voltage variations. In this case, the proposed estimation method accurately predicts the distribution of the SNM. There exists a discrepancy between the proposed model and the conventional linear superposition model, and this discrepancy grows as the variation increases.

After obtaining the distribution of SNM_s, tail of the PDF of the double-sided SNM can be estimated by [10]

$$f\left(SNM_d\right) = 2 f_{SNM_s}\left(1 - F_{SNM_s}\right) \tag{20}$$

The exact distribution of SNM_d can actually be derived from the proposed model as shown in (21).

$$\begin{aligned} &P\left(SNM_D < m\right) \\ &\quad = 2P\left(SNM_{S,H} < m, SNM_{S,H} < SNM_{S,L}\right) \end{aligned} \tag{21}$$

In (21), $SNM_{S,H}$ and $SNM_{S,L}$ represent two single-sided SNMs corresponding to the upper-left and lower-right lobes, respectively. Symmetry of the two single-sided SNM is employed to simplify the equation.

The inequality $SNM_{S,H} < SNM_{S,L}$ is true if (22) holds.

$$g^{-1}\left(T\right) - 2T + 2M < 0 \tag{22}$$

Inequality in (22) corresponds to the unshaded region shown in Fig. 4. Suppose two intersects of the mapping function and the boundary of the region corresponds to (22) are t_A and t_B. As shown in (10) and (11), the slope of the mapping function changes from a value that is larger than $1/2$ to a value that is smaller than $1/2$ considering the small-signal gain of an SRAM cell inverter should be larger than 1 to hold the data. This means the mapping function $g\left(\cdot\right)$ is a concave function and the integration is always taken from t_A and t_B.

Therefore, the PDF of the double-sided SNM is given by

$$f_{SNM_D}\left(m\right) = \int_{t_A}^{t_B} \frac{f_X\left(g^{-1}\left(t\right) + m - t\right) f_Y\left(m - t\right)}{g'\left(g^{-1}\left(t\right)\right)} dt. \tag{23}$$

Once (23) is obtained, the piecewise linear fitting technique described before can be used to calculate the PDF of the double-sided SNM. Equations similar to (15)-(17) can be obtained except, in this case, the sum only runs through knots between t_A and t_B. One thing should be noted is that t_A and t_B are functions of m. Therefore, they have to be calculated for each m that is of interest.

Distributions of the double-sided SNMs obtained from the model in (23) are compared with results estimated from the conventional linear model and MC simulations, as shown in Fig. 7. The proposed model predicts the whole distribution well, whereas the conventional model can only predict the tail with limited accuracy. To show the trend of increasingly worsened variations in advanced technology, results obtained with twice as many variations as the employed technology (65 nm) are also shown. A similar trend is observed in the double-sided SNMs as well: the proposed method can accurately predict positive SNMs. We also show the comparison of results obtained with a supply voltage of 0.5 V. This is, again, a more realistic supply voltage for such serious threshold voltage variations. As the variation gets worse, discrepancies between results obtained from MC simulations and the conventional linear superposition method increase, whereas the proposed method matches well with the MC simulation for regions where SNMs are positive.

To give a quantitative comparison between the proposed model and the conventional model, discrepancies between the two models are compared in Fig. 8. Expectations of the minimum SNMs are obtained through two models. Fig. 8 shows the discrepancies between these two expectations that are normalized to the nominal SNM. The discrepancies corresponding to different array sizes and different levels of variations are illustrated. As the size and variation increases, the discrepancy grows rapidly. Because of the shrinking in transistor sizes, on-chip cache is becoming denser and larger, and the variation is becoming more serious. It can be foreseen

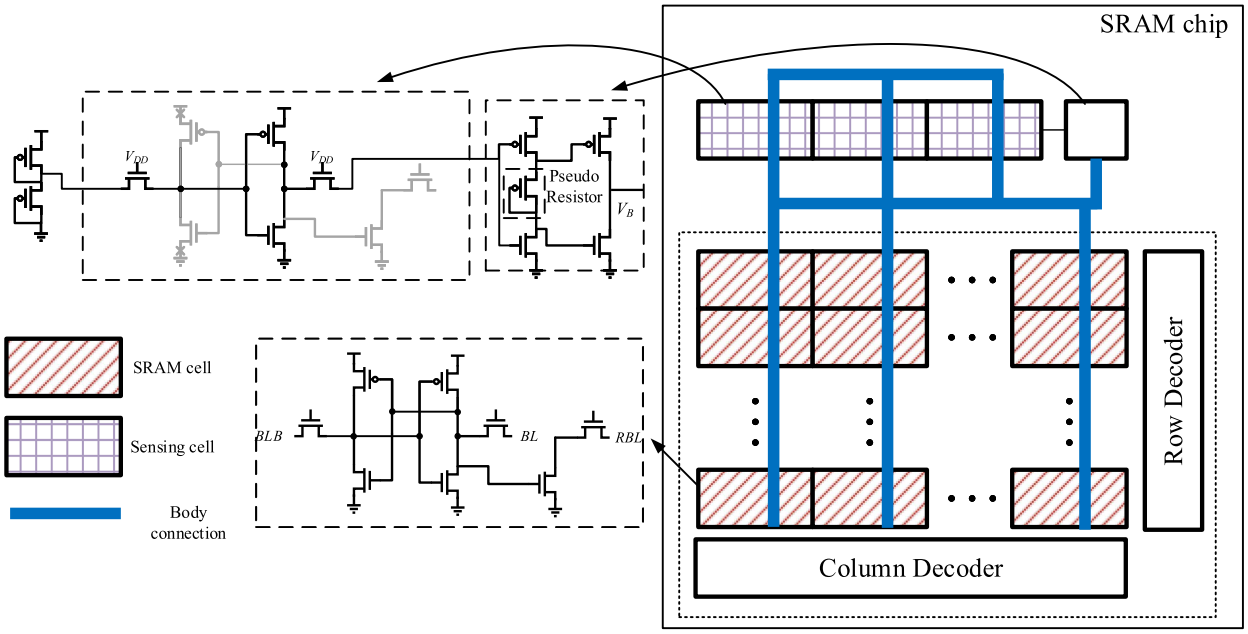


Fig. 15. Configuration of the proposed adaptive body biasing circuit in an SRAM array. Layout of the sensing inverters should resemble the SRAM cell to reduce the systematic mismatch.

in Fig. 8 that both of these trends will make the proposed modeling method more attractive compared to the conventional linear model.

One thing that should be pointed out is that the mapping function $g(\cdot)$is obtained for a specific supply voltage. Even though it is convenient to obtain the exact mapping function for different supply voltages with the help of a circuit simulation tool, the mapping function obtained under one supply voltage can be adapted for different supply voltages. This enables the proposed model to be used conveniently in a parametric analysis. As shown in [21], the small-signal gain of a subthreshold inverter in the transition region is approximately constant as the supply voltage decreases. The gain does not drop significantly until the term $1-exp\left(-V_{DS}/U_T\right)$ in (1) starts playing a role. Therefore, if we treat the slope of the VTC as a constant for different supply voltages, then the drop in supply voltage can be interpreted as the shift in VTC. Fig. 9 illustrates this idea. Two sets of butterfly curves for the supply voltages of 0.4 V and 0.2 V are plotted. In the figure, VTCs obtained with the 0.2 V supply voltage are shifted up by 0.2 V for the purpose of illustration. Following (2), let us assume that $V_T = \beta V_{DD} + C$, where β and C are two constants. In terms of the SNM, lowering the supply voltage behaves as shifting the VTC. Let the mapping function under the supply voltage V_{DD0} and V_{DD1} be $g_{V_{DD0}}(s)$ and $g_{V_{DD1}}(s)$, respectively. Then (24) can be used to obtain the mapping function under the supply voltage V_{DD1}, given $g_{V_{DD0}}(s)$.

$$g_{V_{DD1}}(s) = g_{V_{DD0}}(s) - \beta\left(V_{DD0} - V_{DD1}\right) \tag{24}$$

Such a linear relationship between the supply voltage and the mapping function provides a convenient way to obtain mapping functions corresponding to any supply voltage from mapping functions measured at two different supply voltages. The mapping functions adapted from the case of $V_{DD} = 0.4\ V$

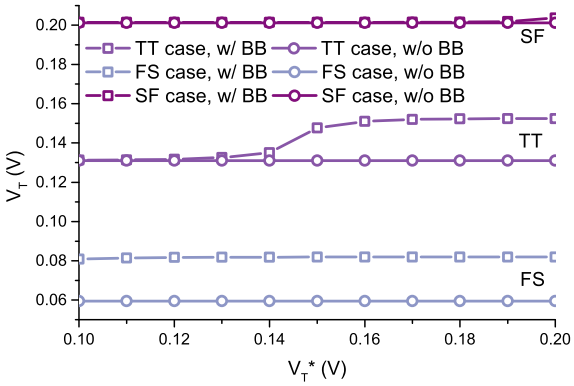


Fig. 16. Transition voltages of an SRAM cell inverter at different process corners. The actual transition voltage V_T tries to follow the desired transition voltage V_T^* with the help of the circuit shown in Fig. 13.

are plotted with exact mapping functions obtained from the simulation tool in Fig. 10. Good agreement is achieved. In the figure, the mapping function obtained under one supply voltage is the shifted version of the mapping function from another supply voltage.

The proposed statistical model can be readily used to estimate data retention voltage (DRV) of a subthreshold SRAM array. (25) should hold for an SRAM array with N cells such that the expectation of the number of cells which have negative hold SNMs is less than one.

$$P\left(SNM_D < 0\right) < \frac{1}{N} \tag{25}$$

One can sweep V_{DD} and calculate the probability of $P\left(SNM_D < 0\right)$ using (23), and DRV is the minimum supply voltage that allows (25) to hold. As an example, V_{DD} is swept for the example SRAM cell in this paper and the DRV for a 64

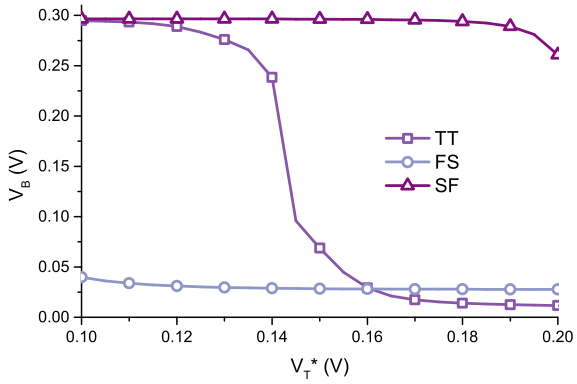


Fig. 17. Body bias voltage outputted by the proposed body biasing circuit in Fig. 13.

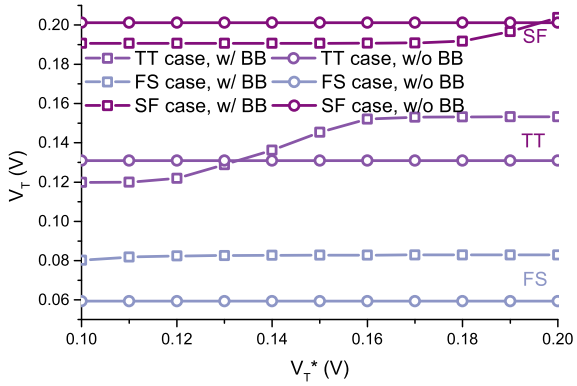


Fig. 18. Transition voltages of an SRAM cell inverter at different process corners. The actual transition voltage V_T tries to follow the desired transition voltage V_T^* with the help of the circuit shown in Fig. 14. V_T can be controlled better compared to the one shown in Fig. 16, as the body bias voltage has a larger tuning range.

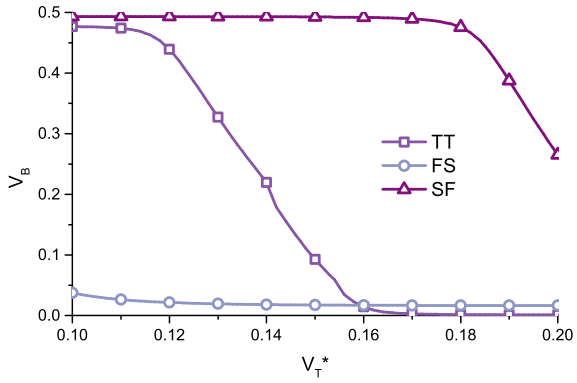


Fig. 19. Body bias voltage outputted by the proposed body biasing circuit in Fig. 14. Compared to Fig. 15, the body bias voltage has a larger tuning range.

KB SRAM is found to be 0.23 V. In Fig. 11, 70000-run MC simulation results are compared with results obtained from (23) using both adapted mapping function and non-adapted mapping function. As shown in the figure, both methods are able to predict the distribution accurately.

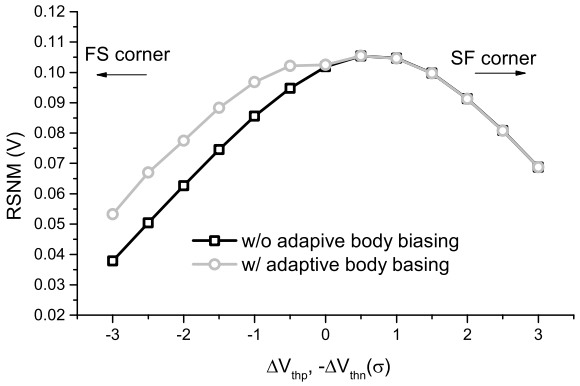


Fig. 20. Comparison of the read SNM obtained with and without the proposed adaptive body biasing circuit shown in Fig. 13 as the threshold voltages of PMOS transistor and NMOS transistor vary.

III. Counteract Global Variation

A. Adaptive Body Biasing Technique

It has been shown in Section II that the variation in SNMs is a major issue in subthreshold memory design and it becomes more and more serious as the technology advances. Certain cell designs are more robust against variations [22]; yet, it is not clear how to generalize this counter-variation feature to other commonly used cells such as the read disturb-free 8T cell. To counteract threshold voltage variations, one straightforward way is to increase the transistor sizing in a cell [18]. This way of remedy, however, introduces a large area penalty. Mismatches in transistors because of random dopant fluctuation are uncorrelated [19]. Therefore, compensation for local variation is not feasible. Global variation, however, can be sensed and compensated accordingly. To mitigate the threshold voltage variation, an adaptive body biasing technique is introduced in this section. A simple yet powerful body bias generation circuit is proposed, which senses global threshold voltage variations and then adjusts the body bias of SRAM cells for compensation. The circuit conceptually consists of only inverters, as shown in Fig. 12(a). To better understand how the proposed circuit functions, its simplified equivalent circuit is shown in Fig. 12(b). The first-stage sensing inverter senses and amplifies the difference between the desired transition voltage V_T^* and the actual transition voltage V_T. Two other inverters are employed as an error amplifier to further amplify the error. The error signal is then fed back to the cell inverter for a feedback control. The block T shown in the figure is the transfer function that models the relationship between the bias voltage of the cell inverter and the transition voltage of that inverter.

In this paper, we use the body bias of PMOS transistors as an example, as this type of body biasing does not require any special option such as the triple-well process. Nevertheless, the proposed technique can be readily applied to the body biasing of NMOS transistors. Two examples of the proposed biasing circuit are shown in Fig. 13 and Fig. 14. The circuit shown in Fig. 14 requires an additional supply voltage so that the tuning range of the body bias can be enlarged. Despite its simplicity, one thing that needs to be ensured for the proposed adaptive

biasing circuit is the stability. Inspired by the stabilization technique employed in a ring amplifier [23], an offset is introduced by the pseudo-resistor in Fig. 13 and Fig. 14 in order for the two transistors at the output stage to conduct a small amount of current in the steady state. Therefore, a dominant pole is formed at the output stage, stabilizing the feedback circuit. In addition, the output of the biasing circuit is connected to bodies of many transistors in an SRAM array. The parasitic capacitance and body leakage current also help stabilize the loop.

B. Overheads

Fig. 15 shows a diagram of how the proposed body biasing circuit can be included in an SRAM array. The layout of the sensing inverters should resemble inverters in an SRAM cell to reduce systematic mismatch. One possible way to achieve this similarity is to have the same device placement and interconnect routing in the sensing cells as those in an SRAM cell. Contacts, however, are used differently to form the connection needed in a sensing inverter, as shown in Fig. 15. Furthermore, in order for the sensing inverters to accurately capture the global threshold voltage, the number of sensing cells K has to be large enough to overcome the local variation. If we assume the biasing circuit has a large enough gain to bias the transition voltage of the sensing inverter to the desired voltage, then the variation (both global and local) of the transition voltage of an SRAM cell inverter with the adaptive biasing is

$$\sigma^{'}(V_T) = \sqrt{\sigma^2_{sense}(V_T) + \sigma^2(V_T)} \tag{26}$$

where $\sigma_{sense}(V_T) = \sigma(V_T) \Big/ \sqrt{K}$.

The above result is obtained by assuming there is no spatial correlation in the threshold voltage variation. This assumption is supported by [19], where no noticeable spatial correlation in threshold voltage was observed. To suppress variation introduced by the sensing cell inverters, we require that $\sigma_{sense}(V_T) \ll \sigma(V_T)$. According to (26), a K of 16 is enough to effectively control variation introduced by inaccurate sensing to be around 3% of the local mismatch in the target technology.

The second stage of the proposed circuit splits input from the sensing stage into two outputs with an offset. The pseudo-resistor employed also helps in tracking variations in supply voltage. The output stage is used to bias bodies of PMOS transistors. Both the huge amount of parasitic capacitance associated with body nodes and leakage current from the body help in stabilizing the circuit. It is found in the implementation that a phase margin of around 90 degrees can be met once the output is applied to an array of SRAM.

Current flowing in the output stage should be designed larger than the junction leakage current such that a moderate gain can be kept to maintain the correct voltage value. Overhead, in terms of the power consumption, can be estimated as

$$\frac{I_{ABB}}{I_{DLeak}} = \frac{K}{2N} e^{\frac{V_{DD}}{2nU_T}} + \frac{BI_B}{I_D} \tag{27}$$

where K is the number of sensing inverters, and N is the number of SRAM cells, B is a design parameter with a vaule around 1, I_B is the junction leakage current, and I_D is the subthreshold leakage current. When the size of the SRAM array is small, the power consumption of the proposed adaptive biasing circuit is dominated by the sensing stage, corresponding to the term $\frac{K}{2N} e^{\frac{V_{DD}}{2nU_T}}$ in (27). This is because a moderate number of inverters are needed to overcome the local variation. For example, a rough estimation according to (27) is that 16 sensing inverters incur a power consumption that is less than 1% of the leakage power of an SRAM array with 40 K cells. The power consumption of sensing inverters is quickly diluted as the size of SRAM array increases. When the size of the SRAM array grows large, power is mostly consumed by the output stage in order to counteract the body leakage. This power consumption scales linearly with the size of the SRAM array. Junction leakage in a transistor is normally much smaller than the subthreshold leakage [24], which keeps the term $I_B \big/ I_D$ small. It is found that power consumption of the proposed biasing circuit is less than 0.4% and 1% of the leakage power of a 100 Kb SRAM under a supply voltage of 0.3 V for circuits shown in Fig. 13 and Fig. 14, respectively.

C. Implementation Examples

Simulations are carried out to demonstrate the effectiveness of the proposed body biasing circuit. Fig. 16 compares the transition voltage V_T of an SRAM cell inverter with and without the proposed body biasing. The comparison is conducted for various desired transition voltage V_T^* under different process corners. Fig. 17 shows how the corresponding body bias varies with V_T^* under different process corners. In the TT (NMOS typical/PMOS typical) case, V_T obtained with the proposed adaptive body bias follows the desired transition voltage V_T^* applied at the input of the biasing circuit. As V_T^* increases, the body bias signal V_B generated by the proposed circuit decreases to strengthen the PMOS transistor in order to shift the VTC towards the right. In this case, Fig. 17 shows that the body bias V_B can swing rail to rail. This is one advantage of the proposed body biasing scheme. In the FS case (NMOS fast/PMOS slow), the generated body bias stays at a low voltage, attempting to compensate for the weak PMOS transistor because of a global threshold voltage variation. Limited by the finite tuning range of the body bias, the actual V_T is not able to follow the desired V_T^* accurately. Nevertheless, the resultant V_T is closer to the desired value, leading to a less spreading V_T. In the SF corner (NMOS slow/ PMOS fast), the body bias signal is close to V_{DD}, leading to a V_T close to the uncompensated value. This is because the body bias is bounded by the supply voltage. In order to provide a more effective compensation for the SF corner, the circuit shown in Fig. 14, which has a wider tuning range, is simulated. Obtained results are compared in Fig. 18 and Fig. 19. The generated body bias and obtained V_T in the FS corner are similar to those achieved in Fig. 16 and Fig. 17, as the lowest body bias available in both circuits are zero. In the TT case, however, the range where V_T follows V_T^* obviously enlarges, thanks to the larger tuning range of V_B.

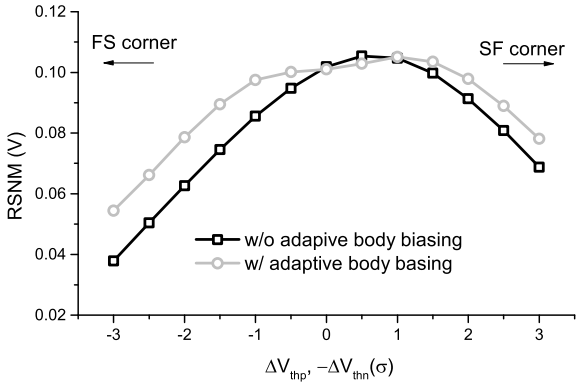


Fig. 21. Comparison of the read SNM obtained with and without the proposed adaptive body biasing circuit shown in Fig. 14 as the threshold voltages of PMOS transistor and NMOS transistor vary. Improvement in SNM is achieved at both SF and FS corners, as the body bias has an extended tuning range.

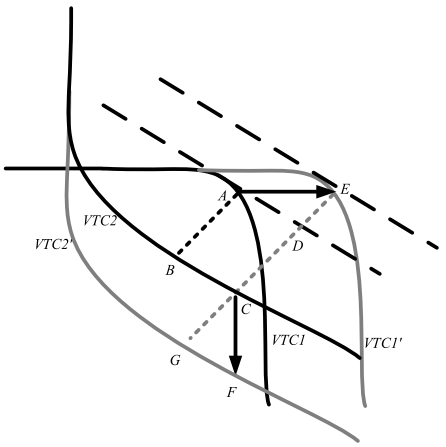


Fig. 22. Illustration of VTCs for two back-to-back inverters in the VTC2-constraint case.

This is illustrated in Fig. 19. In both the TT and SF corner, the body of the PMOS transistor can be reversely biased to weaken the PMOS transistor that is stronger than desired because of process variations.

To study how the proposed adaptive body biasing circuit can be helpful in mitigating the degradation of SNM because of global process variation. Fig. 20 and Fig. 21 compare the double-sided SNMs of an SRAM cell with and without the proposed body bias scheme under different levels of variations. In the figures, variations in threshold voltages of both NMOS and PMOS transistors are swept together to study the worst-case scenario. For example, in Fig. 20 and Fig. 21, the numbers on the x-axis represent the amount of variations ΔV_{thn} and $-\Delta V_{thp}$. As shown Fig. 20, the read SNM is improved wherever a body biasing is effective. 15% improvement is achieved for the worst-case SNM with the proposed biasing technique. The percentage of improvement is normalized with respect to the nominal SNM. In Fig. 21, SNMs at both the SF and FS corners are improved as the body bias voltage can swing both sides.

It is worth pointing out that even though the proposed adaptive biasing circuit is aimed at reducing SNM variation caused by global process variations, it can also be used for variations introduced by other sources, e.g. the negative bias temperature instability (NBTI) effect [25]. In [25], an open-loop compensation scheme was used. Operational amplifiers were used in that work, as a superthreshold supply voltage was considered. It is, however, difficult to employ standard operational amplifiers in a subthreshold design. In addition, the circuit in [25] can only compensate threshold voltage variation for PMOS transistors, whereas the biasing technique proposed in this paper is effective for threshold voltage variations for both PMOS and NMOS transistors. Compared to the body biasing circuit in [15], the stability of the feedback in our proposed circuit is fully considered and discussed. Furthermore, the proposed circuit has better control over the transition voltage of a cell inverter. This is particularly useful in an 8T cell, whereas the transition voltage is typically lower than $V_{DD}/2$ for a more robust write operation.

IV. Conclusions

In this paper, the statistical modeling of SNM for a sub-threshold 8T SRAM cell is presented. Variations in the read SNM are obtained from transition voltage variations with the help of a mapping function. It is shown that the distribution of the single-sided SNM is not necessarily Gaussian. The proposed model can be employed to estimate distributions of hold and read SNMs for SRAM cells such as an 8T memory cell and its derivatives at the design time without resorting to time-consuming Monte Carlo simulations. Compared to the conventional way of estimating SNM as a linear combination of sensitivities with respect to variations in each transistor, the proposed method is more accurate while having a similar or even lower computational complexity, especially when the variation is large. Besides, the mapping function can be easily adapted for different supply voltages, facilitating a parametric analysis for estimating the DRV. To counteract global process variations, an adaptive body bias generation circuit is proposed. Feedback is employed to control the transition voltage of a cell inverter. The stabilization technique used in a ring amplifier is applied to the proposed bias generation circuit. Two examples are provided along with simulations results. It is shown that 15% improvement can be achieved for the worst-case read SNM.

Appendix

Fig. 22 shows two VTC curves: VTC1 and VTC2. A VTC2-constraint scenario is demonstrated in the figure. The transition region on VTC2 that we are interested in is modeled as a straight line. Because of variation, VTCs are shifted by ΔV_{T1} and ΔV_{T2} to VTC1' and VTC2'. In the figure, $\overline{AE}$ and $\overline{CF}$ are equal to ΔV_{T1} and ΔV_{T2}, respectively. $\overline{AB}$ and $\overline{EG}$ are diagonals of two biggest squares that can fit within the butterfly curves. As shown in [8], the diagonal of the square, which determines the SNM, passes two parallel tangents on two VTCs. Therefore, ABCD is a parallelogram, leading us to

$$\Delta SNM = \frac{\overline{DE} + \overline{GC}}{\sqrt{2}}. \tag{28}$$

With the help of the law of sines, (29) can be obtained.

$$\Delta SNM = \frac{1}{G+1}\Delta V_{T1} + \frac{G}{G+1}\Delta V_{T2} \tag{29}$$

where G is the absolute value of the slope of VTC2 before rotating. For an 8T cell in the read (hold) state, the asymptotic behavior of the mapping function shown in (10) and (11) can be obtained from (29) by setting ΔV_{T2} and ΔV_{T1} to zero, respectively.

REFERENCES

[1] L. Chang, R. K. Montoye, Y. Nakamura, K. A. Batson, R. J. Eickemeyer, R. H. Dennard, *et al.*, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *Solid-State Circuits, IEEE Journal of,* vol. 43, pp. 956-963, 2008.

[2] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation," *Solid-State Circuits, IEEE Journal of,* vol. 42, pp. 680-688, 2007.

[3] S. Pal and A. Islam, "9-T SRAM Cell for Reliable Ultralow-Power Applications and Solving Multibit Soft-Error Issue," *IEEE Transactions on Device and Materials Reliability,* vol. 16, pp. 172-182, 2016.

[4] C. Ik Joon, K. Jae-Joon, P. Sang Phill, and K. Roy, "A 32 kb 10T Sub-Threshold SRAM Array With Bit-Interleaving and Differential Read Scheme in 90 nm CMOS," *Solid-State Circuits, IEEE Journal of,* vol. 44, pp. 650-658, 2009.

[5] T. S. Doorn, E. J. W. ter Maten, J. A. Croon, A. Di Bucchianico, and O. Wittich, "Importance sampling Monte Carlo simulations for accurate estimation of SRAM yield," in *Solid-State Circuits Conference, 2008. ESSCIRC 2008. 34th European*, 2008, pp. 230-233.

[6] T. Kida, Y. Tsukamoto, and Y. Kihara, "Optimization of importance sampling Monte Carlo using consecutive mean-shift method and its application to SRAM dynamic stability analysis," in *Quality Electronic Design (ISQED), 2012 13th International Symposium on*, 2012, pp. 572-579.

[7] K. Agarwal and S. Nassif, "The Impact of Random Device Variation on SRAM Cell Stability in Sub-90-nm CMOS Technologies," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 16, pp. 86-97, 2008.

[8] J. Lohstroh, E. Seevinck, and J. de Groot, "Worst-case static noise margin criteria for logic circuits and their mathematical equivalence," *Solid-State Circuits, IEEE Journal of,* vol. 18, pp. 803-807, 1983.

[9] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *Solid-State Circuits, IEEE Journal of,* vol. 22, pp. 748-754, 1987.

[10] B. H. Calhoun and A. P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS," *Solid-State Circuits, IEEE Journal of,* vol. 41, pp. 1673-1679, 2006.

[11] R. Saeidi, M. Sharifkhani, and K. Hajsadeghi, "Statistical Analysis of Read Static Noise Margin for Near/Sub-Threshold SRAM Cell," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 61, pp. 3386-3393, 2014.

[12] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *Solid-State Circuits, IEEE Journal of,* vol. 36, pp. 658-665, 2001.

[13] K. Takeuchi, R. Koh, and T. Mogami, "A study of the threshold voltage variation for ultra-small bulk and SOI CMOS," *Electron Devices, IEEE Transactions on,* vol. 48, pp. 1995-2001, 2001.

[14] M. Sumita, S. Sakiyama, M. Kinoshita, Y. Araki, Y. Ikeda, and K. Fukuoka, "Mixed body-bias techniques with fixed Vt and Ids generation circuits," in *Integrated Circuit Design and Technology, 2005. ICICDT 2005. 2005 International Conference on*, 2005, pp. 233-234.

[15] P. Yu, J. Pineda de Gyvez, H. Corporaal, and H. Yajun, "An Ultra-Low-Energy Multi-Standard JPEG Co-Processor in 65 nm CMOS With Sub/Near Threshold Supply Voltage," *Solid-State Circuits, IEEE Journal of,* vol. 45, pp. 668-680, 2010.

[16] H. Mostafa, M. Anis, and M. Elmasry, "A Novel Low Area Overhead Direct Adaptive Body Bias (D-ABB) Circuit for Die-to-Die and Within-Die Variations Compensation," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on,* vol. 19, pp. 1848-1860, 2011.

[17] M. Alioto, "Ultra-low power VLSI circuit design demystified and explained: A tutorial," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 59, pp. 3-29, 2012.

[18] Z. Bo, S. Hanson, D. Blaauw, and D. Sylvester, "A Variation-Tolerant Sub-200 mV 6-T Subthreshold SRAM," *Solid-State Circuits, IEEE Journal of,* vol. 43, pp. 2338-2348, 2008.

[19] N. Drego, A. Chandrakasan, and D. Boning, "Lack of Spatial Correlation in MOSFET Threshold Voltage Variation and Implications for Voltage Scaling," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 22, pp. 245-255, 2009.

[20] J. H. Drew, D. L. Evans, A. G. Glen, and L. M. Leemis, "Transformations of Random Variables," in *Computational Probability*, ed: Springer, 2017, pp. 47-56.

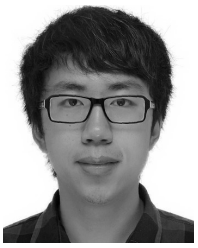
[21] M. Alioto, "Understanding DC Behavior of Subthreshold CMOS Logic Through Closed-Form Analysis," *Circuits and Systems I: Regular Papers, IEEE Transactions on,* vol. 57, pp. 1597-1607, 2010.

[22] S. Pal and A. Islam, "Variation tolerant differential 8T SRAM cell for ultralow power applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,* vol. 35, pp. 549-558, 2016.

[23] L. Yong and M. P. Flynn, "11.5 A 100MS/s 10.5b 2.46mW comparator-less pipeline ADC using self-biased ring amplifiers," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, 2014, pp. 202-203.

[24] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits* vol. 2: Prentice hall Englewood Cliffs, 2002.

[25] H. Mostafa, M. Anis, and M. Elmasry, "Adaptive body bias for reducing the impacts of NBTI and process variations on 6T SRAM cells," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 58, pp. 2859-2871, 2011.



Nan Zheng (S'13) received the B.S. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011, and the M.S degree in electrical engineering from University of Michigan, Ann Arbor, in 2014£¬ where he is currently working toward the Ph.D. degree in electrical engineering.

In the summer of 2012, he had an internship at Qualcomm, CA, where he worked on developing antenna system for the next-generation communication network. His research interests include low-power circuit design, modeling and optimization.

Pinaki Mazumder (S'84–M'87–SM'95–F'99) received the Ph.D. degree from the University of Illinois at Urbana-Champaign, Urbana, in 1988.

He is currently a Professor with the Department of Electrical Engineering and Computer Science, University of Michigan (UM), Ann Arbor. He was for six years with industrial R&D centers that included AT&T Bell Laboratories, where in 1985, he started the CONES Project—the first C modeling-based very large scale integration (VLSI) synthesis tool at India's premier electronics company, Bharat Electronics, Ltd., India, where he had developed several high-speed and high-voltage analog integrated circuits intended for consumer electronics products. He is the author or coauthor of more than 320 technical papers and five books on various aspects of VLSI research works. His current research interests include current problems in nanoscale CMOS VLSI design, computer-aided design tools, and circuit designs for emerging technologies including quantum MOS and resonant tunneling devices, semiconductor memory systems, and physical synthesis of VLSI chips.

Dr. Mazumder is a Fellow of the American Association for the Advancement of Science (2008). He was a recipient of the Digital's Incentives for Excellence Award, BF Goodrich National Collegiate Invention Award, and Defense Advanced Research Projects Agency Research Excellence Award.