2019 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S) | 978-1-7281-3523-6/20/$31.00 ©2019 IEEE | DOI: 10.1109/S3S46989.2019.9320728

# Thermal Noise-Induced Error Simulation Framework for Subthreshold CMOS SRAM

Elahe Rezaei*, Marco Donato†, William Patterson*, Alexander Zaslavsky*, R. Iris Bahar*
*School of Engineering, Brown University, Providence, RI USA
†John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA USA
{elahe_rezaei,william_patterson_iii,alexander_zaslavsky,iris_bahar}@brown.edu, mdonato@seas.harvard.edu.

***Abstract*—Accurate error-rate modeling in ultra-low-power subthreshold CMOS circuitry is needed to predict reliability. In this study, we extend our stochastic time-domain error simulation framework to thermally-induced bit-flip errors in ultimate CMOS SRAM cells. Our approach extracts the dependence of error rate on technological parameters such as operating voltage, threshold variability, and temperature. Our analysis tool extracts behavior that cannot be captured with conventional SPICE-based simulations and provides the first statistically rigorous tool for evaluating ultimate SRAM reliability.**

***Index Terms*—Thermal noise, subthreshold SRAM, time-domain error simulation.**

## I. Introduction

As transistors are scaled deep into the sub-10 nm regime, the small number of electrons on transistor nodes increases vulnerability to intrinsic, thermally-driven, noise voltage fluctuations, especially in subthreshold circuits where $V_{DD} < V_{TH}$, the device threshold voltage. Time-domain analysis of logical circuits in the presence of thermal noise has great utility for designers of noise-immune circuits. Conventional approaches for transient noise analysis assume stationary noise statistics for which thermal noise can be incorporated as small-signal additive white Gaussian noise sources [1]. A more rigorous simulation approach can be developed for subthreshold circuits: since subthreshold MOS currents obey Poisson statistics [2], they can be modeled as stochastic sources dependent on the instantaneous bias point, making it possible to derive charge fluctuations in nonlinear gates by solving stochastic differential equations (SDE) [3].

This study develops a simulation framework to compute the time-to-error (TTE) of a subthreshold SRAM latch due to intrinsic thermal noise, extending the SDE-based model introduced in [3] by considering the coupling between fluctuating nodal voltages in a latch. We present a parallelizable iterative algorithm to detect extremely rare bit-flip errors, which makes it possible to compute TTE as a function of temperature $T$, supply voltage $V_{DD}$, and transistor threshold $V_{TH}$ mismatch. While the noise immunity of circuits has been commonly quantified as mean-time-to-error (MTTE) [4] [5], we also demonstrate a new methodology for statistical TTE analysis. By running massive Monte-Carlo simulations facilitated by the iterative algorithm, we extract the cumulative distribution functions (CDFs) of TTEs to quantitatively evaluate SRAM reliability.

This work was supported by NSF under Grant CCF-1525486.

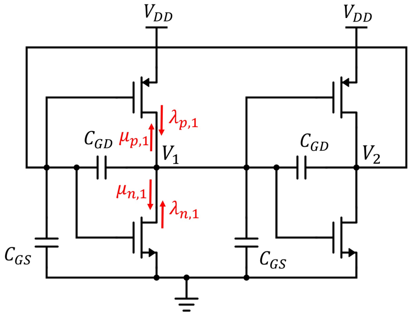

Fig. 1. The structure of CMOS SRAM latch and Poisson charging rates.

## II. Simulation Implementation

### A. SDE-based Formulation for CMOS SRAM

Figure 1 shows a CMOS SRAM cell, where $C_{GD}$ and $C_{GS}$ denote the gate-drain and gate-source capacitances. The Kirchhoff current law (KCL) equations at nodes 1 and 2 in Fig. 1 can be written in matrix form as:

$$\begin{bmatrix} C_{GS}+2C_{GD} & -2C_{GD} \\ -2C_{GD} & C_{GS}+2C_{GD} \end{bmatrix} \begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} = \begin{bmatrix} i_{D_1} \\ i_{D_2} \end{bmatrix} dt, \qquad (1)$$

where $i_{D_i}$ is the difference in the drain currents for PMOS and NMOS transistors of inverter $i$, $i \in \{1, 2\}$. In subthreshold, $i_D$ of a single transistor is formed by two opposing forward and reverse electron currents, described by two independent Poisson processes [2]. The rates for each transistor, shown as $\lambda_{n/p,i}$ and $\mu_{n/p,i}$ in Fig. 1, are derived from the forward and reverse components of the subthreshold drain current [3], as shown here for the NMOS device:

$$\mu_n = \frac{I_0}{q} \exp\left(\frac{qV_{ds}\lambda_D}{kT}\right) \exp\left(\frac{qV_{gs}}{mkT}\right), \qquad (2)$$

$$\lambda_n = \mu_n \exp\left(\frac{-qV_{ds}}{kT}\right), \qquad (3)$$

where $\lambda_D$ is the DIBL parameter and $I_0$ and $m$ are technology-dependent parameters. Therefore, in Eqn. (1), $i_{D_i}$ currents are aggregates of opposite electron flows (entering and exiting node $i$) described by two Poisson-distributed random variables with rates $(\lambda_{n,i} + \lambda_{p,i})$ and $(\mu_{n,i} + \mu_{p,i})$, as illustrated in Fig. 1. Converting Eqn. (1) to stochastic form, we compute nodal voltages $V_1(t)$ and $V_2(t)$ by repeatedly sampling Poisson distributions characterized by rates $\lambda_{n/p,i}$ and $\mu_{n/p,i}$ that are continuously updated using the instantaneous operating point for each transistor. To speed up the solution of the SDEs in

978-1-7281-3523-6/19/$31.00 ©2019 IEEE

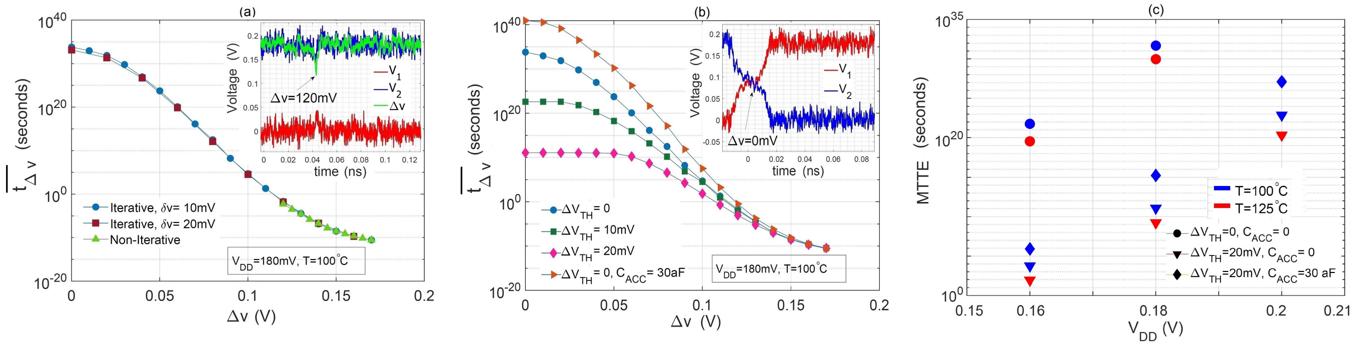

Fig. 2. (a) Thermal noise-driven $\overline{t_{\Delta v}}$ (the mean time to reach $\Delta v$ averaged over 50,000 runs) of a subthreshold CMOS latch at $V_{DD} = 180$ mV, inset shows an event of $\Delta v(t) = 120mV$ captured by our simulator; (b) $\overline{t_{\Delta v}}$ of mismatched CMOS SRAMs and a bit flip captured by the iterative algorithm; (c) MTTE as a function of $V_{DD}$, $T$ and $\Delta V_{TH}$, with and without access capacitance.

Eqn. (1), the rates are precomputed and saved in look-up tables for later use.

### B. Extraction of Thermal Noise-Induced Errors in an SRAM

The new simulator combines this improved statistical model of thermal noise response in SRAM cells with three orders of magnitude speed up compared to SPICE [3]. In equilibrium, the inter-nodal voltage difference in the latch, $\Delta v(t)$, can be expressed as $\Delta v(t) \equiv V_1(t) - V_2(t) \simeq V_{DD}$. A bit-flip error occurs when thermal noise drives $\Delta v(t)$ to zero and positive feedback then drives the latch to its other stable state. The Poisson processes governing the drain current imply that the evolution of $\Delta v(t)$ is independent of history. We exploit this in an iterative procedure for extracting the TTE as follows:

- We start the transient simulation by looking for a deviation from equilibrium, defined by $\Delta v(t) \leq \Delta v_1 \equiv V_{DD} - \delta v$, where $\delta v$ is a user-defined decrement step. Once the desired event is captured, the event time is recorded, and the corresponding nodal voltage values, $V_1[1]$ and $V_2[1]$, are saved as a checkpoint.
- The simulator progresses iteratively to the next stages by updating the voltage deviation as $\Delta v_n \equiv \Delta v_{n-1} - \delta v$. While tracking $\Delta v(t)$ at stage $n$, if $\Delta v(t) > \Delta v_{n-2}$, indicating a return towards equilibrium, a loop exit condition is triggered. Then, the simulation is reset to the last checkpoint, $V_1 = V_1[n-1]$ and $V_2 = V_2[n-1]$.
- This iterative process is repeated until $\Delta v(t) \leq 0$. The simulator then completes the run allowing the voltages to reach their opposite equilibrium point (a bit flip).

At any given stage, as long as $\Delta v(t)$ is much larger than 0, the internal feedback of the latch tends to return it to equilibrium, which is why a bit-flip error is exponentially unlikely. However, by counting loop exit conditions and discarding their data at each stage, the simulation run-time can be reduced greatly. The time to reach stage $n$ can be found via the recursive formula:

$$t_{\Delta v}[n] \approx (M_n + 1) \times t_{\Delta v}[n-1] + b_n, \qquad (4)$$

where $t_{\Delta v}[n-1]$ is the time to required to reach stage $(n-1)$, $b_n$ is the time spent in stage $n$, and $M_n$ counts the loop exits at stage $n$. In addition to the key improvement in simulation run-time provided by the iterative algorithm, multi-threading can be used to run multiple simulations in parallel and the computation time for bit-flip detection in an SRAM latch can be reduced to minutes.

### C. Simulator Setup

The simulation tool is entirely built in a C++ environment. To speed circuit simulation in the time domain, some parameter values such as mean currents and node capacitances as functions of bias point are extracted from the DC transfer curves of the inverter using SPICE. This process is done only once for a given transistor technology and device size. Two run-time parameters, the time step and the decrement step size $\delta v$, are user choices. During execution, Poisson distributed samples are generated to model electron flows, and inter-arrival times are counted within each time step, yielding the net change of charge from subtracting the discharging process from the charging process. The $i_{Di}$ values in Eqn. (1) are then updated, the SDEs solved, new values of $V_1$ and $V_2$ calculated, and the rates updated accordingly.

## III. Results and Discussion

For analysis of thermally induced errors in an SRAM with advanced CMOS fabrication nodes, we employed the 7nm predictive technology model from the ASAP7 PDK [6]. The SRAM circuit is designed based on minimum-sized low-$V_{TH}$ (LVT) transistors, for which $V_{TH} \simeq 250$ mV, and NMOS and PMOS $I_{ON}$ currents well-matched at $V_G = V_D = V_{DD}/2 =$ 90 mV. The initial results were generated as a set of Monte-Carlo simulations in subthreshold operation at $V_{DD} = 180$ mV, $T = 100°C$, and $C_{GS} = C_{GD} = 30$ aF (extracted from [6]). We set the time step as 0.5 ps. Figure 2(a) compares the variations of mean time $\overline{t_{\Delta v}}$ to reach $\Delta v$ for iterative and non-iterative algorithms. In the non-iterative direct simulation, the minimum $\Delta v(t)$ captured by our simulator after more than three weeks of continuous simulation time was 120 mV, far short of a bit flip. But the $\overline{t_{\Delta v}}$ curves from the iterative algorithm calculated via Eqn. (4) with $\delta v$ = 10 and 20 mV, can be obtained in minutes all the way to a bit-flip condition, while agreeing perfectly with the non-iterative case down to $\Delta v(t)$ = 120 mV (see inset of Fig. 2(a) for a representative set of $V_1(t)$, $V_2(t)$, and $\Delta v(t)$ traces).

The iterative procedure also provides MTTE to a bit flip, which for perfectly matched transistors at $T = 100°C$ occurs

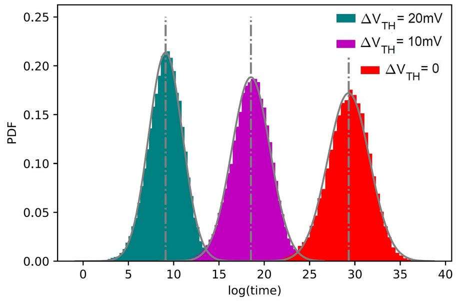

Fig. 3. The statistical distribution of TTE to a bit flip of subthreshold CMOS latch extracted by the iterative algorithm.

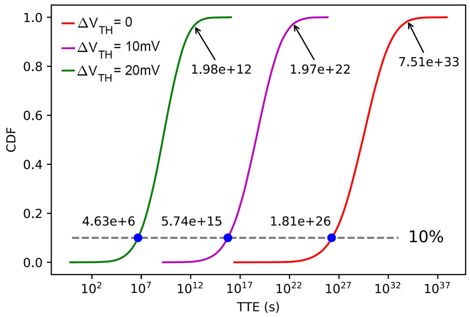

Fig. 4. The CDF of TTE to a bit flip of subthreshold CMOS latch extracted by the iterative algorithm. The upper values (with arrows) represent MTTE and lower numbers show the $10^{th}$ percentile for TTE.

after $\sim 10^{34}$ s (such a latch would be unconditionally stable). However, reliability degradation due to process variability and particularly to transistor $V_{TH}$ mismatch, is a major concern for nanoscale SRAM [7]. Allowing for an asymmetric worst-case threshold mismatch $\Delta V_{TH}$ between NMOS and PMOS transistors in the latch severely degrades its thermal noise immunity. Figure 2(b) compares the $\overline{t_{\Delta v}}$ for matched transistors and those with asymmetric $\Delta V_{TH} = 10$ and 20 mV, where we observe that $\Delta V_{TH} = 20$ mV mismatch (less than $10\%$ of the nominal $V_{TH}$) brings MTTE to a bit flip down to $\sim 10^{10}$ seconds per device (unacceptable in a memory containing $10^6$ latches). An inset in Fig. 2(b) shows an actual bit flip captured by our simulator. For a conventional 6T-SRAM bitcell with two access transistors in addition to the cross-coupled inverters [8], we include additional access transistor capacitance $C_{ACC} = 30$ aF. Figure 2(b) shows how additional $C_{ACC}$ partially stabilizes the latch. Finally, Fig. 2(c) summarizes the effect of various parameters on the MTTE. As expected, higher $V_{DD}$ provides additional stability, higher $T$ increases the bit-flip rate, and in all cases threshold mismatch $\Delta V_{TH}$ makes subthreshold SRAM susceptible to thermal noise-driven bit-flip errors. Our simulation framework provides a tool to evaluate the trade-off between low-$V_{DD}$ operation and noise immunity as a function of $T$ and technological parameters.

The large number of simulations possible with our technique provides a full statistical model of error rates over the technologically important time scales. Figure 3 compares the distributions of the logarithm of TTE for matched transistors and those with asymmetric $\Delta V_{TH} = 10$ and 20 mV. As shown in the figure, the TTE to a bit flip follows a log-normal distribution. Any statistical process like that described by Eqn. (4), which realizes a multiplicative product of many positive independent random variables $\{(M_n + 1)\}$, is log-normal as we have also found in previous results [3]. The number of simulations is also sufficient to be a representative distribution from which to derive the cumulative distribution function (CDF) of the TTEs. Figure 4 compares the CDFs for perfectly matched transistors and those with asymmetric $\Delta V_{TH} = 10$ and 20 mV. Each curve represents 50,000 simulations. MTTE values (shown by arrows in Fig. 4) are dominated by the largest terms in the distribution, whereas reliability is affected by the smallest terms. For example, the $10^{th}$ percentile line of TTE in Fig. 4 shows that $10\%$ of all latches in a memory bank with matched transistors are likely to suffer a bit flip in $\sim 10^{26}$ seconds, but this time falls to $\sim 50$ days for 20 mV transistor mismatch. Common practice places tens to thousands of kilobytes of memory on a chip. Careful examination of the 20 mV CDF, shows that it is probable that any of those kilobytes with mismatched devices will have an erroneous bit within 90 seconds.

## IV. Conclusion

In this paper, we have described a new approach for analyzing thermal noise-driven transients in subthreshold SRAMs. We have also introduced a parallelizable iterative algorithm that allows us to run massive Monte-Carlo simulations to detect exponentially rare bit-flip errors. This would be unfeasible with conventional transient simulation methods. In addition, we can compute the CDF curves of TTE distributions to fully express the impact of error rates and their dependence on technological parameters. Our analysis of the TTE distributions shows that although the bit-flip errors in SRAMs made up of perfectly matched transistors are extremely rare, this reliability is significantly degraded by modest $V_{TH}$ mismatch. Our simulator provides the first statistically rigorous tool for evaluating ultimate SRAM reliability.

## References

[1] P. R. Gray, R. G. Meyer, P. J. Hurst, and S. H. Lewis, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[2] R. Sarpeshkar, T. Delbruck, and C. A. Mead, "White noise in MOS transistors and resistors," *IEEE Circ. Dev. Mag.*, vol. 9, pp. 23–29, 1993.

[3] M. Donato, R. I. Bahar, W. R. Patterson, and A. Zaslavsky, "A subthreshold noise transient simulator based on integrated random telegraph and thermal noise modeling," *IEEE TCAD*, vol. 37, pp. 643–656, 2018.

[4] P. Jannaty, F. C. Sabou, S. T. Le, M. Donato, R. I. Bahar, W. R. Patterson, J. Mundy, and A. Zaslavsky, "Shot-noise-induced failure in nanoscale flip-flops part II: Failure rates in 10-nm ultimate CMOS," *IEEE TED*, vol. 59, pp. 807–812, 2012.

[5] N. Miskov-Zivanov and D. Marculescu, "MARS-C: Modeling and reduction of soft errors in combinational circuits," in *Proceedings of the 43rd Annual Design Automation Conference*. ACM, 2006, pp. 767–772.

[6] L. T. Clark, V. Vashishtha, L. Shifren, A. Gujja, S. Sinha, B. Cline, C. Ramamurthy, and G. Yeric, "ASAP7: A 7-nm finFET predictive process design kit," *Microelectronics J.*, vol. 53, pp. 105–115, 2016.

[7] J. Han, E. Taylor, J. Gao, and J. Fortes, "Faults, error bounds and reliability of nanoelectronic circuits," in *2005 IEEE International Conference on Application-Specific Systems, Architecture Processors (ASAP'05)*. IEEE, 2005, pp. 247–253.

[8] M. Qazi, M. Sinangil, and A. Chandrakasan, "Challenges and directions for low-voltage SRAM," *IEEE Design & Test of Computers*, vol. 28, no. 1, pp. 32–43, 2010.