Analyzing Static Noise Margin for Sub-threshold SRAM in 65nm CMOS

Benton H. Calhoun and Anantha Chandrakasan

MIT, 50 Vassar St 38-107, Cambridge, MA, 02139 USA {bcalhoun,anantha}@mtl.mit.edu

Abstract:

This paper evaluates the static noise margin (SNM) of 6T SRAM bitcells operating in sub-threshold. We analyze the dependence of SNM during both hold and read modes on supply voltage, temperature, transistor sizes, local transistor mismatch due to random doping variation, and global process variation in a commercial 65nm technology. We analyze the statistical distribution of SNM with process variation and provide a model for the tail of the PDF that dominates SNM failures.

1. Introduction

This paper provides a thorough evaluation of static noise margin (SNM) for SRAM bitcells functioning in the sub-threshold region. Sub-threshold digital circuit design has emerged as a low energy solution for applications with strict energy constraints, such as micro-sensor networks. Analysis of sub-threshold designs has focused on logic circuits (e.g. [1]). The large fraction of chip area often devoted to SRAM makes low power SRAM design very important as well. Recent low power memories show a trend of lower voltages with some designs holding state on the edge of the sub-threshold region (e.g. [2]). This scaling will continue, leading to sub-threshold storage modes and even sub-threshold operation for SRAM's operating in tandem with sub-threshold logic.

The minimum voltage for retaining bistability was theorized in [3] and modeled for SRAM in [4], but degraded SNM can limit voltage scaling for SRAM designs above this minimum voltage. SNM quantifies the amount of voltage noise required at the internal nodes of a bitcell to flip the cell's contents. Figure 1(a) shows the location of the noise sources in the 6 transistor (6T) bitcell schematic. Figure 1(b) provides the common graphical representation of SNM for a cell during read access and while holding data (un-accessed). The voltage transfer curves (VTCs) of the two inverters are shown with one VTC inverted. The SNM is found graphically

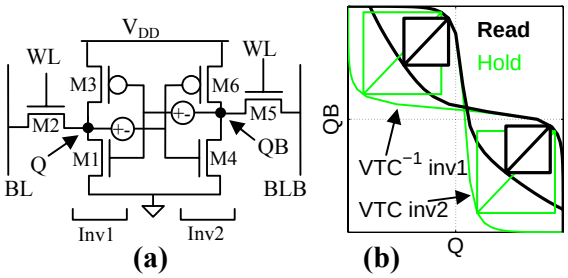

Figure 1: Schematic for 6T bitcell showing voltage noise sources for finding SNM (a). SNM plots (b).

as the length of the side of a square fitted between the VTCs and having the longest diagonal [5]. As the value of the noise sources increases, the VTC^{-1} for inverter 1 in the figure moves upward and the VTC for inverter 2 moves to the left. Once they both move by the SNM value, the curves meet at only two points. Any further noise flips the cell. Figure 1(b) also shows that the SNM is lower during read access because the VTC is degraded by the voltage divider across the access transistor (M2,M5) and drive transistor (M1,M4).

An expression for above-threshold SNM based on long-channel models is given in [5], and [6] models above-threshold SNM for modern processes with process variation. This paper builds on previous work by examining SNM for sub-threshold SRAM. The next section provides 1st-order equations for the VTCs in sub-threshold and evaluates their accuracy. Section 3 describes the impact of various parameters on SNM. Local mismatch due to random doping variation and global process variation provide the focus since these have the dominant effect on SNM.

2. Modeling Sub-V_T Static Noise Margin

It is well-known that lowering V_{DD} reduces gate current much more rapidly than sub-threshold current, so total current in the sub-threshold region is given by (1).

$$I_D = I_S \exp\left(\frac{V_{GS}-V_T}{nV_{th}}\right)\left(1-\exp\left(\frac{-V_{DS}}{V_{th}}\right)\right) \quad (1)$$

The sub-threshold factor $n=1+C_{ds}/C_{ox}$, $V_{th}=kT/q$, and I_S is the current when V_{GS} equals V_T. For simplicity, we treat PMOS parameters as positive values. For the 65nm technology used in this paper, the NMOS drive current is higher in above-threshold than the PMOS for iso-width, but the PMOS current is higher in sub-threshold. During hold mode, WL=0 so M2 and M5 have $V_{GS} \le 0$ and thus negligible current. We can model the cell VTCs ($V_{OUT} = f_{VTC}(V_{IN})$) as those of a simple inverter in sub-threshold.

$$QB = \begin{aligned} &V_{th}\frac{n_1 n_3}{n_1+n_3}\left(\ln\frac{I_{S3}}{I_{S1}}+\ln\left(\frac{1-\exp((-V_{DD}+Q)/V_{th})}{1-\exp(-Q/V_{th})}\right)\right) \\ &+\frac{n_1 V_{DD}}{n_1+n_3}+\frac{n_1 n_3}{n_1+n_3}\left(\frac{V_{T1}}{n_1}-\frac{V_{T3}}{n_3}\right)\end{aligned} \quad (2)$$

Equation (2) [3] gives the inverse VTC for inverter 1 ($V_{IN}=f_{VTC}^{-1}(V_{OUT})$). The inverse of (2) is given in [7] for matched PMOS and NMOS (same n, V_T, I_S). We give a full solution for $V_{OUT}=f_{VTC}(V_{IN})$ for inverter 2 in (3).

$$QB = V_{DD}+V_{th}\ln\left(0.5*\left(1-G+\sqrt{(G-1)^2+4e^{\frac{-V_{DD}}{V_{th}}}G}\right)\right),$$
$$G = \exp\left(\frac{n_4+n_6}{n_4 n_6 V_{th}}Q-\ln\frac{I_{S6}}{I_{S4}}-\frac{V_{DD}}{n_6 V_{th}}-\frac{1}{V_{th}}\left(\frac{V_{T4}}{n_4}-\frac{V_{T6}}{n_6}\right)\right) \quad (3)$$

Figure 2(a) plots (2) and (3) against simulation curves for no local mismatch and for 1σ V_T mismatch in M6.

During a read access, WL=V_{DD} and the bitlines are precharged to V_{DD} so, if Q=0 prior to access, M1 and M2 are both on. This creates a voltage division that raises the voltage at Q. Assuming PMOS current is negligible in the region of interest, (4) shows the inverse VTC equation near the SNM [2] for inverter 1.

$$QB = \begin{aligned} & n_1 V_{th} \ln\frac{I_{S2}}{I_{S1}} + n_1 V_{th} \ln\left(\frac{1-\exp((-V_{DD}+Q)/V_{th})}{1-\exp(-Q/V_{th})}\right) \\ & + V_{T1} + \frac{n_1}{n_2}(V_{DD} - V_{T2} - Q) \end{aligned} \quad (4)$$

This equation cannot be inverted analytically, and it applies only to the region of the VTC where V_{OUT} is low. Figure 2(b) shows (4) and its graphical inverse combined piecewise with (2) and (3) and plotted against simulation for no local mismatch and for 1σ V_T mismatch in M1 for minimum device sizes at 25°C.

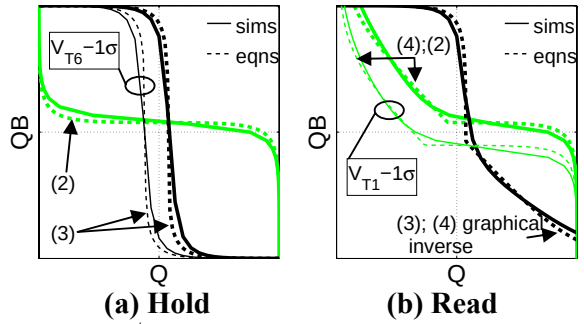

Figure 2: 1st-order VTC equations versus simulations.

Graphical or numerical solutions for SNM are easily derived from the VTC equations, although no direct analytical solution exists. The equations provide a good estimate of the behavior of the SNM based on key parameters. One shortcoming of (2)-(4) is the assumption that sub-threshold slope ($S=nV_{th}\ln 10$) is constant for each transistor. Figure 3(a) shows that S varies with V_{GS}, and Figure 3(b) shows S changing with temperature without the expected constant slope due to V_{th}. A more crucial problem with (2)-(4) is the assumption that certain currents are negligible. These assumptions break down under certain combinations of V_T variation, rendering the 1st-order equations inaccurate.

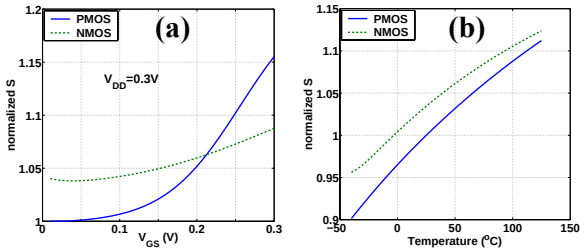

Figure 3: Changes in sub-threshold slope (S) versus V_{GS} (a) and temperature (b).

3. Sub-V_T SNM Dependencies

With embedded SRAM often providing multiple megabits of storage, the SNM of the nominal bitcell becomes largely irrelevant. Variations in processing and in the chip's environment create a distribution of SNM across the bitcells in a given memory, and the worst-case tail of this distribution determines the yield. This section examines the impact of different parameters on SNM in sub-threshold and offers a model for estimating the tail of the SNM density function for process variation.

3.1 Dependence on V_{DD}

SNM for a bitcell with ideal VTCs is still limited to $V_{DD}/2$ because of the two sides of the butterfly curve. An upper limit on the change in SNM with V_{DD} is thus ½. Figure 4 shows how SNM varies with V_{DD} for both hold and read mode. The slopes of the curves confirm that less than ½ of V_{DD} noise will translate into SNM changes.

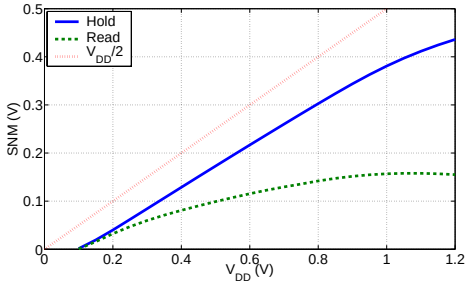

Figure 4: SNM versus V_{DD}.

3.2 Dependence on Temperature

Varying temperature from -40°C to 125°C only alters read and hold SNM by 21mV and 6mV, respectively. Higher temperatures lower SNM in sub-threshold due to the degraded gain in the inverters that results from worse sub-threshold slope (see Figure 3(b)). Also, PMOS devices weaken relative to NMOS at higher temperature.

3.3 Dependence on Sizing

In contrast to above-threshold [8], Figure 5 shows that cell ratio ($(W/L)_1/(W/L)_2$ or $(W/L)_4/(W/L)_5$) has very little impact on SNM during sub-threshold read. In fact, sub-threshold SNM sensitivity to any sizing changes is reduced. The lower impact of sizing is intuitively reasonable considering the exponential dependence of sub-threshold current on other parameters. Mathematically, we can see from (2)-(4) that sizing changes affect $I_{S,i}$ linearly and only have a logarithmic impact on the VTCs. One point of caution here is that V_T for deep submicron devices tends to vary with size as a result of narrow or short channel effects. The impact of this V_T change that might accompany a sizing change is more pronounced. These effects depend on the technology and make general SNM modeling more complicated.

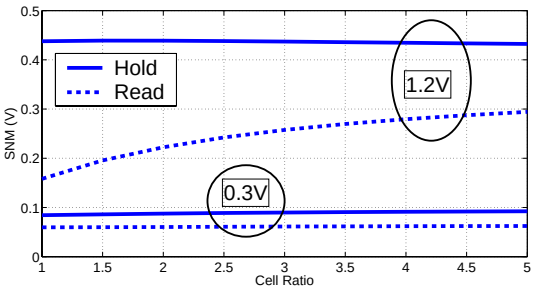

Figure 5: Cell ratio affects SNM less in sub-threshold

3.4 Dependence on Random Doping Variation

The randomness of the number of doping atoms and their placement in a MOSFET channel causes random mismatch even in transistors with identical layout [9]. The impact on threshold voltage, whose σ is proportional to $(WL)^{-½}$, is the worst for minimum sized devices which are common in SRAM. The exponential dependence of current on V_T in sub-threshold operation makes this random variation even more influential. Furthermore, the large number of bitcells in many SRAMs makes the tails (5-6σ) of the probability density function (PDF) more critical for modeling since the extreme cases are the limiting factor for yield.

Previous work has shown that above-threshold SNM is nearly linear with V_T, and modeling $\partial SNM / \partial VTi$ as a constant allows an approximation of the joint PDF for SNM [6]. Likewise, the sensitivity of above-threshold SNM to V_T is linearized for each transistor in [10]. Figure 6(a) shows that the sensitivity of SNM high (the upper-left box in Figure 1(b)) is nearly linear with each individual V_T. However, Figure 6(b) shows the relationship between SNM and V_{T4} for a few different values of the other V_Ts. The obvious dependence of the slope on the other V_Ts prevents using a model of the form $SNM = SNM_0 + \Sigma c_i V_{Ti}$ for sub-threshold SNM.

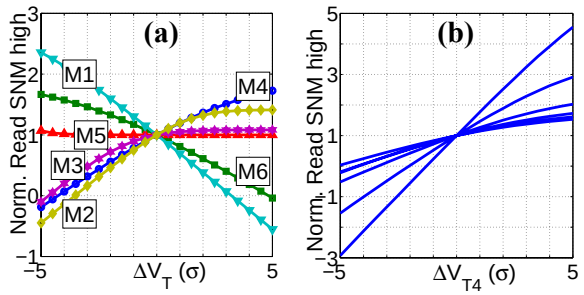

Figure 6: Dependence of SNM high on single FETs is nearly linear (a) but slope depends on other V_Ts (b).

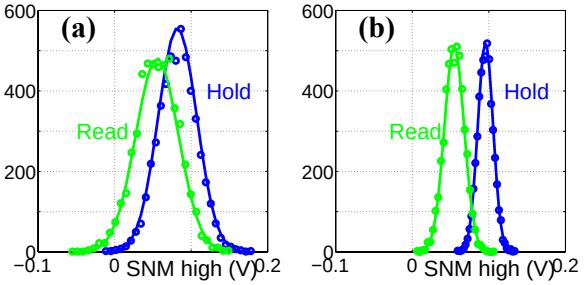

Figure 7: SNM high and low (not shown) for a min. sized cell (a) and for 4*WL (b) is normally distributed with random V_T mismatch in all transistors.

Figure 7 shows the results of 5k-point Monte-Carlo (M-C) simulations with random independent V_T mismatch in all transistors. These histograms confirm that sub-threshold SNM at the upper lobe of the butterfly curve (SNM high) is normally distributed. The solid lines show a fitted Gaussian PDF, and the markers show simulation results. Larger sizes for the bitcell clearly have the advertised effect of lowering the variance of V_T as seen in Figure 7 (b). The SNM low PDFs are very similar. The scatter plot in Figure 8 shows that SNM high and SNM low are correlated. The dependencies for mismatch in each single transistor are overlaid in white. The hold SNM shows a saturation effect along the upper edge. SNM high and SNM low are not independent because any change to a VTC that increases the SNM at one side tends to decrease SNM at the other side.

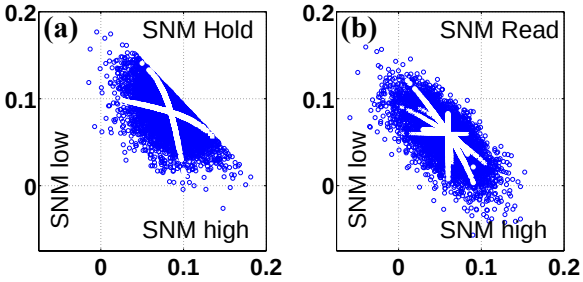

Figure 8: Scatter plots for SNM high vs. SNM low with single FET dependencies overlaid in white.

The actual SNM that matters for a bitcell is the minimum of SNM high and SNM low. Thus, the random variable $X_{SNM} = \min(X_{SNMhigh}, X_{SNMlow})$. Order statistics can provide us with the PDF for the minimum of *n* independent, identically distributed (*iid*) random variables, X_i. If *f* is the PDF, and *F* is the CDF for X_i, the PDF of the minimum of two *iid* variables is given in (5).

$$f(\min(X_1, X_2)) = 2 f_X (1 - F_X) \quad (5)$$

Although SNM high and SNM low are normally distributed with approximately the same mean and variance, we have previously shown that they are not independent. However, we are less interested in modeling the entire PDF for SNM than we are in modeling the worst-case tail. As previously stated, the tail toward lower SNM is the limiting factor. Let us assume that they are *iid*. Then we can write:

$$f_{SNM} = 2 f_{SNMhigh} (1 - F_{SNMhigh}) \quad (6)$$

Figure 9 shows the histogram for a 5k-point M-C simulation of read SNM plotted on linear axes (a) and semilog axes (b). Clearly, SNM is not normally distributed, and its mean is lower than the mean of SNM high and SNM low. Figure 9(b) shows that a Gaussian PDF does not match the worst-case tail on the left side of the PDF. On the other hand, the PDF based on (6) provides a good estimate of the tail. This PDF gives the powerful option of estimating the SNM at the worst-case end of the PDF without using extremely long M-C simulations.

Figure 10 shows several estimated PDFs using (6) that

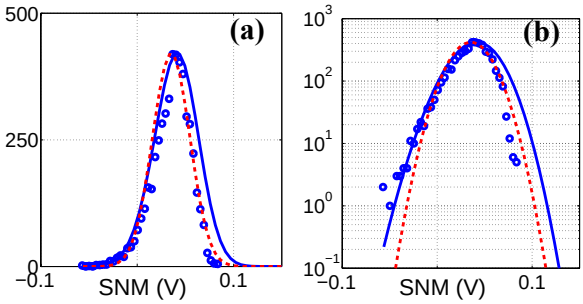

Figure 9: Histogram of SNM M-C simulation (circles) with normal PDF (dash) and PDF based on (6) (solid) over-laid. The semilog plot (b) shows that the PDF based on (6) matches the worst-case tail quite well.

are based on data sets of different lengths. These estimates are plotted over a 50k-point M-C simulation. A 1000-point M-C simulation gives an estimate that overlays the estimate from the 50k-point case on the plot (< 3% error). Using this approach allows a designer to reliably estimate the tail of the SNM PDF for a large memory with relatively few samples.

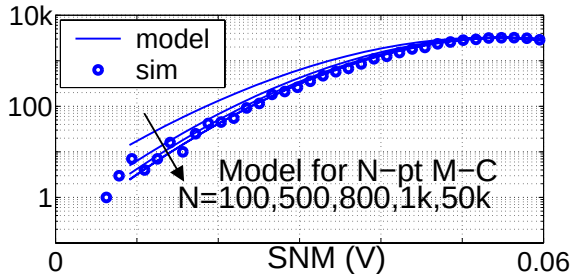

Figure 10: 50k-pt. M-C simulation for SNM with 4*WL sized transistors. Model based on 1k-pt. M-C overlays model based on 50k-pt. M-C.

3.5 Impact of Global Process Variation

Thus far we have assumed that device mismatch occurs in transistors that start off as typical for the process. In addition to the inter-die V_T mismatch that we have described is an intra-die process variation that sets the process corner (e.g. fast NMOS, slow PMOS, etc.). Even for no mismatch, the process corner impacts the SNM. Figure 11 shows the SNM PDF for a minimum sized 6T bitcell from a M-C simulation of global process corner in which nine process parameters are varied. Here again, the tail of the PDF is the limiting factor.

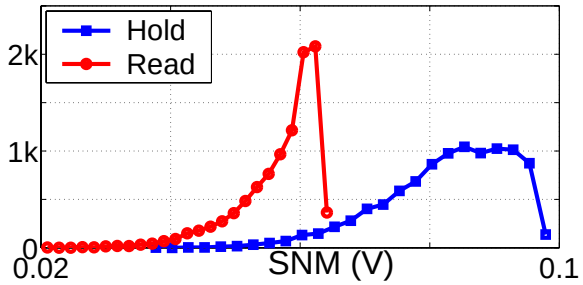

Figure 11: M-C simulation showing global variation impact on SNM for minimum sized bitcell.

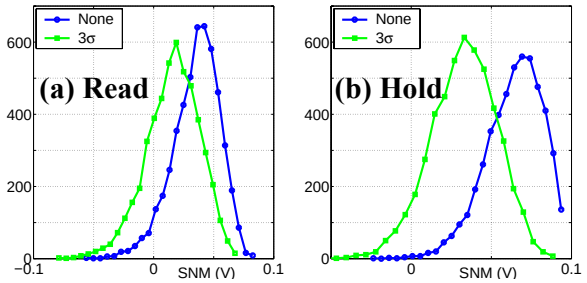

Figure 12: SNM M-C sims for mismatch on top of global variation (in legend).

In a production framework, each die containing a given SRAM will have a global process corner that affects SNM as in Figure 11. On top of this, mismatch in each cell will result from random doping variation. Assuming that any die within 3σ of the mean is usable, we found the global process corner that gives an SNM yield equivalent to -3σ for both hold and read cases. Figure 12 shows that the impact of mismatch at this 3σ process corner is essentially to shift the mean of the PDF by the offset caused by global variation. This means that the models we have presented remain valid for the case of combined global and local variation.

4. Conclusions

Static noise margin is a critical metric for SRAM bitcell stability. This paper has explored the impact of different parameters on SNM for SRAM bitcells in sub-threshold. The dominant factor affecting sub-threshold circuits in general and SNM specifically is V_T mismatch due to random doping variation, and the critical region for examination is the tail of the SNM PDF. We have shown that first-order theoretical models for calculating SNM are accurate close to the nominal values of V_T, but they cannot accurately account for all of the mismatch cases. We have shown that SNM high and SNM low are normally distributed with V_T mismatch and correlated. Despite their correlation, we have shown that treating them as *iid* leads to a PDF for SNM that gives an accurate model of the tail cases. This estimate is invaluable for avoiding long Monte-Carlo simulations in the design of large SRAMs for sub-threshold operation.

Acknowledgements

This work was funded by Texas Instruments and by the Defense Advanced Research Projects Agency (DARPA) through a subcontract with MIT Lincoln Laboratory. Thanks to Raúl Blázquez for many helpful discussions.

References

[1] A. Wang and A. Chandrakasan, "A 180mV FFT Processor Using Sub-threshold Circuit Techniques," *ISSCC*, 2004.

[2] A. Bhavnagarwala, et al., "A Transregional CMOS SRAM with Single, Logic VDD and Dynamic Power Rails, *Symp. on VLSI Circuits*, 2004.

[3] R. Swanson and J. Meindl, "Ion-Implanted Complementary MOS Transistors in Low-Voltage Circuits," *JSSC*, Vol. SC-7, No.2, pps 146-153, Apr. 1972.

[4] H. Qin, et al., "SRAM Leakage Supressiong by Minimizing Standby Supply Voltage," *ISQED*, 2004.

[5] E. Seevinck, F. List, and J. Lohstroh, "Static-Noise Margin Analysis of MOS SRAM Cells," *JSSC*, Vol. SC-22, No. 5, pps 748-754, Oct. 1987.

[6] A. Bhavnagarwala, D. Tang, and J. D. Meindl, "The Impact of Intrinsic Device Fluctuations on CMOS SRAM Cell Stability," *JSSC*, Vol. 36, No. 4, April 2001.

[7] E. Vittoz, "Weak Inversion for Ultimate Low-Power Logic," in Low-Power Electronics Design, ed. C. Piguet, CRC Press, 2005.

[8] B. Cheng, S. Roy, and A. Asenov, "The Impact of Random Doping Effects on CMOS SRAM Cell," *ESSCIRC*, 2004.

[9] R. Keyes, "The Effect of Randomness in the Distribution of Impurity Atoms on FET Threshold," *Appl. Phys.*, Vol. 8, pp 251-259, 1975.

[10] K. Takeuchi, R. Koh, and T. Mogami, "A Study of the Threshold Voltage Variation for Ultra-Small Bulk and SOI CMOS," *IEEE Transactions on Electron Devices*, Vol. 48, No. 9, September 2001.