Stochastic Nonlinear Dynamical Modelling of SRAM Bitcells in Retention Mode

Léopold Van Brandt, Denis Flandre and Jean-Charles Delvenne
ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium, leopold.vanbrandt@uclouvain.be

2024 8th IEEE Electron Devices Technology & Manufacturing Conference (EDTM) | 979-8-3503-7152-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EDTM58488.2024.10512067

Abstract

SRAM bitcells in retention mode behave as autonomous stochastic nonlinear dynamical systems. From observation of variability-aware transient noise simulations, we provide an unidimensional model, fully characterizable by conventional deterministic SPICE simulations, insightfully explaining the mechanism of intrinsic noise-induced bit flips. The proposed model is exploited to, first, explain the reported inaccuracy of existing closed-form near-equilibrium formulas aimed at predicting the mean time to failure and, secondly, to propose a closer estimate attractive in terms of CPU time.
Keywords: Ultra-Low-Voltage SRAM, Noise-Induced Failures, SRAM Dynamic Stability, Stochastic Modelling.

Introduction

Integrated *Static Random Access Memory* (*SRAM*) arrays contains hundred of thousands bitcells, e.g 32 kB = 262 144 bits for an ultra low voltage (ULV) microcontroller [1], whose functionality is statistically endangered by the combined effects of *process variability* and *intrinsic noise* [2]. is twofold. On one hand, a small yet non-negligible fraction (e.g. that must be guaranteed below ~ 100 ppm [3]) of bitcells are defective as fabricated [4]. On the other hand, some surviving bitcells severely affected by variability have reduced noise immunity and are thereby prone to *dynamic instability* [5]. Although these are deemed functional at time zero, the intrinsic noise is likely to induce *transient failures* or *bit flips* in short times unacceptable for data retention in practical applications [5].

Simulating bit flips in SRAM bitcells in *retention* mode requires transient noise analyses [2], [5], [6], i.e. time-domain simulations however with independent random current noise sources added in parallel to each dissipative devices (resistors, MOS transistors,...). A robust methodology, compatible with industrial tools, was described in [5]. Whereas transient noise simulations are suitable for nonlinear circuits operating in out-of-equilibrium, large-signal conditions such as SRAM bitcells on the verge of instability, the CPU time quickly explodes with the number of cases (e.g. process variations) [2], [5].

Reference [6] developed an accelerated simulator aimed at efficiently estimating the *mean time to failure* ($MTTF$) in SRAMs though not straightforward to extend other memory architectures and technologies. We also find a few analytical attempts in the literature. In [7], the $MTTF$ is calculated from stochastic thermodynamic considerations, however assuming simplified transistor model and constant capacitances. We owe the only existing closed-form formula in the electronics literature to Kish [8], with preliminary attempts of application reported in [2], [9]. The previous work [2] has shown by numerical experiments that Kish's and the similar but more rigorous Nobile's formulas [10] lack accuracy. Both indeed rely on coarse near-equilibrium approximation around the presumed stable point of the bitcell in retention, as will be revisited in the present paper. Ignoring technological and application aspects (notably variability), SRAM dynamic stability has been investigated in the past within the mathematical framework of nonlinear dynamical systems [11], [12]. Paper [11] used simplified analytical inverter model and only considered deterministic pulse noise as disturbance, i.e. not the thermal noise of the transistors [2], [5], [6].

The purpose of the present work is to propose a stochastic and nonlinear dynamical model of an SRAM bitcell in retention, insightfully explaining the observed transient bit-flip mechanism. The parameters are fully extracted from conventional deterministic SPICE simulations, compatible with industrial transistor compact models and advanced technologies. While pointing out the limits of existing analytical formulas at the light of the presented model, we open promising avenues for fast and accurate semi-analytical approaches.

Modelling and Analysis

In Six-Transistor (6T) and other ULV SRAM bitcell architectures, data *retention* is ensured by a cross-coupled inverter pair (Figure 1(a)) implementing a feedback loop counteracting moderate disturbances [2]. We simulate the noise in the variability-aware setup proposed in [2]. Following [4], process variations are introduced as series-voltage sources δV_1 and δV_2 applied at the input of the inverters (Figure 1(a)). Their effect is incorporated in the *modified* voltage transfer characteristics (VTCs, in blue in Figure 1(c)), shifted from the nominal VTCs (in green). Without loss of generality, we focus the noise analyses on the special case $\delta V_1 = -\delta V_2$ corresponding to the worse-case scenario where both inverters are adversely affected [2].

The two output node voltages are denoted $v_1(t)$ and $v_2(t)$, respectively. If we adopt the convention $\delta V_1 = -\delta V_2 > 0$, the threatened memory state is $(v_2, v_1) = (X_0, Y_0)$ (see Figure 1(c)) due to the degraded noise margins.

A. Observation and Analysis of the Bit-Flip Mechanism

Figure 1(b) presents one typical simulation of intrinsic noise-induced bit flip [2]. The underlying mechanism is more easily understood with the *state space* representation [11], [12] of Figure 1(c), where the *state vectors* $(v_2(t), v_1(t))$ are plotted at various times to yield the *state trajectory*. The VTCs of the affected bitcell are also represented to locate the two *stable equilibrium points* (X_0, Y_0) and (X_1, Y_1), which slightly deviate from the nominal $(0, V_{DD})$ and $(V_{DD}, 0)$ due to process variations and to highlight the out-of-equilibrium behaviour of the SRAM during the transient bit flip.

The *unstable* equilibrium point (X_M, Y_M) is precisely located on the *stability boundary* (or *separatrix* [12]) separating the two stability regions. As long as $(v_2(t), v_1(t))$ lies before (X_M, Y_M), the natural dynamics of the SRAM bitcell tends to bring it to back to its initial logic state (X_0, Y_0) in absence of continuous disturbance. A state flip occurs when $(v_2(t), v_1(t))$ crosses (X_M, Y_M) due to simultaneous large voltage noise fluctuations (Figure 1(b)). Once $(v_2(t), v_1(t))$ has fallen in the region of attraction of the other equilibrium, (X_1, Y_1), the two cross-coupled inverters enter in positive feedback loop that quickly completes the bit flip. We define the random TTF as the time when $v_2(t)$ and $v_1(t)$ cross (Figure 1(b)).

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on December 12,2024 at 14:35:55 UTC from IEEE Xplore. Restrictions apply.

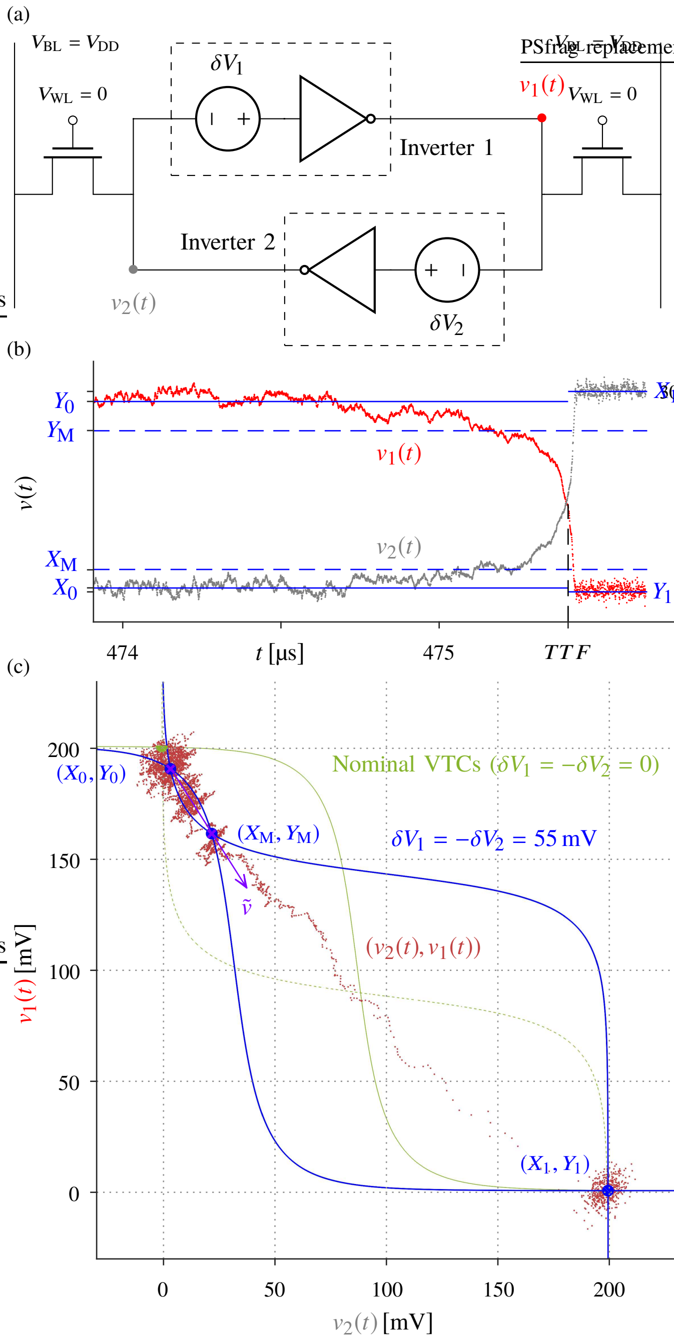

Figure 1. (a) SRAM bitcell in retention mode, with series-voltage sources modelling process variations. Illustrated below: $\delta V_1 = -\delta V_2 = 55\,\text{mV}$.
(b) Transient simulation of a noise-induced bit flip in the 6T SRAM bitcell.
(c) State trajectory of the bit flip of (b) in the state space. The nominal and modified VTCs of the inverters are shown in blue and green, respectively. The preferential reaction coordinate $\tilde{v}$ is shown in violet.
Illustrated case: 28 nm FD-SOI Single-P-Well (SPW) SRAM cell (inverters made of RVT nMOS and LVT pMOS; RVT nMOS access transistors) of minimal transistor dimensions $L_n = L_p = 30\,\text{nm}$ and $W_n = W_p = 80\,\text{nm}$, SPW bias $V_B = 0$, operating at $V_{DD} = 200\,\text{mV}$ and $T = 300\,\text{K}$.
Bandwidth of the generated noise: $f_{max} = 1\,\text{GHz}$ ($dt = 500\,\text{ps}$) [2].

B. Stochastic Nonlinear Dynamical Model

Upon observation of the experience in Figures 1(b) and 1(c) (other trajectories simulated for other cases showed similar behaviour), it seems reasonable to assume that a bit flip occurs according to the *preferential direction* given by the straight line connecting the two nearby points (X_0, Y_0) and (X_M, Y_M). This justifies the introduction of the unidimensional coordinate $\tilde{v}$ (shown in violet in Figure 1(c)) constructed by affine transformation (orthogonal projection) of (v_2, v_1):

$$\tilde{v} \equiv \frac{X_M - X_0}{\Delta\tilde{v}} \cdot v_2 + \frac{Y_M - Y_0}{\Delta\tilde{v}} \cdot v_1, \qquad (1)$$

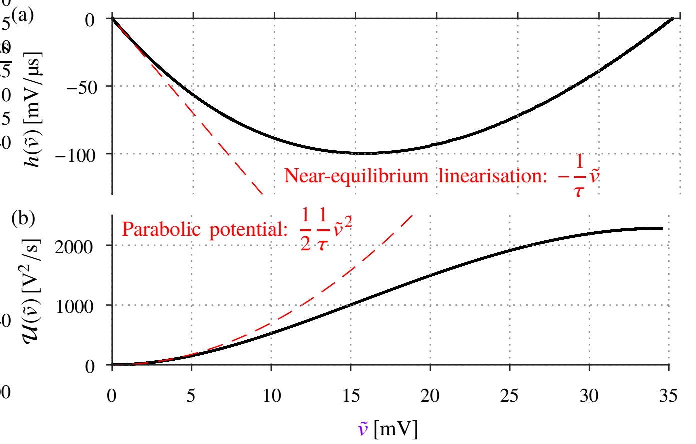

Figure 2. Extraction of the deterministic drift term in (2). Near-equilibrium approximations inherent to Kish's formula are shown in red. (a) $h(\tilde{v})$. (b) Quasi potential $\mathcal{U}(\tilde{v}) = -\int_0^{\tilde{v}} h(\tilde{v}')\,d\tilde{v}'$. Illustrated case: same as Figure 1.

where $Y_M - Y_0 < 0$ and $\Delta\tilde{v} \equiv \sqrt{(X_M - X_0)^2 + (Y_0 - Y_M)^2}$.

We assume a first-order dynamics:

$$d\tilde{v}/dt = h(\tilde{v}) + \sigma_w w(t), \qquad (2)$$

which may be thought as formally describing a *nonlinear RC* circuit, i.e. an overdamped single-state system. The function $h(\tilde{v})$ absorbs the nonlinearities of both the resistive and capacitive components of the SRAM. It can be cheaply and unambiguously extracted from a noiseless transient simulation: 1) starting from an initial condition $(v_2(0), v_1(0)) = (X_M - \epsilon, Y_M + \epsilon)$; the deterministic evolution of $(v_2(t), v_1(t))$ toward (X_0, Y_0) is recorded; 2) $\tilde{v}(t)$ is calculated through (1); 3) $d\tilde{v}/dt$ is computed numerically and mapped to $\tilde{v}(t)$ to provide $h(\tilde{v})$ between 0 and $\Delta\tilde{v}$ (depicted in Figure 2(a)).

In (2), $\sigma_w w(t)$ is the thermal noise term, σ_w^2 being the variance per unit time (i.e. in V^2/s) and $w(t)$ the white noise process of unit variance (in $s^{-1/2}$). Equation (2) may also be regarded as a nonlinear *drift* ($h(\tilde{v})$) - *diffusion* ($\sigma_w w(t)$) dynamics, considering that the drift term derives from a scalar *quasi potential* $\mathcal{U}(\tilde{v})$ (in V^2/s), $h(\tilde{v}) \equiv -d\mathcal{U}/d\tilde{v}$, as represented in Figure 2(b). The stable point $\tilde{v} = 0$ corresponds to the valley of $\mathcal{U}(\tilde{v})$, whereas the hill characterizes the unstable point at $\tilde{v} = \Delta\tilde{v}$.

Thereafter, we assume that σ_w^2 does *not* vary with $\tilde{v}$, and equals to its equilibrium value at $\tilde{v} = 0$. Only at the stable equilibrium point, the SRAM dynamics can be formally assimilated to a linear *RC* circuit and $\sigma_{\tilde{v}}^2$, the time-independent variance of $\tilde{v}(t)$, can be soundly defined and extracted by combining noise AC simulations with relation (1) (see [2, (2)]). For an *RC* circuit, we have $\tau = RC$, $\sigma_{\tilde{v}}^2 = kT/C$ and $\sigma_w^2 = 2kT/(RC^2)$ (Johnson-Nyquist formula). These combine into $\sigma_w^2 = 2\sigma_{\tilde{v}}^2/\tau$, where τ is the time constant of the linearised system (extraction illustrated in Figure 2(a)).

Discussion of Existing Analytical Predictions

Reference values for the *MTTF* were obtained by averaging 100 *TTF* extracted from SPICE transient noise simulations like Figure 1(b) [2], for several process variations $\delta V_1 = -\delta V_2 = 55$ to $58\,\text{mV}$ (Figure 3). The total CPU time reaches no less than a few *days* [2]. The model (2) is a stochastic differential equation for which trajectories can be efficiently simulated in a Monte-Carlo fashion by resorting to a dedicated numerical integration schema like the *Euler–Maruyama method* [13]. For each $\tilde{v}(t)$ sample paths,

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on December 12,2024 at 14:35:55 UTC from IEEE Xplore. Restrictions apply.

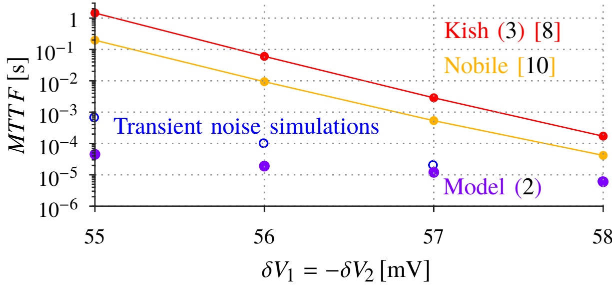

Figure 3. Comparison between the $MTTF$ estimated empirically from transient noise simulations (100 like Figure 1(b) for each δV case [2]), from the model (2), and predictions of analytical near-equilibrium formulas.

one TTF is obtained by recording the first time to reach $\Delta\tilde{v}$, multiplied by 2 since at the top of the hill (Figure 2(b)) the state can still flip left or right with probability $1/2$. The $MTTF$ thus estimated are shown in violet in Figure 3. Very good agreement can be observed, excepted for $\delta V = 55\,\text{mV}$. The distance $\Delta\tilde{v}$ is larger for this case, and the discrepancy is likely to be attributed to the constant-σ_w assumption which, yet convenient, is physically questionable. We can appreciate a CPU time reduced to barely a few *minutes*.

We now review existing closed-form formulas from the literature. Kish proposed a simplified Rice formula for the mean frequency of crossing ($1/MTTF$) a given threshold voltage ($\Delta\tilde{v}$) by a band-limited Gaussian white noise process ($\tilde{v}(t)$) [8, (9)]:

$$\frac{1}{MTTF} = \frac{2}{\sqrt{3}} \exp\left(-\frac{1}{2}\left(\frac{\Delta\tilde{v}}{\sigma_{\tilde{v}}}\right)^2\right) f_\text{p} \propto \exp\left(-\frac{\Delta\tilde{v}^2/(2\tau)}{\sigma_{\tilde{v}}^2/\tau}\right)\frac{1}{\tau}. \quad (3)$$

$f_\text{p} = 1/(2\pi\tau)$ is the $\tilde{v}$ noise bandwidth, defined at $\tilde{v} = 0$. Despite lack of detailed proof, we understand (3) as linearising the true SRAM dynamics around equilibrium to reduce it to an RC-like circuit with Gaussian voltage distribution. $\Delta\tilde{v}^2/(2\tau)$ is the value of the quasi potential barrier evaluated at the unstable point ($\tilde{v} = \Delta\tilde{v}$) under the harmonic (parabolic) potential approximation (Figure 2(b)). Equation (3) finally highlights a Maxwell-Boltzmann probability:
$1/MTTF \propto \exp(-\Delta E/kT)/\tau$, with $\Delta E = C\Delta\tilde{v}^2/2$.

A similar but more mathematically rigorous $MTTF$ formula can be obtained from Nobile's mean *first passage time* of an Ornstein-Uhlenbeck process [10, (6a) multiplied by 2][2, (5)], which is is defined by a linearised drift term $h(\tilde{v}) \approx -\tilde{v}/\tau$ (Figure 2(a)) [2]. Consequently, Nobile relies on exactly the *same* near-equilibrium approximation as Kish.

The above analytical predictions were added in Figure 3, respectively in red and orange, for comparison. We emphasize that both formula totally fail in accurately predicting the $MTTF$, the discrepancy increasing to more than two orders of magnitude for larger $MTTF$ (more moderate variability conditions). At the light of the model (2) we attribute the reported inaccuracy to the near-equilibrium approximation on which the formulas are based, inappropriate to capture the nonlinear SRAM dynamics (Figure 2). Through the parabolic approximation, Kish overestimates the quasi potential barrier that the process $\tilde{v}(t)$ must cross. For the $\delta V_1 = -\delta V_2 = 55\,\text{mV}$ case illustrated in Figure 2(b), we find that $\Delta\tilde{v}^2/(2\tau)$ exceeds the actual $\mathcal{U}(\Delta\tilde{v})$ by a factor close to 4. The Boltzmann probability $\propto \exp(-\Delta E/kT)$ is therefore strongly underestimated. Although this partially explains the significant overestimation of the $MTTF$, we must mention that the τ present in (3) is also a quantity defined and extracted at equilibrium, hence likely to be a source of discrepancy. Remedy these existing formulas would require more general out-of-equilibrium concepts that would be consistent with the stochastic nonlinear dynamics modelled by (2) and valid from $\tilde{v} = 0$ to $\tilde{v} = \Delta\tilde{v}$, like the Monte-Carlo trials proposed in this work. A subsequent promising approach will be to *calculate* the $MTTF$ piecewise, over small interval over which the dynamics can be *locally* approximated.

Conclusions

In ULV SRAM design, bitcells with lowered noise immunity due to process variations are prone to dynamic instability. Fast and accurate prediction of the time to noise-induced transient failures, compatible with industrial models and simulation tools, remains a challenge. We have proposed a stochastic nonlinear dynamical model of SRAM bitcells in retention mode. Existing closed-form formulas are based on a near-equilibrium approximation, interpreted either as a parabolic approximation of the quasi potential barrier or, equivalently, as a linearised drift term. We have shown that they struggle to accurately predicted the $MTTF$ extracted from computational intensive SPICE transient noise simulations. We expect from preliminary Monte-Carlo trials exploiting our model that a semi-analytical formulation accommodating the true SRAM dynamics and involving parameters that can be cheaply extracted from deterministic SPICE simulations is the most promising avenue for future work on that topic.

References

[1] D. Bol *et al.*, "SleepRunner: A 28-nm FDSOI ULP Cortex-M0 MCU With ULL SRAM and UFBR PVT Compensation for 2.6-3.6-μW/DMIPS 40-80-MHz Active Mode and 131-nW/kB Fully Retentive Deep-Sleep Mode," *IEEE J. of Solid-State Circuits*, 2021.

[2] L. Van Brandt *et al.*, "Variability-Aware Noise-Induced Dynamic Instability of Ultra-Low-Voltage SRAM Bitcells," in *IEEE LASCAS 2024, Uruguay*, 2024, submitted.

[3] R. G. Dreslinski *et al.*, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits," *Proc. of the IEEE*, vol. 98, no. 2, pp. 253–266, 2010.

[4] L. Van Brandt *et al.*, "Accurate and Insightful Closed-Form Prediction of Subthreshold SRAM Hold Failure Rate," *IEEE Trans. on Circuits and Systems I: Regular Papers*, pp. 1–14, 2022.

[5] ——, "On Noise-Induced Transient Bit Flips in Subthreshold SRAM," *Solid-State Electronics*, vol. 208, p. 108715, 2023.

[6] E. Rezaei *et al.*, "Fundamental Thermal Limits on Data Retention in Low-Voltage CMOS Latches and SRAM," *IEEE Trans. on Device and Materials Reliability*, vol. 20, no. 3, pp. 488–497, 2020.

[7] N. Freitas *et al.*, "Reliability and entropy production in nonequilibrium electronic memories," *Physical Review E*, vol. 105, no. 3, p. 034107, 2022.

[8] L. B. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics," *Physics Letters A*, vol. 305, no. 3-4, pp. 144–149, 2002.

[9] F. Veirano *et al.*, "Minimum Operating Voltage Due to Intrinsic Noise in Subthreshold Digital Logic in Nanoscale CMOS," *J. of Low Power Electronics*, vol. 12, no. 1, pp. 74–81, 2016.

[10] A. Nobile *et al.*, "Exponential trends of Ornstein–Uhlenbeck first-passage-time densities," *J. of Applied Probability*, vol. 22, no. 2, pp. 360–369, 1985.

[11] B. Zhang *et al.*, "Analytical Modeling of SRAM Dynamic Stability," in *Proceedings of the 2006 IEEE/ACM International Conf. on Computer-Aided design*, 2006, pp. 315–322.

[12] W. Dong *et al.*, "SRAM Dynamic Stability: Theory, Variability and Analysis," in *2008 IEEE/ACM International Conf. on Computer-Aided Design*. IEEE, 2008, pp. 378–385.

[13] D. J. Higham, "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations," *SIAM review*, vol. 43, no. 3, pp. 525–546, 2001.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on December 12,2024 at 14:35:55 UTC from IEEE Xplore. Restrictions apply.