Modeling and Predicting Noise-Induced Failure Rates in Ultra-Low-Voltage SRAM Bitcells Affected by Process Variations

Léopold Van Brandt, Michele Bonnin, Mauricio Banaszeski da Silva, *Member, IEEE*, Pascal Bolcato, Gilson I. Wirth, *Senior Member, IEEE*, Denis Flandre, *Senior Member, IEEE*, and Jean-Charles Delvenne

***Abstract*— Stability of ultra-low-voltage SRAM bitcells in retention mode is threatened by two types of uncertainty: process variability and intrinsic noise. While variability dominates the failure probability, noise-induced bit flips in weakened bitcells lead to dynamic instability. We study both effects jointly in a unified SPICE simulation framework. Starting from a synthetic representation of process variations introduced in a previous work, we identify the cases of poor noise immunity that require thorough noise analyses. Relying on a rigorous and systematic methodology, we simulate them in the time domain so as to emulate a true data retention operation. Short times to failure, unacceptable for a practical ultra-low-power memory system application, are recorded. The transient bit-flip mechanism is analyzed and a dynamic failure criterion involving the unstable steady state is established. We conclude that, beyond static variability, the dynamic noise inflates defectiveness among SRAM bitcells. Then, a stochastic nonlinear model, fully characterizable from conventional deterministic SPICE simulations, is presented. We then leverage it to efficiently and accurately predict the mean time to failure with an analytical Eyring-Kramers formula, recently extended to account for the varying-noise behavior of nonlinear systems.**

***Index Terms*— Ultra-low-voltage SRAM, noise-induced failures, SRAM dynamic stability, stochastic modeling.**

I. Introduction

THE need for *ultra-low-power* (ULP) circuits and systems is notably motivated by the massive deployment of connected autonomous IoT nodes [1], translating into *ultra-low voltage* (ULV) design [2]. Processors operating at a *supply voltage* (V_{DD}) lowered below 200 mV are demonstrated [3]. *Static Random Access Memory* (*SRAM*) arrays are essential blocks of ULP systems [2], typically ranging from a few kB [2], [3] to 32 kB = 262 144 bits [4]. The functionality of these bitcells must be statistically guaranteed and thereby predicted.

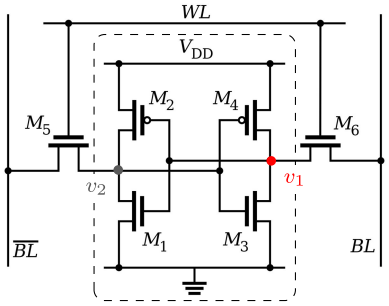

Fig. 1. 6T SRAM bitcell. Data retention is ensured by the cross-coupled inverter pair (M_1 and M_2, M_3 and M_4), like in 8T and 10T architectures.

Whereas the MOS transistors of smallest gate length offer higher-density SRAM arrays and faster read/write operations, they are also more sensitive to uncertainties like *process variability* and *intrinsic noise*. The effect of process variations is twofold. On one hand, a small yet non-negligible fraction (e.g. that must be guaranteed below $\sim$ 100 ppm for the above mentioned SRAM array sizes [5]) of bitcells are defective as fabricated [6]. On the other hand, some surviving bitcells severely affected by variability have reduced noise immunity and are thereby prone to *dynamic instability* [7], [8], [9]. Although these are deemed functional at time zero, the intrinsic noise is likely to induce *failures* during retention or *bit flips* in short times unacceptable for data retention in practical applications [7], [8], [9].

Generally, robustness of SRAM bitcells against all read/write/hold failures due to the combined effects of process variations and noise is a major concern for ULV design [2], [10]. To alleviate the tradeoff and to overcome the limitations of the Six-Transistor (6T) SRAM bitcell (sketched at the transistor level in Figure 1) dedicated ULV architectures like the 8T [11] and the 10T [12], [13], [14], [15] are provided with a read buffer. In this configuration, the hold mode becomes the critical one. For all these bitcells, data *retention* is ensured by a cross-coupled inverter pair (*latch*, dotted box in Figure 1)

Received 23 June 2024; revised 24 October 2024; accepted 17 December 2024. Date of publication 10 January 2025; date of current version 27 February 2025. This work was supported in part by the Research Project "Thermodynamics of Circuits for Computation" of the National Fund for Scientific Research (FNRS) of Belgium. This article was recommended by Associate Editor L. Trojman. *(Corresponding author: Léopold Van Brandt.)*

Léopold Van Brandt, Denis Flandre, and Jean-Charles Delvenne are with the Institute for Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), Université catholique de Louvain (UCLouvain), 1340 Ottignies-Louvain-la-Neuve, Belgium (e-mail: leopold.vanbrandt@uclouvain.be).

Michele Bonnin is with the Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy.

Mauricio Banaszeski da Silva is with the Department of Electronics and Computing, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil.

Pascal Bolcato is with Siemens DISW-ICS-CICV, 38240 Meylan, France.

Gilson I. Wirth is with the Electrical Engineering Department, Universidade Federal do Rio Grande do Sul, 90040-060 Porto Alegre, Brazil.

Digital Object Identifier 10.1109/TCSI.2024.3525387

1549-8328 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

[14, Fig. 1], which implements a feedback loop counteracting moderate disturbances, thereby ensuring a certain stability and noise immunity.

One can find a many ULP applications, especially biomedical [16], [17], for which the activity factor of the SRAM is low so that the typical retention time is long. This makes the retention mode all the more critical in terms of energy minimisation and dynamic stability. From a power-aware design point of view, the key issue is to further reduce V_{DD} in retention (lower than for the read/write operations [16], [17]), while maintaining the cell failure probability below an acceptable value.

Simulating failures in retention, i.e. bit flips induced by the intrinsic noise of the transistors, requires computational intensive transient noise analyses [7], [8], [9], [18]. Those are conventional 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 [7]. Whereas transient noise simulations are suitable and insightful (as emulating real-world conditions) 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) [7], [8], [9]. The brute-force approach that couples Monte-Carlo simulations with transient noise analyses is therefore prohibitively expensive. This has motivated former attempts of faster prediction methods. The authors of [18] and [19] developed an accelerated simulator aimed at efficiently estimating the *mean time to failure* ($MTTF$) in CMOS latches though not straightforward to extend to other memory architectures and technologies. We also find a few analytical attempts in the literature. The framework proposed in [20] and [21], based on Markov chains and queue theory, suffers from the same limitations as [18] and [19] and lacks validation with SPICE transient noise simulations. In [22], 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 [23], with preliminary attempts of application reported in [8] and [24]. The previous works [8], [9], [25] have nevertheless shown by numerical experiments that Kish's and the similar but more rigorous Nobile's formulas [26] lack accuracy. Both indeed rely on a coarse near-steady-state approximation around the presumed stable point of the bitcell in retention, which is highly inappropriate in describing the stochastic nonlinear dynamics of the SRAM [25] as will be further revisited in the present paper.

Very recently, an extended Eyring-Kramers formula has been derived from stochastic calculus [25]. Relying on a unidimensional stochastic model fully characterizable with SPICE simulations [9], it does account for both the nonlinear SRAM dynamics and the varying noise intensity along the bit-flip trajectory. Remarkable in terms of accuracy, the formula aims at predicting the mean transition time in a wide and general class of autonomous nonlinear bistable systems, in particular the $MTTF$ in SRAM bitcell in retention. Combined with the process variability-aware framework introduced in [8], we believe that such a semi-analytical approach is a promising avenue for efficient SPICE-compliant reliability assessment.

Extending the proceedings [8], [9], we thus propose a methodology for fast and accurate prediction of the $MTTF$ in ULV SRAM bitcells in retention, in presence of both process variations and intrinsic noise. The rest of this paper is structured as follows. Section II present the process variability-aware simulation framework of [8] in order to insightfully identify the worst-cases of endangered functionality. Transient noise simulations is then the purpose of Section III. Starting from the observation, the bit-flip mechanism is extensively explained in Section III-A. The choice of the simulation parameters and the related CPU time are addressed in Section III-B. In Section III-C, we analyse and model the TTF distribution to show that, in agreement with the general theory of Markovian bistable system, the $MTTF$ metric alone fully parametrises it and hence is relevant in quantifying the SRAM dynamic stability. Stochastic modeling is extensively developed in Section IV, first generally (Section IV-A) and then formally reduced to unidimensional dynamics (Section IV-B) as conjectured in [9]. We place particular emphasis on the characterisation from conventional deterministic SPICE simulations, compatible with industrial transistor compact models and advanced technologies. Subsequently, approximate closed-form formulas to predict the $MTTF$, valid when the thermal noise is dominant, are summarised in Section V. Their accuracies are discussed in Section VI, dedicated to the results. Section VII finally draws the conclusions and opens perspectives regarding the future incorporation of the low-frequency noise, in addition to process variations and thermal noise.

II. Process Variability-Aware Simulation Framework

The voltage limit in ULV circuits is mainly dictated by process variations [2, Fig. 17]. As extensively reviewed in [6], enhanced Monte Carlo methods rely on the most recent advances in importance sampling to speed up the simulations whose aim is to empirically estimate the SRAM failure probability. The intrinsic noise of the transistors (including the access transistors M_5 and M_6 shown Figure 1) and of peripheral circuits comes as an additional uncertainty, taking on the design margins. The same goes for the supply-voltage (droop) noise. The V_{DD} referred to as below may therefore be thought as the minimal supply voltage, reduced compared to its nominal value.

In [6], a novel non-Monte-Carlo semi-analytical methodology was introduced the detect the hold failures caused by static variability within ULV SRAM bitcells. It was shown that the dominant effect of process variations affecting the MOS transistors can be suitably and accurately modelled by two series-voltage sources δV_1 and δV_2, each applied at the input of one inverter of the latch [6, Figure 4]. The noise margin of a CMOS inverter operating in subthreshold is indeed dominantly degraded by the imbalance between nMOS and pMOS transistors [2], [27]. These δV may notably be related to the individual V_{th} shifts and the same goes for their statistics [6]. This is why we understand these sources of variations

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

as "intrinsic" to the physical inverters: for instance, "Inverter 1" refers to a nominal inverter affected by δV_1 (see dashed box in Figure 2(a)) throughout the text. Their effect is incorporated in the *modified* voltage transfer characteristics (VTCs), shifted from the nominal VTCs (in green in Figure 2(c)). The double DC sweep of the variations $(\delta V_1, \delta V_2)$ yields the two-dimensional (2D) representation of Figure 2(b). To each $(\delta V_1, \delta V_2)$ point is thoughtfully associated an SRAM bitcell, whose functionality has been assessed with the traditional "butterfly plot" [28] (Figure 2(c)). A tested bitcell is functional (green and orange), at time zero, if the number of cross points is exactly equal to three; defective (red) otherwise. The (positive) *static noise margin* (SNM) of the functional bitcells can be extracted with the SPICE-compliant method from List and Seevinck [28]. The procedure is graphically illustrated for three special cases of variations in Figure 2(c): nominal bitcell ($SNM = 61$ mV, comfortably stable), $\delta V_1 = -\delta V_2 = 55$ mV ($SNM = 5$ mV, barely functional), 65 mV (defective).

Whereas any SRAM bitcell exhibiting a non-negative noise margin ($SNM \geq 0$) would be classified as functional based on purely static considerations, we expect those with lowest SNM to have poor noise immunity and to be dynamically *unstable*. Rigorously, the SNM only quantifies the robustness of a bitcell against DC sources of variations, in a particular scenario where both inverters are adversely affected [6] (dashed line $\delta V_1 = -\delta V_2$). The dynamic noise margin, i.e. the robustness of the bitcell against transient noise, is substantially larger than the SNM as a consequence of the non-zero response time of the inverter gates (low-pass filtering effect) [7], [29]. The SNM nevertheless remains an indicative metric of low noise immunity and we can identify cells to be treated in priority for noise analysis.

In Figure 2(b), we have highlighted in orange the region corresponding to bitcells whose SNM lies between 0 (verge of instability, corresponding to the failure boundary shown in black dots) and 10 mV. Crucially, short TTF [18], [24] are likely observed for the bitcells in the orange crown, possibly shorter than the typical retention time of the memory application and hence unacceptable. The intrinsic noise therefore inflates the cell failure probability, in addition to variability, and one wants to assess its contribution through the probability

$$P(TTF < \text{retention time} \mid SNM \geq 0). \tag{1}$$

whereas the inner boundary of the orange crown of Figure 2(b) is firstly guessed by $SNM \leq 10$ mV, an iterative procedure would ultimately compare the obtained TTF to the retention time to enlarge the statistically relevant area. The whole area should be covered for a sound calculation of (1) in a longer-term reliability perspective. The naive approach would rely on repeated trials of transient noise simulations, in a Monte-Carlo fashion. As will be evidenced in the next Section III, this route is totally unaffordable without accelerated framework.

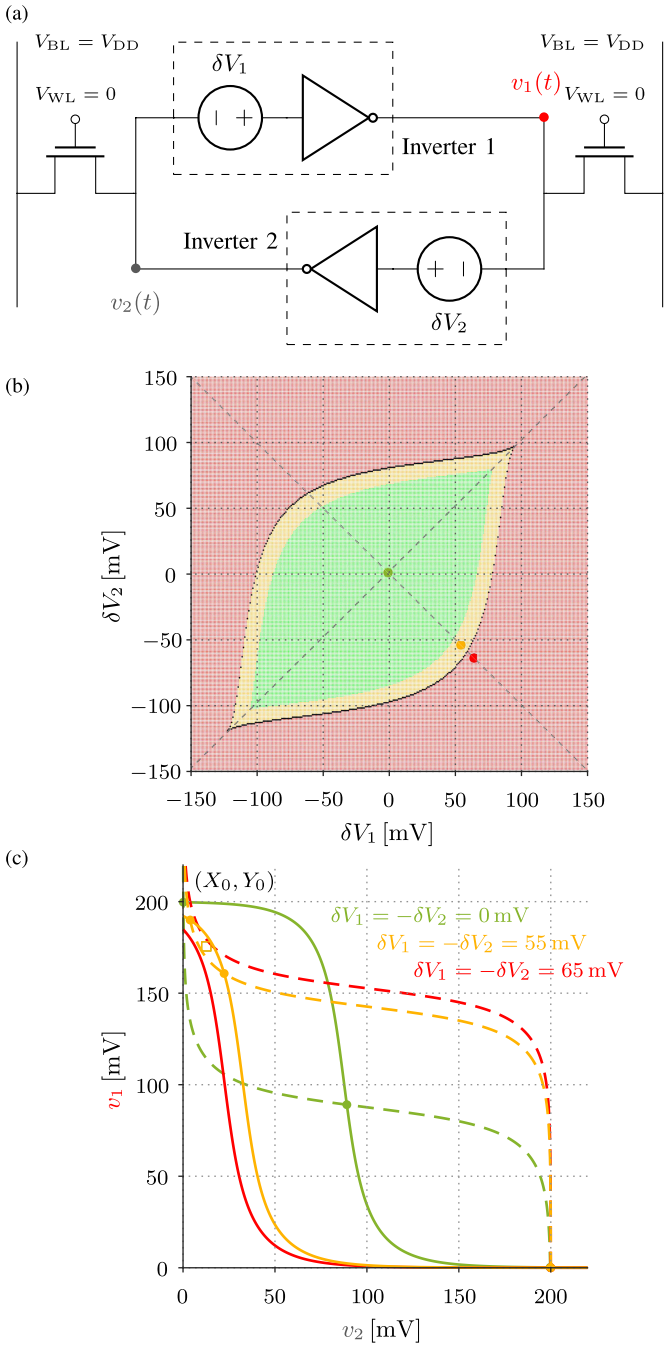

Fig. 2. (a) SRAM bitcell in retention mode, with process variations modelled as series-voltage sources applied at the inputs of the two inverters (M_1 and M_2, M_3 and M_4 in Figure 1). (b) 2D representation of functional (green and orange) and defective (red) SRAM bitcells in presence of process variability, deterministically simulated with a double DC sweep of variations $(\delta V_1, \delta V_2)$. The orange crown contains the bitcells of positive but low SNM (≤ 10 mV). Voltage step of the double DC sweep : $\Delta\delta V = 1$ mV. (c) Butterfly plots of three special cases marked by dots in (b), along the line $\delta V_1 = -\delta V_2$ corresponding to the worse-case scenario where both inverters are adversely affected. For functional bitcells, the SNM is the width of the largest inscribed square. 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 $L_n = L_p = 30$ nm and $W_n = W_p = 80$ nm, SPW bias $V_B = 0$, operating at $V_{DD} = 200$ mV and $T = 300$ K.

III. Transient Noise Simulations

Simulating the noise in the time domain requires to: (i) add individual physical noise sources in parallel with each device (e.g. transistor); (ii) generate random values at each time step; (iii) perform conventional transient simulations [7].

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

While insightful and accurate as emulating real-world large-signal conditions for arbitrary nonlinear circuits [30], transient noise analyses are rarely encountered in the literature. The first reason is, undoubtedly, the high computation time, difficult to fit in a design schedule; the second is, possibly, a lack of understanding of the specific simulation parameters. These points will be addressed in Section III-B.

The concern is further reinforced by an additional challenge: the incorporation of process variability in noise analyses, or conversely. The rarity of the failure events makes the brute-force approach coupling Monte-Carlo simulations with transient noise simulations even more prohibitively expensive. Previous work [7] mainly focused on symmetrical latches (neglecting access transistors M_5 and M_6 in Figure 1) operating at extremely low V_{DD}, however in the somewhat artificial situation of zero process variations. Provided with home-made accelerated noise simulation tools, the authors of [18], [20], and [21] introduced variability as a global deterministic imbalance between nMOS and pMOS transistors (asymmetrical process corner), like [24]). Instead, we rely on the framework presented in Section II, according to which we leverage on Figure 2(b) to select the weakest SRAM bitcells prone to dynamic instability.

Because of the CPU-time constraint, we have focused the transient noise analyses on a few limit cases belonging to the worst-case line $\delta V_1 = -\delta V_2$ (dashed line in Figure 2(b)); the case 55 mV presented earlier in orange in Figure 2(c) is one of them. From statistical considerations involving a two-dimensional Gaussian variability distribution [6], we can show that such selected points lie within a 10 ppm-equiprobability circle [6], i.e. frequently encountered among Monte-Carlo samples or fabricated bitcells. The choice $\delta V_1 = -\delta V_2$ does not affect the generality of the presented methodology and subsequent analyses. We emphasis that, whereas we first *observe* the bit-flip mechanism in the time domain, the expensive transient noise simulations are *not* part of the proposed predictive methodology.

The two output node voltages are denoted $v_1(t)$ and $v_2(t)$ (Figure 2(a)), respectively. The role of the two cross-coupled inverters of Figure 2(a) is totally interchangeable. Having adopted the convention $\delta V_1 = -\delta V_2 > 0$, the endangered memory state is $(v_2, v_1) = (X_0, Y_0)$ (see Figure 2). The exact high and low logic levels X_0 and Y_0 depend on the process variations affecting the particular bitcell (see again Figure 2). We have used the right initial data $(v_2(0), v_1(0)) = (X_0, Y_0)$ as the initial condition for all the transient experiments of the retention operation (i.e. not the approximate $(0, V_{DD})$ as [8], which resulted in a short settling time of about 1 μs to reach the actual steady state).

A. Observation and Analysis of the Bit-Flip Mechanism

One typical transient simulation of a bit-flip caused by intrinsic transistor noise is shown in Figure 3(a) [8]. The bit-flip mechanism in SRAM bitcells is better understood within the mathematical formalism of nonlinear dynamical systems [31], [32]. We call *state vectors* the pairs of voltages $(v_2(t), v_1(t))$. The set of all the possible values of those

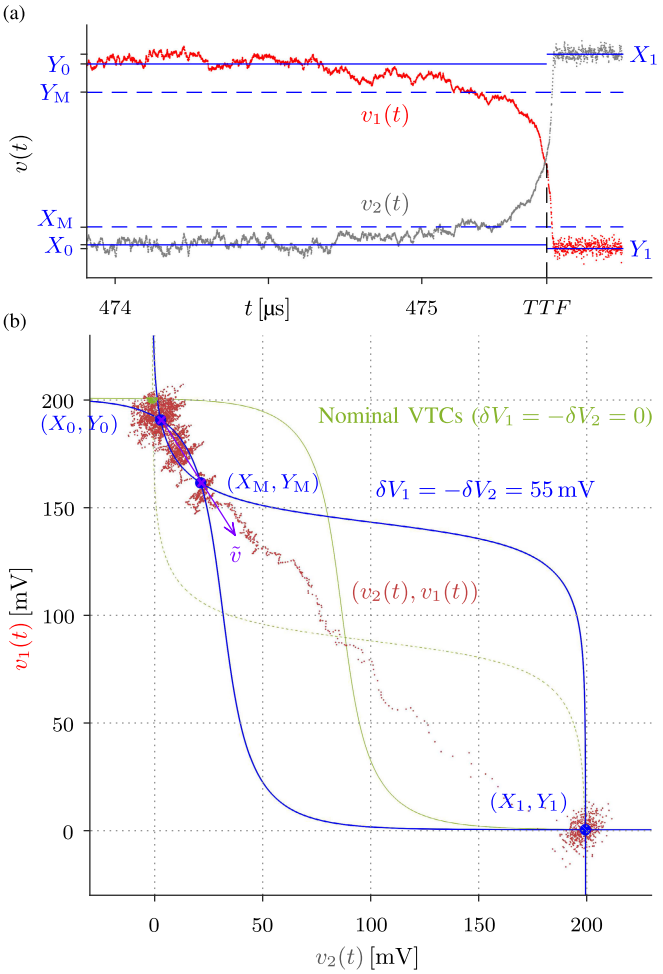

Fig. 3. (a) Transient simulation of a noise-induced bit flip (failure) in the 6T SRAM bitcell in retention (Figure 2(a)). Adapted from [9]. (b) State trajectory of the bit flip of (a) in the state space. The nominal and modified VTCs of the inverters are shown in blue and green, respectively. The unidimensional reaction coordinate $\tilde{v}$ is shown in violet. Illustrated case: same SRAM design as Figure 2, with process variations $\delta V_1 = -\delta V_2 = 55$ mV [9]. Bandwidth of the generated noise: $f_{max} = 1$ GHz ($\Delta t = 500$ ps) [8].

vectors forms the *state space* [31]. The *state trajectory*, obtained by plotting the state vectors at various times in the state space, is depicted in Figure 3(b). The butterfly of the affected SRAM bitcell is also represented in the state space in order to locate the two stable steady states or points (X_0, Y_0) and (X_1, Y_1), which slightly deviate from the nominal and ideal $(0, V_{DD})$ and $(V_{DD}, 0)$ due to process variations δV_1 and δV_2, and to emphasize the out-of-equilibrium behavior of the system during the transient bit flip. To each stable state is associated a *stability region* or *region of attraction* [32] The *stability boundary*, or *separatrix* [32], which separates the two stability regions, necessarily includes the saddle (*unstable*) point (X_M, Y_M).

At the beginning of the represented time segment, the two node voltages $v_2(t)$ and $v_1(t)$ (defined in Figure 2(a)) fluctuate quietly around the logic levels X_0 and Y_0, the data initially retained by the SRAM bitcell. Starting from about 474.5 μs, $v_2(t)$ gradually increases and $v_1(t)$ decreases due to hazardous and simultaneous large voltage noise fluctuations. This process goes against the deterministic regenerative property of the

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

inverters, which in absence of continuous disturbance would restore the logic levels X_0 and Y_0. Once $v_2(t)$ and $v_1(t)$ have crossed specific thresholds, respectively X_M and Y_M (Figure 3(a)), i.e. $(v_2(t), v_1(t))$ has gone beyond the unstable steady state (X_M, Y_M) and thereby has crossed the separatrix (Figure 3(b)), the two cross-coupled inverters enter in positive feedback loop. $(v_2(t), v_1(t))$ falls in the region of attraction of the other stable steady state, (X_1, Y_1), the bit flip becomes highly likely and rapid as dictated by the natural dynamics of the SRAM bitcell. The positive feedback makes the node voltages vary almost exponentially [31] as we can observe in Figure 3(a). This mode has sometimes been referred to as "strong feedback mode" [31].

Formally, the TTF is the time taken by the state to cross the separatrix for the last time, when the transition from (X_0, Y_0) to (X_1, Y_1) is completed. Experimentally, we define the completion of the state flip and we record the TTF when $v_2(t)$ and $v_1(t)$ cross, corresponding to the logical switch $0 \rightarrow 1$; it can be noticed in Figure 3(a) that, at this stage, crossing backward becomes unlikely due to the asymmetry of the bitcell. Because of the fast exponential evolution of the voltages beyond (X_M, Y_M), the two definitions are consistent. This TTF is a random variable for a given bitcell, which takes a different value for each experiment carried out.

Although determining the exact shape of the full separatrix (as addressed in [32]) is not required in this work, it is important to understand that (X_M, Y_M) is the threshold point. If we assume, after observation of Figure 3 (other trajectories simulated for other cases behaved similarly), that a bit flip occurs according to the preferential direction given by the line connecting the two nearby points (X_0, Y_0) and (X_M, Y_M), *the necessary failure criterion is that* $v_2(t)$ *and* $v_1(t)$ *cross the thresholds* X_M *and* Y_M, respectively. The largest the individual distances $\Delta X \equiv X_M - X_0$ and $\Delta Y \equiv Y_1 - Y_M$, the statistically rarest the bit-flip event (at fixed noise intensity) and the most robust the SRAM bitcell is. The deleterious effect of process variability is to reduce the noise margins (like the SNM) and, similarly, ΔX and ΔY. This leads to short TTF. The relevance of this hypothesis of unidimensional dynamics in estimating the TTF, here phenomenologically observed, will be further formalized and quantitatively justified in Section IV.

B. Simulation Parameters and CPU Time

Careful setting of the transient simulations parameters like the generated noise bandwidth (f_{max}), time step (Δt), and duration optimizes the CPU-time tradeoff while ensuring accuracy [7]. Preliminary to transient simulations, we have extracted, for each $(\delta V_1, \delta V_2)$ case studies, the noise bandwidth f_p of the inverters in the closed-loop configuration of Figure 2(a) (unlike [7] which suggested an open-loop procedure). Such `NOISE(AC)` simulations and their importance in determining the noise intensities will be described in Section IV. We make several important comments at this stage. First of all, the two inverters may have significantly different bandwidths (typically by a factor ≈ 2), because the input of one is low (X_0, close to 0) while the input of the other is high (Y_0, close to V_{DD}). Secondly, the bandwidth decreases from about 100 MHz for the bitcell free of any process variation down to a only few MHz for the largest $\delta V_1 = -\delta V_2$.

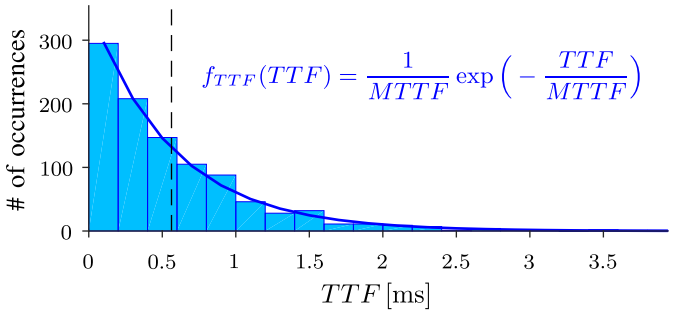

Fig. 4. Empirical histogram of the TTF, constructed from 1000 TTF realisations of transient noise simulations like Figure 3. The $MTTF \approx 0.56$ ms is indicated by the vertical dashed line. Solid line represents the theoretical exponential distribution (2) with the empirical sample mean plugged into it. Illustrated case: same as Figure 3. The bin width of 0.2 ms was chosen for illustration but is not directly related to any simulation parameter.

In order to take into account all contributing noise components, we usually use as generated noise bandwidth $f_{max} \approx 10 \cdot f_p$ [7]. The $f_{max} = 1$ GHz used in [8] is thus an excessively loose choice, computationally inefficient. To better optimise the CPU-time tradeoff [7], we here use $f_{max} = 50$ MHz. The maximum time step of the transient simulation (Δt) is related to f_{max} through $\Delta t = 1/(2\, f_{max})$ [7].

Setting the appropriate duration of the transient simulation is difficult for a highly asymmetrical bitcell, because the typically measured TTF are not known and hardly guessed *a priori*. As shown in previous work [7] and as will be illustrated in Section VI, the $MTTF$ metric evolves across several orders of magnitude with SNM variations of a few mV. With the aim of being able to observe longer TTF, corresponding to the right tail of the TTF distribution [18] (also studied in the next Section III-C), we ideally want the duration to exceed at least $10 \cdot MTTF$, approximately. This may require an iteratively refined simulation plan. In practice and since we focus on the shortest $MTTF$ cases, we limit and set the duration of all the transient experiments to 5 ms. Finally, let us mention that, for each $(\delta V_1, \delta V_2)$, we must perform 1000 repeated experiments and record a bit flip occurring at a random TTF for each of them, to guarantee the statistical accuracy of the empirical $MTTF$ estimator (relative standard error of the mean is 3 %).

Let us mention that, despite this finely optimised set of parameters, we still end up with a huge CPU time of a few hours per single bit-flip experiments and of three days to go through four selected $\delta V_1 = -\delta V_2$ variability cases. This, despite the use of a high-performance work station and parallel multi-core computing. Though, the whole orange crown of Figure 2(b) should be covered by massive transient noise simulations in order to estimate all the $MTTF$ and to evaluate (1). Without accelerator or semi-analytical framework, this would take several weeks or months of computation, prohibiting the naive brute-force approach.

C. Distribution of the Time to Failure

Using the 1000 recorded realisations for each variability case, we can empirically construct the probability density

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

function (PDF) of the random variable TTF as an histogram (depicted in Figure 4). In agreement with the theory of Markovian bistable system [33], the distribution is found to be exponential:

$$f_{TTF}(TTF) = \frac{1}{MTTF} \exp\Big(-\frac{TTF}{MTTF}\Big), \tag{2}$$

like [19, Fig. 11]. The authors of [18, Fig. 6] found, instead, a log-normal distribution. This could be an artefact of their iterative threshold-crossing algorithm, according to [19].

Importantly, the theoretical PDF (2) is entirely characterized by one single parameter: the mean, i.e. the $MTTF$ whose estimation is the purpose of the present paper. Within the usual notations in probability theory, $\lambda = 1/MTTF$ [33]. In particular, the standard deviation of the TTF is nothing but the $MTTF$ too. Thus, the $MTTF$ alone contains all the information about the statistics of the noise-induced failures, which are rare events. For a given $MTTF$ (e.g. the process variations are known and fixed), the probability (1) can be explicitly evaluated from (2):

$$P(TTF < \text{retention time}) = 1 - \exp\Big(-\frac{\text{retention time}}{MTTF}\Big). \tag{3}$$

Equation (3) clearly highlights the relevance of the $MTTF$ metric, compared to the typical retention time of the SRAM. The left tail of the TTF distribution corresponds to the shortest $MTTF$, unacceptable when smaller than the retention time, as (3) quantifies in probability.

This said, we emphasize that the demanding part of the reliability analysis is the estimation of the $MTTF$, for some *given* process variations ($(\delta V_1, \delta V_2)$ selected from Figure 2(b)), to be eventually plugged into (2) or (3). Subtly, $MTTF$ varies from one bitcell to another due to process variations and hence should be itself thought as a random variable within a sample space made of a large number of bitcells.

IV. Stochastic Nonlinear Dynamical Modeling

An SRAM bitcell in retention is undriven, i.e. isolated from peripheral circuitry (except for the leakage currents through the access transistors, which are taken into account in our study). It can therefore, generally, be assimilated to an autonomous stochastic nonlinear dynamical system, whose modeling is the purpose of this section.

A. Two-Dimensional Stochastic State-Space Model

1) Nodal Analysis of the SRAM Bitcell: We reasonably assume that the state of the system is fully determined by the node voltages $v_1(t)$ and $v_2(t)$ (as indicated in Figure 2(a)). The set of coupled nonlinear differential equations governing the dynamics, i.e. the time evolution of $v_1(t)$ and $v_2(t)$, results from the application of Kirchhoff current law at both nodes:

$$\begin{pmatrix} C_{22}(v_2, v_1) & C_{21}(v_2, v_1) \\ C_{12}(v_2, v_1) & C_{11}(v_2, v_1) \end{pmatrix} \begin{pmatrix} \mathrm{d}v_2/\mathrm{d}t \\ \mathrm{d}v_1/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} I_2(v_2, v_1) \\ I_1(v_2, v_1) \end{pmatrix} + \begin{pmatrix} i_{\mathrm{n},2}(t) \\ i_{\mathrm{n},1}(t) \end{pmatrix}. \tag{4}$$

The stochastic differential algebraic equation (4) is understood in Itô' sense, consistently with [25]. The importance and the relevance of this convention is discussed in more detail below.

The general model (4) can be obtained from a modified nodal analysis of the circuit of Figure 2(a), taking into account the intrinsic capacitances and the random noise sources of the MOS transistors. $I_2(v_2, v_1)$ denotes the static component of the current flowing out of Inverter 2, a nonlinear function of v_2 and v_1 through the v_{GS} and v_{DS} of the MOS transistors. It also incorporates the contribution of the leakage of the access transistor.

All the dynamic currents of the transistors have been regrouped in the form $C_{ij}\,\mathrm{d}v_j/\mathrm{d}t$, as included in the left side of (4). For instance, C_{22} may be thought as the sum of the output capacitance of Inverter 2 and the input capacitance of Inverter 1 referred to the ground. The capacitive coupling between input and output of each inverter comes from the gate-to-drain capacitances of the transistors, embedded in C_{12} and C_{21}. Most generally, all the intrinsic capacitances depend on the instantaneous voltages and are nonlinear. Extracting them from SPICE simulations is neither trivial nor practical, nor even useful for our purposes as will be explained below. It is, however, important to recognise the role of the capacitive currents in determining SRAM bitcell dynamics by relating the voltage variations to currents.

The noise is intrinsic to all the transistors (including the access). In (4), all the individual noise contributions are collected into single parallel-current sources $i_{\mathrm{n},2}(t)$ and $i_{\mathrm{n},1}(t)$. The white thermal noise intensities (i.e. the variances of $i_{\mathrm{n},2}(t)$ and $i_{\mathrm{n},1}(t)$) do vary with (v_2, v_1) in general, because the MOS transistors are nonlinear dissipative devices [34].

In former work [31, eq. (2)-(3)], single constant output node capacitance was assumed: $C_{22} = C_{11} = C =$ constant, together with $C_{21} = C_{12} = 0$. Subsequently simplified and inaccurate analytical transistor and inverter models were used. Moreover, only deterministic pulse noise were considered as disturbance $i_n(t)$ in [31], inappropriate in modeling the thermal noise. The accelerated thermal noise simulator of [18] and [19] assume constant C_{ij}. In the present work, we instead rely on a general reformulation of (4) as a *stochastic state-space representation*, characterizable without approximation from conventional SPICE simulations.

2) State-Space Representation: Equation (4) can be compactly rewritten in matrix form:

$$\boldsymbol{C}(\boldsymbol{v})\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{I}(\boldsymbol{v}) + \boldsymbol{i}_{\mathbf{n}}(t) \tag{5}$$

denoting $\boldsymbol{v} = (v_2, v_1)^{\mathsf{T}}$. The matrix $\boldsymbol{C}$ is diagonally dominant because the load capacitances C_{22} and C_{11} are substantially larger than the coupling capacitances C_{21} and C_{12}. This ensures $\det \boldsymbol{C} = C_{22}C_{11} - C_{21}C_{12} > 0$ as $C_{22}C_{11} \gg C_{21}C_{12}$, i.e. the matrix $\boldsymbol{C}$ is known to be invertible. We may thus formally transform (5) into

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \boldsymbol{C}(\boldsymbol{v})^{-1}\boldsymbol{I}(\boldsymbol{v}) + \boldsymbol{C}(\boldsymbol{v})^{-1}\boldsymbol{i}_{\mathbf{n}}(t) \\ &\equiv \boldsymbol{h}(\boldsymbol{v}) + \boldsymbol{\sigma}(\boldsymbol{v})\boldsymbol{w}(t). \end{aligned} \tag{6}$$

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

We refer to (6) as the (stochastic) state-space representation of the system [25]. $\boldsymbol{h}(\boldsymbol{v})$ is a two-dimensional (2D) *drift vector field*, further analyzed here below. The vector $\boldsymbol{w}(t) = (w_1(t), w_2(t))$ contains two independent (since generated by different devices) white noises of unit power spectral density (PSD); $\boldsymbol{\sigma}(\boldsymbol{v})$ is the square *diffusion matrix*, non-diagonal as a consequence of the capacitive coupling involved in (6) through the $\boldsymbol{C}(\boldsymbol{v})^{-1}$. The $\boldsymbol{\sigma}(\boldsymbol{v})$ matrix will be extensively explained thereafter, at the light of dedicated SPICE extractions.

Mathematically, a continuous-time white noise $w(t)$ of infinite bandwidth does not rigorously exist, because its variance is infinite. Only the integral of a white noise, i.e. a *Wiener process* (or Brownian noise) $W(t) = \int_0^t w(t)\,\mathrm{d}t$, is soundly defined [35]. This is why some authors [25], [36] prefer to adopt the more rigorous notation of *stochastic differential equation* (SDE): $\mathrm{d}\boldsymbol{v} = \boldsymbol{h}(\boldsymbol{v})\,\mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{W}(t)$. Written as (6), the stochastic state-space model is a (vectorial) nonlinear *Langevin equation*. It may only be formally interpreted as an SDE by adopting an interpretation rule for stochastic integral calculus [37], such as Itô or Stratonovich's scheme [36]. The question is relevant because the SRAM bitcell exhibits a nonlinear behavior and the noise intensity $\boldsymbol{\sigma}(\boldsymbol{v})$, modulated by the state $\boldsymbol{v}$, is not constant. As widely discussed elsewhere [36], [37], the two formulations imply, for a given physical system, a different function $\boldsymbol{h}(\boldsymbol{v})$ but lead to the *same* solution (statistical distribution of $\boldsymbol{v}(t)$). Previous works dedicated to electronic circuit modeling [35], [38] suggest, however, that Itô's interpretation follows naturally from conventional formulation like (4) and (6) and is especially consistent with the SPICE implementation of transient noise simulations. Thus, all through this paper, SDE like (6) must be understood in Itô's sense as assumed in [25] when performing stochastic calculus.

3) SPICE Characterization of the Drift: If the noise term vanishes in (6), the evolution of $\boldsymbol{v}(t)$ is only dictated by the drift field $\boldsymbol{h}(\boldsymbol{v})$:

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{h}(\boldsymbol{v}). \tag{7}$$

The interpretation is that, in the theoretical absence of noise, the state trajectory follows the natural "deterministic" dynamics of the silent cross-coupled inverters. As explained earlier in Section III-A, their regenerative action brings $(v_2(t), v_1(t))$ to either (X_0, Y_0) or (X_1, Y_1) depending on the initial state. Equation (7) provides a straightforward tool to extract $\boldsymbol{h}(\boldsymbol{v})$ from conventional noiseless transient simulations (TRAN in SPICE). For this purpose, we use a uniform grid in the state space of Figure 3(b) as initial conditions. We record the evolutions of the different trajectories $(v_2(t), v_1(t))$ (Figure 5(a)) and compute the two time derivatives $\mathrm{d}v_2/\mathrm{d}t$ and $\mathrm{d}v_1/\mathrm{d}t$ numerically (the two scalar components of (7)). A discretised $\boldsymbol{h}(\boldsymbol{v})$ is finally obtained by mapping the vectors $\mathrm{d}\boldsymbol{v}/\mathrm{d}t$ and $\boldsymbol{v}(t)$. It is depicted in Figure 5(b) as a vector field, with a zoom on the region relevant for bit-flip analysis.

Figure 5(a) shows the usual characteristics of a bistable system. The two stable steady states (X_0, Y_0) or (X_1, Y_1) attract, from all directions, the trajectories in their respective basin of attraction. The unstable steady state $(X_\mathrm{M}, Y_\mathrm{M})$ exhibits a typical saddle-point behavior [39]. It is *repulsive* along its *unstable manifold*, which coincides with the *stable manifold* of (X_0, Y_0). The *heteroclinic orbit* from (X_0, Y_0) to (X_1, Y_1) [40] is roughly the $\tilde{v}$ straight line connecting them (parametrised by $\tilde{v}$ in violet in Figure 3(b)). The orbit is,

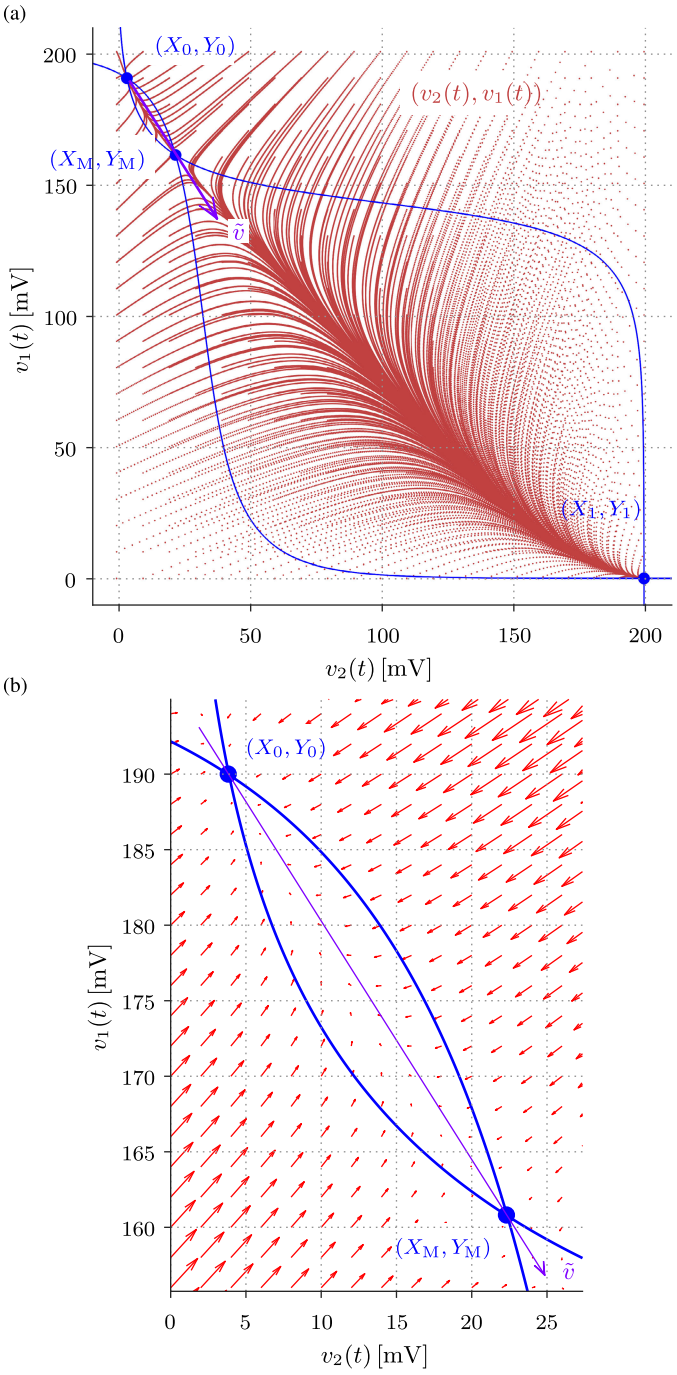

Fig. 5. (a) Simulated deterministic $(v_2(t), v_1(t))$ trajectories in the state space, starting from uniformly gridded initial conditions. The stable steady states (X_0, Y_0) and (X_1, Y_1) (valley points) are attractive. The unstable steady state $(X_\mathrm{M}, Y_\mathrm{M})$ (saddle point) is repulsive along its manifold closely approximated by the straight line parametrised by $\tilde{v}$ (in violet), and attractive in the orthogonal direction (separatrix). (b) Two-dimensional drift field $\boldsymbol{h}$, with a zoom on the stable steady state (X_0, Y_0) weakened by process variability. The intensity of $\boldsymbol{h}$ (i.e. arrow size) in the orthogonal direction tends to quickly push $(v_2(t), v_1(t))$ toward the $\tilde{v}$ axis (shown in violet), which supports the observation of the time-scale separation and the underlying assumption of reduction of dimensionality. Illustrated case: same as Figure 3.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

in the region between (X_0, Y_0) and (X_M, Y_M), contained in the stable manifold of (X_0, Y_0) and in the unstable manifold of (X_M, Y_M). The unstable steady state (X_M, Y_M) is *attractive* along its stable manifold, corresponding to the more or less straight separatrix [32] in the orthogonal direction.

More specific to the SRAM bitcell and useful for our purposes is the observation of a *time-scale separation*: whatever the exact initial condition, the trajectory rapidly converges, in a very short time, toward the heteroclinic orbit ($\tilde{v}$ straight line connecting (X_0, Y_0) and (X_M, Y_M) in Figures 3(b) and 5(b)). Then, the evolution toward either attractive steady state along this orbit is *slower*.

Those notions of velocity and time are more easily appreciated in Figure 5(b). Outside the wing of the butterfly, the drift vector field points almost perpendicularly toward the $\tilde{v}$ line, with large arrows translating high velocity and very short time constant. Along the unstable manifold of (X_M, Y_M), the tiny arrows show that the unidimensional dynamics in $\tilde{v}$ is slow and hence is expected to dominate the transition time. It is possible to make those observations more quantitative, at least near the steady states. One may for instance linearize (7) near the stable point $\boldsymbol{v}_0$

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \approx \frac{\mathrm{d}\boldsymbol{h}}{\mathrm{d}\boldsymbol{v}}_{\boldsymbol{v}=\boldsymbol{v}_0} (\boldsymbol{v} - \boldsymbol{v}_0) \equiv \boldsymbol{A}(\boldsymbol{v} - \boldsymbol{v}_0), \tag{8}$$

since $\boldsymbol{h}(\boldsymbol{v}_0) = \boldsymbol{0}$ by definition, and to compute the eigendecomposition of the Jacobian matrix $\boldsymbol{A}$, i.e. calculate the eigenvalues and eigenvectors to diagonalize it: $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$. The time constants of the linearized system in the orthogonal directions are the inverse of the purely real eigenvalues contained in $\boldsymbol{D}$:

$$\boldsymbol{D} = \begin{pmatrix} -1/\tau_{0,\perp} & 0 \\ 0 & -1/\tau_{0,\tilde{v}} \end{pmatrix}. \tag{9}$$

The negative sign of the eigenvalues indicates the stability, attractiveness of the steady state $\boldsymbol{v}_0$. For the case illustrated throughout the text, we find $\tau_{0,\perp} \approx 10\,\mathrm{ns}$ much shorter than $\tau_{0,\tilde{v}} \approx 70\,\mathrm{ns}$. Last but not least, we find that the eigenvector associated to the eigenvalue $-1/\tau_{0,\tilde{v}}$ (slow time constant) closely matches the direction of the straight line connecting (X_0, Y_0) and (X_M, Y_M) whose unit vector is

$$\tilde{\boldsymbol{v}} = \left(\frac{X_M - X_0}{\Delta\tilde{v}} \;\; \frac{Y_M - Y_0}{\Delta\tilde{v}} \right)^{\mathsf{T}}, \tag{10}$$

with

$$\begin{aligned} \Delta\tilde{v} &= \sqrt{(X_M - X_0)^2 + (Y_0 - Y_M)^2} \\ &= \sqrt{\Delta X^2 + \Delta Y^2}. \end{aligned} \tag{11}$$

We say that the eigenvectors span the stable linear space, which is locally tangent to the manifolds of the steady state [40].

A similar analysis near the unstable point leads to the same conclusion, except that the eigenvalue associated to $\tilde{\boldsymbol{v}}$ is *positive*: $+1/\tau_{M,\tilde{v}}$, indicating the repulsiveness of this steady state along its unstable manifold. In the intermediate region between (X_0, Y_0) and (X_M, Y_M), the magnitudes of the drift field arrows (Figure 5) makes the time-scale separation of the slow dynamics along $\tilde{\boldsymbol{v}}$ and the fast one along the orthogonal directions even more valid.

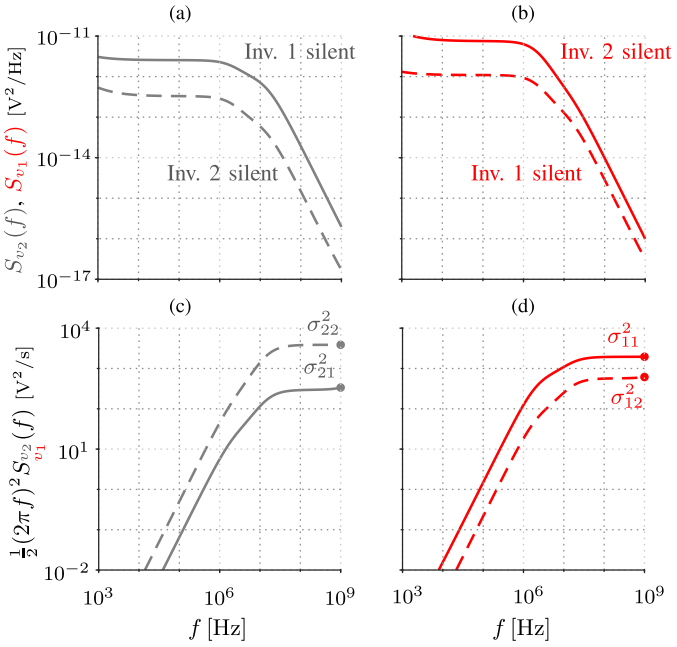

Fig. 6. Power spectral densities of the node voltages (a) v_2 (gray) and (b) v_1 (red) obtained from `NOISE(AC)` simulations of the circuit of Figure 2(a), here around the stable steady state $(v_2, v_1) = (X_0, Y_0)$. The individual contribution of each inverter (plus the access transistor connected to its output) is extracted separately by disabling the noise (`NONOISE` command in SPICE) of the other inverter-transistor pair (then said silent). For a given node (e.g. v_2 in (a)), they are thus two configurations: in solid line, the dominant noise contribution from the *driving* inverter (e.g. Inverter 2); in dashed line, the smaller (yet non negligible) contribution of the *load* inverter (e.g. Inverter 1). Illustrated case: same as Figure 3. (c) and (d) Characterization of $\boldsymbol{\sigma}(X_0, Y_0)$ through the separate extraction of the four σ^2_{ij}: $\sigma^2_{22} = 3872\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{21} = 336\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{11} = 1957\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{12} = 614\,\mathrm{V^2\,s^{-1}}$. Similar simulations were performed around the *unstable* steady state to get $\boldsymbol{\sigma}(X_M, Y_M)$ (not shown): $\sigma^2_{22} = 1705\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{21} = 271\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{11} = 1320\,\mathrm{V^2\,s^{-1}}$; $\sigma^2_{12} = 301\,\mathrm{V^2\,s^{-1}}$. Extracted infinitesimal variances of 1D white noise of (19) at steady states: $\sigma^2_0 = 910\,\mathrm{V^2\,s^{-1}}$ (stable) and $\sigma^2_M = 565\,\mathrm{V^2\,s^{-1}}$ (unstable).

While the two-dimensional drift was extracted and depicted in Figure 5(b) for illustrative purposes, as well as to validate the hypothesis of time-scale separation and unidimensional dynamics, it will not be used to predict the $MTTF$. Instead, we will rely on the reduced unidimensional model presented in Section IV-B.

4) SPICE Characterization of the Noise: We now explain how to characterise $\boldsymbol{\sigma}(\boldsymbol{v})$ in the noise term of the SDE (6). First of all, let us emphasize that the noise can only be inferred, from fluctuations of the observables v_2 and v_1 at the steady states of the circuit, i.e. not at any arbitrary (v_2, v_1) of the state space. This problem was generally referred to as "hidden stochasticity" in [25], as encountered in many physical systems. Our goal is thus to extract $\boldsymbol{\sigma}(X_0, Y_0)$ and $\boldsymbol{\sigma}(X_M, Y_M)$, i.e. four quantities for each steady state (either 0 or M):

$$\boldsymbol{\sigma} \equiv \begin{pmatrix} \sigma_{22} & \sigma_{21} \\ \sigma_{12} & \sigma_{11} \end{pmatrix}. \tag{12}$$

From (6) we develop, for instance, the equation of $\mathrm{d}v_2/\mathrm{d}t$:

$$\frac{\mathrm{d}v_2}{\mathrm{d}t} = h_2(v_2, v_1) + \sigma_{22} w_2(t) + \sigma_{21} w_1(t). \tag{13}$$

Equation (13) clearly shows that $v_2(t)$ receives two independent noise contributions: the dominant $\sigma_{22} w_2(t)$ from its

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

driving Inverter 2; the smaller $\sigma_{21} w_1(t)$ from the load Inverter 1. An important ingredient is that we are going to extract σ_{22} and σ_{21} separately from two different SPICE simulations by disabling the noise of the other inverter (`NONOISE` command in SPICE). For instance, if Inverter 1 (as well as the access transistor connected to its output) is silent, (13) becomes

$$\frac{\mathrm{d}v_2}{\mathrm{d}t} = h_2(v_2, v_1) + \sigma_{22} w_2(t), \tag{14}$$

i.e. reduces to a scalar nonlinear Langevin equation with one single white noise.

Previous works [8], [9] attempted to subsequently extract the infinitesimal variance of the white noise (σ_{22}^2 in (14)) by assimilating the SRAM bitcell to a linear *RC* circuit near the steady state. Under this single-pole assumption, σ_{22}^2 relates to the *integrated* voltage noise power

$$\sigma_{v_2}^2 = \int_0^\infty S_{v_2}(f)\,\mathrm{d}f, \tag{15}$$

according to [9]:

$$\sigma_{22}^2 = 4\pi f_{\mathrm{p},22}\,\sigma_{v_2}^2 \tag{16}$$

where $f_{\mathrm{p},22}$ would be the fitted value of the only presumed dominant pole of $S_{v_2}(f)$. In passing, we would like to point out that, for device modeling and SPICE simulations, PSD like $S_{v_2}(f)$ are, by convention, *one-sided*, i.e. only defined and integrated over positive frequencies to retrieve the noise power as in (15). At first sight of the voltage noise PSD extracted from `NOISE(AC)` simulations (Figures 6(a) and 6(b)), the *RC* low-pass-filter assumption, whose signature is a first-order transfer function with a dominant pole, seems to be fairly well-founded. However, the method (16) has turned out to be very sensitive to slight perturbations of the PSD attributed to higher-order poles and zeros introduced by the various capacitances of the SRAM bitcell. The usage of (16) should thus be avoided.

Fortunately, it is possible to directly relate σ_{22}^2 to the high-frequency behavior of the PSD $S_{v_2}(f)$, as explained below:

$$\sigma_{22}^2 = \frac{1}{2}(2\pi f^*)^2 S_{v_2}(f^*) \quad \text{(Inv. 1 silent)}. \tag{17}$$

The f^* selected for the extraction is a sufficiently high frequency where $f^2 S_{v_2}(f)$ (corresponding, within constants, to the one-sided PSD of the output current noise) is nearly flat (e.g. 1 GHz in Figures 6(c) and 6(d)). Equation (17) is trivial for an *RC* circuit, and can be justified to be valid more generally.

The key ingredient is to relate $S_{v_2}(f)$, the PSD of $v_2(t)$, to the one-sided PSD of the white noise $\sigma_{22} w_2(t)$, which is $2\sigma_{22}^2$. Near the steady state, $v_2(t)$ may be regarded as a first-order Markov process resulting from the filtering of the white noise $\sigma_{22} w_2(t)$ by the system (13) (yet with $h_2(v_2, v_1)$ linearized). The PSD is then given by [41, Section II-D,eq. (34)]. For a high frequency beyond the dominant pole of the linearised system, i.e. $f^* \gg f_{\mathrm{p},22}$, $S_{v_2}(f^*)$ converges to the PSD of the Wiener process $W_2(t) = \int_0^t \sigma_{22} w_2(t)\,\mathrm{d}t$, which can be shown to be $(2\sigma_{22}^2)/(2\pi f^*)^2$ (see [41, eq. (36)] or [42, eq. (7)]). Isolating σ_{22}^2 finally leads to (17).

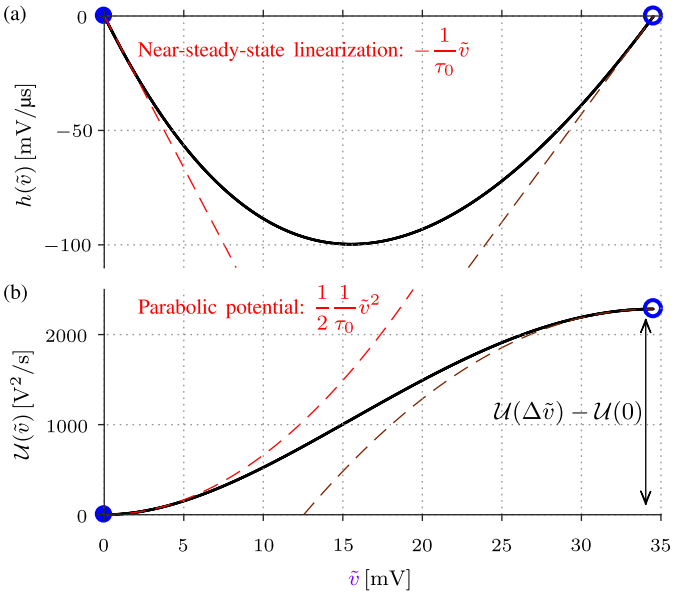

Fig. 7. (a) Extracted deterministic drift as appears in (19), which derives from (b) quasi potential $\mathcal{U}(\tilde{v}) = -\int_0^{\tilde{v}} h(\tilde{v}')\,\mathrm{d}\tilde{v}'$. The near-stable-steady-state ((X_0, Y_0), full blue dot) approximations inherent to [23] and [26]'s formulas are shown (dashed red line). Eyring-Kramers formula also exploits the behavior near the unstable point ($(X_\mathrm{M}, Y_\mathrm{M})$, open blue dot and dashed brown curve). Illustrated case: same as Figure 3. $\Delta\tilde{v} = 34.5$ mV.

Intuitively, in high frequency, the dynamics $h_2(v_2, v_1)$ is ineffective in varying $v_2(t)$ compared to the white noise, whose intensity is thereby contained in the tail of $S_{v_2}(f)$. The $(2\pi f^*)^2$ factor in (17) results from the integral relationship between $v_2(t)$ and $\sigma_{22} w_2(t)$.

The three other σ_{ij}^2 are extracted similarly, as illustrated in Figures 6(c) and 6(d) (the numerical values both at the stable and unstable steady states are provided in the caption). This procedure to characterise $\boldsymbol{\sigma}(X_0, Y_0)$ and $\boldsymbol{\sigma}(X_\mathrm{M}, Y_\mathrm{M})$ is fundamental in the *MTTF* prediction methodology, since those quantities will come into play in the closed-form analytical formulas.

B. Reduced Unidimensional Stochastic Model

Upon observation of transient experiments like Figure 3 (other trajectories simulated for other cases showed similar behavior) and following the argument of time-scale separation supported by the previous section (Figure 5), it seems reasonable to reduce the dimensionality of the system (4). To this end, we introduce an unidimensional (1D) coordinate $\tilde{v}$ (previously shown in violet in Figures 3(b) and 5), constructed by affine transformation of (v_2, v_1) obtained by orthogonal projection of the 2D state vector $\boldsymbol{v}$ on the $\tilde{\boldsymbol{v}}$ defined in (10):

$$\begin{aligned}\tilde{v} &= \tilde{\boldsymbol{v}}^\mathsf{T} \cdot \boldsymbol{v} \\ &\equiv \frac{X_\mathrm{M} - X_0}{\Delta\tilde{v}} \cdot v_2 + \frac{Y_\mathrm{M} - Y_0}{\Delta\tilde{v}} \cdot v_1,\end{aligned} \tag{18}$$

where $Y_\mathrm{M} - Y_0 < 0$ and $\Delta\tilde{v}$ is the Euclidean distance between (X_0, Y_0) and $(X_\mathrm{M}, Y_\mathrm{M})$, as defined earlier in (11).

This is equivalent to studying the nonlinear SRAM bitcell dynamics in the main direction $\tilde{\boldsymbol{v}}$ dominating the *TTF*,

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

modelled by the *scalar* Langevin equation, assuming a first-order dynamics [9]:

$$\frac{\mathrm{d}\tilde{v}}{\mathrm{d}t} = h(\tilde{v}) + \sigma(\tilde{v})w(t). \tag{19}$$

Equation (19) may be thought as formally describing a *non-linear RC* circuit, i.e. an overdamped single-state system with drift-diffusion dynamics [25]. We refer to (19) to as the *unidimensional stochastic model* of the SRAM bitcell [9], as a special case of single-state nonlinear bistable system [25]. Below, we explain how $h(\tilde{v})$ and σ are characterized from the SPICE simulations presented in Section IV-A.

1) Characterization of $h(\tilde{v})$: The function $h(\tilde{v})$ (in $\mathrm{V\,s^{-1}}$), determining the natural dynamics of the 1D system in absence of noise, absorbs the nonlinearities of both the resistive and capacitive components of the SRAM bitcell. It can be cheaply and unambiguously extracted from a noiseless transient simulation (`TRAN` in SPICE, similarly to its 2D field counterpart in Section IV-A). The procedure goes as follows [9]: 1) starting from an initial condition $(v_2(0), v_1(0)) = (X_\mathrm{M} - \epsilon, Y_\mathrm{M} + \epsilon)$; the deterministic evolution of $(v_2(t), v_1(t))$ toward (X_0, Y_0) is recorded; 2) $\tilde{v}(t)$ is calculated through (18); 3) $\mathrm{d}\tilde{v}/\mathrm{d}t$ is computed numerically and mapped to $\tilde{v}(t)$ to provide $h(\tilde{v})$ between 0 and $\Delta\tilde{v}$ (depicted in Figure 7(a)). Again, this extraction is nothing but a special case, one single trajectory of Figure 5(b) originating from $(X_\mathrm{M} - \epsilon, Y_\mathrm{M} + \epsilon)$. In 1D, drift $h(\tilde{v})$ may always be thought as deriving from a scalar *quasi potential* $\mathcal{U}(\tilde{v})$ [39] (extracted and represented on $[0, \Delta\tilde{v}]$ in Figure 7(b)):

$$h(\tilde{v}) \equiv -\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}\tilde{v}} \iff \mathcal{U}(\tilde{v}) - \mathcal{U}(0) = -\int_0^{\tilde{v}} h(\tilde{v}')\,\mathrm{d}\tilde{v}'. \tag{20}$$

The stable steady states (notably the fragile $\tilde{v} = 0$) corresponds to the valleys of $\mathcal{U}(\tilde{v})$, whereas the hill characterises the unstable point ($\tilde{v} = \Delta\tilde{v}$, labelled M). The quasi potential landscape directly relates to the steady-state occupation probability (e.g. [20], [21]) through a Boltzmann relation.

2) Characterization of $\sigma(\tilde{v})$: In (19), $\sigma(\tilde{v})w(t)$ is the thermal noise term, $\sigma^2(\tilde{v})$ being the intensity, or infinitesimal variance, or variance per unit time (i.e. in $\mathrm{V^2\,s^{-1}}$) and $w(t)$ the white Gaussian noise process (in $\mathrm{s^{-1/2}}$) [9] of unit two-sided PSD (i.e. the one-sided PSD of $w(t)$ is 2, dimensionless). The meaning to be attributed to the notion of "infinitesimal variance" is the following: $\mathrm{Var}\{\int_0^t \sigma\, w(t)\,\mathrm{d}t\} = \sigma^2 t$, i.e. $\mathrm{Var}\{\int_0^t \sigma\, w(t)\,\mathrm{d}t\}$ is proportional to the time window. An example is the variance of a white current noise (e.g. physically generated within a CMOS inverter) integrated on an lossless capacitance [43].

As mentioned in Section IV-A, the noise can only be inferred, in our setup, at the steady states of the circuit. Consequently, $\sigma(\tilde{v})$ is, at best, known in $\tilde{v} = 0$ and $\tilde{v} = \Delta\tilde{v}$, where we respectively denote it σ_0 and σ_M. Most previous analytical formulas coarsely assumes that $\sigma^2(\tilde{v})$ is *constant*, i.e. that it does *not* vary with $\tilde{v}$ throughout the bit flip (e.g. assumed equals to σ_0^2). Finer prediction would rely on the knowledge of both σ_0^2 and σ_M^2 as will be reviewed in Section V.

The σ (evaluated at either steady states) of (19) relates to the matrix $\boldsymbol{\sigma}$ of (6), whose four entries have been extracted from `NOISE(AC)` simulations according to Figure 6. To establish the link, we first note from (6) and (18) that

$$\begin{aligned}\frac{\mathrm{d}\tilde{v}}{\mathrm{d}t} &= \frac{\mathrm{d}(\tilde{\boldsymbol{v}}^\intercal \cdot \boldsymbol{v})}{\mathrm{d}t} \\ &= \tilde{\boldsymbol{v}}^\intercal \cdot \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \\ &= \tilde{\boldsymbol{v}}^\intercal \cdot \boldsymbol{h}(\boldsymbol{v}) + \tilde{\boldsymbol{v}}^\intercal \boldsymbol{\sigma}\, \boldsymbol{w}(t).\end{aligned} \tag{21}$$

By identification with (19), we find

$$\sigma\, w(t) = \tilde{\boldsymbol{v}}^\intercal \boldsymbol{\sigma}(\boldsymbol{v}) \boldsymbol{w}(t), \tag{22}$$

or, written in terms of white noise variance,

$$\begin{aligned}\sigma^2 &= \tilde{\boldsymbol{v}}^\intercal \boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal \tilde{\boldsymbol{v}} \\ &= \begin{pmatrix} \frac{X_\mathrm{M}-X_0}{\Delta\tilde{v}} & \frac{Y_\mathrm{M}-Y_0}{\Delta\tilde{v}} \end{pmatrix} \begin{pmatrix} \sigma_{22} & \sigma_{21} \\ \sigma_{12} & \sigma_{11} \end{pmatrix} \begin{pmatrix} \sigma_{22} & \sigma_{12} \\ \sigma_{21} & \sigma_{11} \end{pmatrix} \begin{pmatrix} \frac{X_\mathrm{M}-X_0}{\Delta\tilde{v}} \\ \frac{Y_\mathrm{M}-Y_0}{\Delta\tilde{v}} \end{pmatrix}.\end{aligned} \tag{23}$$

Equation (23) is used to calculate σ_0^2 and σ_M^2 from the previously extracted quantities.

V. Approximate Closed-Form Prediction Formulas

Now characterized, the unidimensional stochastic model (19) points to the hope of analytically predicting the $MTTF$. There are, however, two obstacles that should not be overlooked: (i) the SDE (19) is nonlinear through $h(\tilde{v})$ (Figure 7(a)), and any attempt of linearization is expected to result in significant discrepancies, as will be evidenced; (ii) even if the parameters of (19) have been accurately characterized, including the noise, deducing a (semi-)analytical general formula for the $MTTF$ from (19) is not a straightforward mathematical task. We will therefore draw on literature from various scientific fields.

Let us first mention our preliminary work [9]. A partially empirical estimate of the $MTTF$ was obtained by efficiently simulating trajectories of $\tilde{v}(t)$ from the SDE (19), in a Monte-Carlo fashion. *Euler–Maruyama method* [44], a dedicated numerical integration scheme, was used. For each $\tilde{v}(t)$ trace, the TTF corresponds to twice the first time to reach $\Delta\tilde{v}$, since at the top of the hill (Figure 7(b)) the state can still flip left or right with probability $1/2$. Although attractive in CPU time (reduced to barely a few minutes compared to the three days of transient noise simulations in SPICE), the accuracy of this attempt was quite low due to the incorrect constant-σ assumption. Moreover, such empirical approach, based on repeated trials, does not offer an insightful closed-form formula for the $MTTF$.

A. Near-Stable-Steady-State Approximate Formulas

Our proceedings [8], [9] already pointed out the severe inaccuracy of some analytical formulas [23], [26] that rely on a linearization of the SRAM bitcell dynamics, $h(\tilde{v})$ in (19). We review them in the present section.

1) First-Passage Time of an Ornstein-Uhlenbeck Process: Translating the definition of Section III-A in 1D, the TTF as the time taken by $\tilde{v}(t)$ to, starting from 0, exceed $\Delta\tilde{v}$ for the last time before successfully transitioning to the other stable point. The mean TTF ($MTTF$) is twice the mean *first*

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

passage time (FPT), which is the mean time to reach $\Delta\tilde{v}$ for the *first* time starting from 0 [25]; the factor 2 accounts for the probability 1/2 of falling left or right when reaching the unstable point at $\Delta\tilde{v}$.

A closed-form solution of such FPT problem exists, unfortunately only for the special case of linearized drift term

$$h(\tilde{v}) \approx -\frac{1}{\tau_0}\tilde{v}(t) \tag{24}$$

(dashed red line in Figure 7(a)), with

$$\frac{1}{\tau_0} = -\frac{\mathrm{d}h}{\mathrm{d}\tilde{v}}(0) = \frac{\mathrm{d}^2\mathcal{U}}{\mathrm{d}\tilde{v}^2}(0), \tag{25}$$

the time constant of the 1D system linearized at $\tilde{v} = 0$. The $\tilde{v}$ defined by (19) and (24) is called an *Ornstein-Uhlenbeck process*. By construction, the τ_0 defined in (25) is the $\tau_{0,\tilde{v}}$ found in (9) when searching for the eigenvalues of the Jacobian matrix of the drift field of the 2D system.

Nobile [26] has rigorously solved the FPT problem of an Ornstein-Uhlenbeck process, leading to [26, (6a)]:

$$MTTF \approx 2\,\tau_0 \left(\sqrt{\pi}\int_0^{\Delta\tilde{v}/(\sigma_0\sqrt{\tau_0})} \mathrm{d}u\ \exp\left(u^2\right) + \int_0^{\Delta\tilde{v}/(\sigma_0\sqrt{\tau_0})} \mathrm{d}u\ \exp\left(u^2\right)\mathrm{erf}(u)\right). \tag{26}$$

2) Kish's Formula: Kish [23] proposed a simplified Rice formula for the mean frequency of crossing ($1/MTTF$ in our notations) a given threshold voltage ($\Delta\tilde{v}$) by a Gaussian noise process (as $\tilde{v}(t)$ is roughly assumed) [23, (8)]. Kish's formula in integral form [23, (8)] diverges when applied to the PSD of the output voltage noise of an inverter (see, for instance, Figures 6(a) and 6(b)). It is most often used (notably [24, (4)]) in a simplified form [23, (9)] supposing band ($f_\mathrm{p} = 1/(2\pi\,\tau_0)$)-limited white (thermal) noise, written with the notations of this paper as:

$$\frac{1}{MTTF} = \frac{2}{\sqrt{3}}\exp\left(-\frac{1}{2}\left(\frac{\Delta\tilde{v}}{\sigma_{\tilde{v}}}\right)^2\right) f_\mathrm{p} \propto \exp\left(-\frac{\Delta\tilde{v}^2/(2\tau_0)}{\sigma_{\tilde{v}}^2/\tau_0}\right)\frac{1}{\tau_0}. \tag{27}$$

Despite lack of detailed proof, we understand (27) as linearizing the true SRAM bitcell dynamics around the stable steady state to reduce it to an RC-like circuit with Gaussian voltage distribution. It can be regarded as a simplified version of (26) [25].

Within the exponential, $\Delta\tilde{v}^2/(2\tau_0)$ is the value of the quasi-potential barrier evaluated at the unstable point ($\tilde{v} = \Delta\tilde{v}$) under the coarse harmonic (parabolic) potential approximation (dashed red line in Figure 7(b)), equivalent to linearized drift (24). Remembering that for an RC circuit we have $\tau_0 = RC$ and $\sigma_{\tilde{v}}^2 = kT/C$, (27) finally highlights a Maxwell-Boltzmann probability [9]:

$$1/MTTF \propto \exp(-\Delta E/kT)/\tau_0, \text{with} \Delta E = C\,\Delta\tilde{v}^2/2.$$

B. Eyring-Kramers Formulas

The calculation of transition rates ($1/MTTF$) is a classical problem in chemical physics. The old Eyring-Kramers formula [45] is valid for a wide class of nonlinear bistable systems but requires a constant white noise intensity. Our recently published extended formula [25] overcomes this limitation.

1) Classical Eyring-Kramers Formula: Constant σ^2: The near-stable-steady-state approximation inherent to [23] and [26] is twofold: the deterministic drift is linearized around $\tilde{v} = 0$ only; the noise is assumed constant and also extracted at $\tilde{v} = 0$ only. As shown numerically in [25], drastic accuracy improvement is achieved by Eyring-Kramers formula [45], revisited in [39] and applied to the SRAM bitcell model (19) in [25]:

$$MTTF \approx 2\pi\sqrt{\tau_0\,\tau_\mathrm{M}}\,\exp\left(\frac{\mathcal{U}(\Delta\tilde{v}) - \mathcal{U}(0)}{\frac{1}{2}\sigma^2}\right). \tag{28}$$

Equation (28) also takes into account the drift (quasi-potential) behavior of the system near the *unstable* point (dashed brown lines in Figure 7).

We explain the newly introduced quantities in (28). $\mathcal{U}(\Delta\tilde{v}) - \mathcal{U}(0)$ is the *true* quasi-potential barrier (computed according to (20) and indicated in Figure 7(b)) whose crossing by the random diffusion process leads to a bit flip.

$$\frac{1}{\tau_\mathrm{M}} = \frac{\mathrm{d}h}{\mathrm{d}\tilde{v}}(\Delta\tilde{v}) = \left|\frac{\mathrm{d}^2\mathcal{U}}{\mathrm{d}\tilde{v}^2}(\Delta\tilde{v})\right| \tag{29}$$

is the time constant of the system linearized at the unstable (and repulsive) point ($\tilde{v} = \Delta\tilde{v}$), straightforwardly computed from either $h(\tilde{v})$ or $\mathcal{U}(\tilde{v})$. Hence, a new effective time constant of the diffusion process, defined as a geometrical mean $\tau \equiv \sqrt{\tau_0\tau_\mathrm{M}}$, appears in the pre-factor of (28). It somehow incorporates the drift behavior in both the near stable- and unstable-steady-state regions.

Unfortunately, the classical formula (28) still assumes constant noise variance σ^2. The numerical values, for instance those provided for the illustrated case in the caption of Figure 6, show that σ_0 can be significantly different from σ_M, i.e. that the noise intensity *does* significantly vary along the bit-flip trajectory. This behavior, for nonlinear devices, was expected from physical considerations [34]. No observation or theoretical argument justifies favouring either $\sigma = \sigma_0$ or $\sigma = \sigma_\mathrm{M}$ for our purposes. Thus, these two exclusive choices could be inaccurate, when attempting to apply (28) to predict the $MTTF$ in SRAM bitcells.

2) Extended Eyring-Kramers Formula: σ_0 and σ_M: In [25], we have addressed the situation of "hidden stochasticity", i.e. when $\sigma(\tilde{v})$ is only known in 0 and $\Delta\tilde{v}$ and hidden outside the steady states. The main result is

$$MTTF \approx 2\pi\sqrt{\tau_0\,\tau_\mathrm{M}}\,\frac{\sigma_0}{\sigma_\mathrm{M}}\,\exp\left(\frac{\mathcal{U}(\Delta\tilde{v})}{\frac{1}{2}\sigma_\mathrm{M}^2}\right)\exp\left(-2\left(\frac{1}{\sigma_\mathrm{M}^2} - \frac{1}{\sigma_0^2}\right)\overline{\mathcal{U}}\right), \tag{30}$$

where

$$\overline{\mathcal{U}} \equiv \frac{1}{\Delta\tilde{v}}\int_0^{\Delta\tilde{v}} \mathcal{U}(\tilde{v}')\ \mathrm{d}\tilde{v}'. \tag{31}$$

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

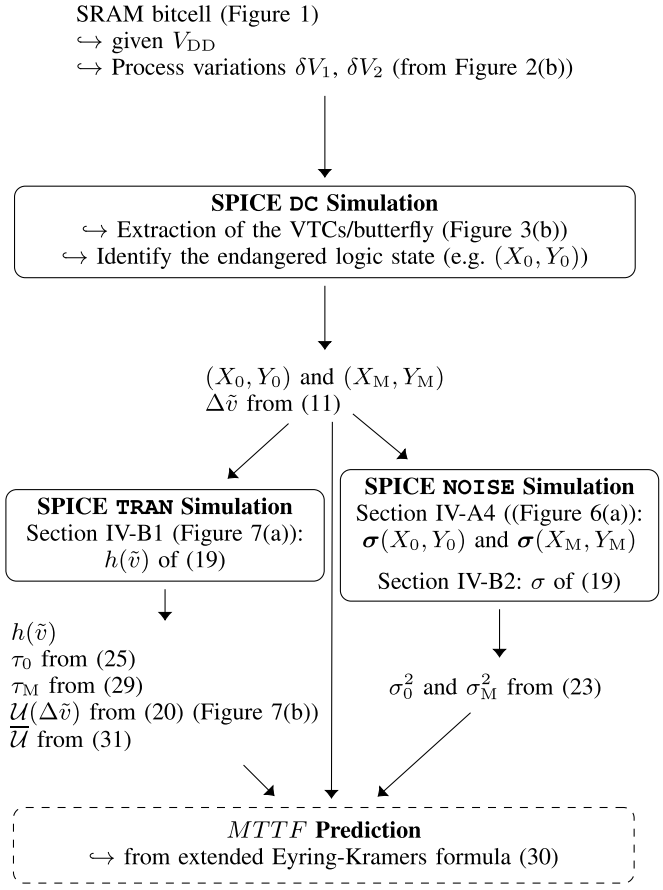

Fig. 8. Flow chart of the proposed methodology, aimed at predicting the mean time to failure due to the noise in ULV SRAM bitcells affected by process variations (assumed given, as selected from Figure 2(b)).

Equation (30) does incorporate the noise behavior near both the stable (σ_0) and the unstable (σ_M) steady states. It is somewhat similar to the classical formula (28) where $\sigma = \sigma_M$ would be used, yet with additional factors:

- σ_0/σ_M naturally appears following Itô's rule for stochastic calculus [25]. It is expected to have a negligible impact on the order of magnitude of the $MTTF$, compared to the exponentials.
- The factor

$$\exp\left(-2\left(\frac{1}{\sigma_M^2}-\frac{1}{\sigma_0^2}\right)\overline{\mathcal{U}}\right) \tag{32}$$

may be regarded as correcting for the fact that the noise variance is not constant throughout the state flip. It is unity only when $\sigma(\tilde{v}) = \sigma_0 = \sigma_M =$ constant. If $\sigma_M < \sigma_0$, the argument within the exponential is *positive* and the factor is larger than 1, thereby compensating the excessive first exponential involving a too small σ. Conversely, if $\sigma_M > \sigma_0$, the correcting factor is smaller than 1.

The complete methodology, aimed at predicting the $MTTF$ in a given SRAM bitcell at a given V_{DD} affected by process variations, from the characterization of the stochastic model based on SPICE simulations to the extractions of the different quantities involved in the analytical extended Eyring-Kramers formula (30), has been summarized as a flow chart in Figure 8.

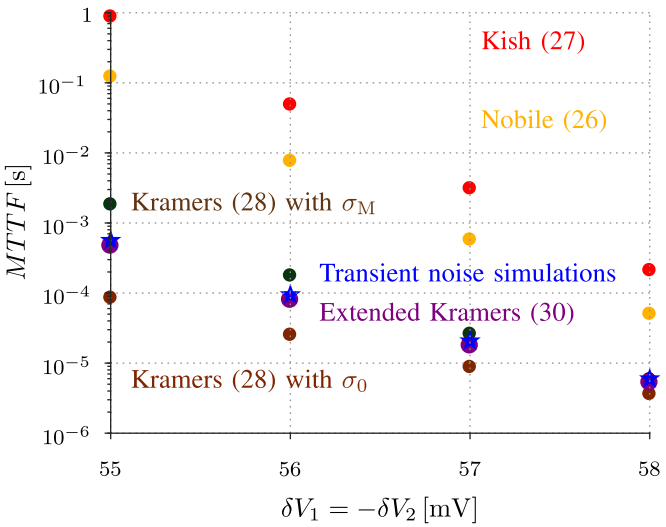

Fig. 9. $MTTF$ empirically estimated from transient noise simulations (averaged over 1000 experiments like Figure 3(a), for each point) compared to the predictions of the analytical formulas.

VI. Prediction of the $MTTF$: Results and Discussion

In Figure 9, we assess the accuracy of the different closed-form formulas for the $MTTF$, taking the massive and finely-tuned transient noise simulations of Section III as reference golden values. We exploit the unidimensional stochastic model (19) of the SRAM bitcell to insightfully point out and explain the origins of the discrepancies.

Reference values for the $MTTF$ were obtained by averaging 1000 TTF (which leads to histograms like Figure 4), extracted from SPICE transient noise simulations like Figure 3(a) [8] for several severe process variations $\delta V_1 = -\delta V_2 = 55$ to 58 mV (blue pentagrams in Figure 9). As previously observed in [7], the metric spans across orders of magnitude: it drops from half a ms (as provided earlier in Figure 4). at $\delta V_1 = -\delta V_2 = 55$ mV to a only few μs at 58 mV. This exponential trend has been explained earlier by the reduction of the SNM with δV, from 5 mV to 1 mV [8], and could have been anticipated from the various approximate formulas such as (27) or (28).

The low reported $MTTF$ values confirm that, while they were considered functional at time zero ($SNM > 0$), all these bitcells weakened by static variations are prone to dynamic instability and should be classified as defective when the typical retention time of the SRAM is for example longer than 1 ms in the ULP system application. An extrapolation of the robust (as validated with transient noise simulations) predictions (30) of Figure 9 to smaller process variations, e.g. $\delta V_1 = -\delta V_2 = 50$ mV corresponding to an $SNM = 10$ mV, estimates $MTTF$ greater than $1-10$ s. Because this value is larger than any practical retention time, it justifies *a posteriori* the choice of $SNM = 10$ mV as first guess to define the boundary of the orange crown in Figure 2(b). Similarly, no bit flip would ever be observed or practically measured on silicon in SRAM bitcells in conditions closer to nominal (e.g. $V_{DD} = 1$ V).

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

We first review the predictions of (26) and (27), respectively depicted in red and orange in Figure 9. As extensively discussed in previous publications [8], [9], [25], they totally fail in accurately predicting the $MTTF$, the discrepancy reaching two to three orders of magnitude for larger $MTTF$. This comes from the near-stable-steady-state linearization from which these formulas derive. In particular, the approximation (24) is inappropriate in describing the nonlinear SRAM dynamics (Figure 7(a)). Ornstein–Uhlenbeck model assumes a linear drift term $-\tilde{v}/\tau_0$ all along the trajectory and thereby severely overestimates the recall effect to $\tilde{v}=0$. In particular, it gives non-zero drift term $-\Delta\tilde{v}/\tau_0$ at $\tilde{v}=\Delta\tilde{v}$ while it is exactly zero within the full nonlinear model (19) (see Figure 7(a)). Equivalently, through the parabolic approximation, Kish overestimates the quasi-potential barrier that the process $\tilde{v}(t)$ must cross Figure 7(b)):

$$\frac{1}{2}\frac{1}{\tau_0}\Delta\tilde{v}^2 \gg \mathcal{U}(\Delta\tilde{v})-\mathcal{U}(0). \tag{33}$$

This clearly questions the failure rate predictions of [24], based on Kish formula.

The major improvement brought by the different Eyring-Kramers formula of Section V-B is the careful treatment of the drift $h(\tilde{v})$. The classical formula (28) requires one single constant noise variance σ^2. Depending on whether σ_0 (light brown in Figure 9) or σ_M (dark brown in Figure 9) is used, the noise is either over- or under-estimated, causing the $MTTF$ to be either under- or over-estimated compared to transient noise simulations. Interestingly, we can observe in Figure 9 that the gap between the σ_0- and σ_M-based predictions widens for the case of more moderate process variations, i.e. the variations of $\sigma^2(\tilde{v})$ with the state of the SRAM bitcell becomes more significant. This is because the steady states (X_0, Y_0) and (X_M, Y_M) becomes further apart (larger $\Delta\tilde{v}$ range), so that the influence of the change in bias point from 0 to M gets substantial.

This is why we recommend harnessing the extended law (30) to predict the $MTTF$ in SRAM bitcells with much greater accuracy. It was indeed constructed to remedy these important shortcomings of the former analytical approaches [25]. It achieves, at least for the investigated cases, a remarkable accuracy: compared to SPICE transient noise simulations the error remains below 20 %, which is outstanding for a metric spanning across orders of magnitude. This combines with a major reduction of the CPU time: from three days (Section III) to a few minutes for four $(\delta V_1, \delta V_2)$ cases, accounting for the cheap deterministic SPICE simulations required to calibrate the stochastic model and to apply the closed-form formula for the $MTTF$. We then estimate to two days the time required to estimate the $MTTF$ of all the cases in the orange crown of Figure 2(b). A benchmark with the few previous accelerated simulation frameworks [18], [19], [20], [21] is complicated for two reasons: the authors did not report the CPU time; more importantly, the extracted TTF was never compared to reliable reference values as provided by general SPICE transient noise simulations in our work.

VII. Conclusion and Perspectives

ULV SRAM arrays are essential blocks of ULP systems. The data retention is threatened by two random phenomena affecting the MOS transistors and thus the SRAM bitcell functionality: static process variations and dynamic intrinsic noise. Whereas most simulation and modeling efforts have been rightly focused on variability, we have shown that, because the noise immunity of the bitcells severely degrades with large process variations, dynamic instability comes as a non-negligible additional concern when the typical retention time is long. An efficient variability-aware noise simulation framework, compatible with industrial SPICE tools and compact models, is therefore needed. We have exploited a 2D variability representation, mapping of the situation at time zero, to select cases to be treated in priority for noise analysis.

Because the brute-force approach that would combine Monte-Carlo simulations with transient noise analyses is unaffordable in CPU time, we have developed and presented an hybrid semi-analytical methodology aimed at accurately predicting the $MTTF$, compatible with industrial transistor compact models and simulation tools. The different steps are summarized in Figure 8. SRAM bitcells in retention are autonomous stochastic nonlinear dynamical systems, for which stochastic model faithful to its strongly nonlinear behavior accounting for varying white noise intensity can be fully characterized from a limited number of fast conventional SPICE simulations. This allows to finally apply the extended Eyring-Kramers formula to estimate the $MTTF$ with a relative error 20 %, remarkably.

As a perspective, we should mention that, toward comprehensive reliability assessment, the contribution of the low-frequency noise (which includes both $1/f$ noise and random telegraph noise [46], [47]), in addition to process variations and thermal noise, remains to be investigated. The $1/f$ noise, yet present in our transient noise simulations, was negligible in the bandwidth of interest. The $1/f$ corner can indeed be discretely noticed near 1 kHz in Figures 6(a) and 6(b). This is because, in this work, the PSDs of the noise are evaluated in nominal conditions, i.e. ignoring the device-to-device variability of the statistics of the low-frequency noise itself [46], [47]. The effect of the *average* $1/f$ turns out to be negligible compared to the thermal noise in the simulated noise bandwidth. This conclusion is very likely to be revised by considering the variations of the $1/f$ (e.g. at 3 or even 6σ as made in [47] when studying analog circuits like comparators and oscillators). Thus, in a longer-term perspective, the stochastic model and analytical prediction tools will have to be enriched with colored and non-Gaussian noise sources. By providing a quantitative answer to the timely scientific question of whether the intrinsic noise sources substantially inflate the SRAM cell failure probability, optimal ULV design margins (i.e. lowering the supply voltage below the 300 mV limits reported in [2]) can be achieved.

References

[1] D. Bol, G. de Streel, and D. Flandre, "Can we connect trillions of IoT sensors in a sustainable way? A technology/circuit perspective (Invited)," in *Proc. IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. (S S)*, Oct. 2015, pp. 1–3.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.

[2] M. Alioto, "Ultra-low power VLSI circuit design demystified and explained: A tutorial," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 1, pp. 3–29, Jan. 2012.

[3] A. P. Chandrakasan et al., "Technologies for ultradynamic voltage scaling," *Proc. IEEE*, vol. 98, no. 2, pp. 191–214, Feb. 2010.

[4] 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. Solid-State Circuits*, vol. 56, no. 7, pp. 2256–2269, Jul. 2021.

[5] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-threshold computing: Reclaiming Moore's law through energy efficient integrated circuits," *Proc. IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.

[6] L. Van Brandt, R. Saeidi, D. Bol, and D. Flandre, "Accurate and insightful closed-form prediction of subthreshold SRAM hold failure rate," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 7, pp. 2767–2780, Jul. 2022.

[7] L. Van Brandt, F. Silveira, J.-C. Delvenne, and D. Flandre, "On noise-induced transient bit flips in subthreshold SRAM," *Solid-State Electron.*, vol. 208, Oct. 2023, Art. no. 108715.

[8] L. V. Brandt, J. Delvenne, and D. Flandre, "Variability-aware noise-induced dynamic instability of ultra-low-voltage SRAM bitcells," in *Proc. IEEE LASCAS*, Punta del Este, Uruguay, Feb. 2024, pp. 1–5.

[9] L. Van Brandt, D. Flandre, and J.-C. Delvenne, "Stochastic nonlinear dynamical modelling of SRAM bitcells in retention mode," in *Proc. IEEE EDTM*, Bengaluru, India, Mar. 2024, pp. 1–3.

[10] D. Bol, R. Ambroise, D. Flandre, and J.-D. Legat, "Interests and limitations of technology scaling for subthreshold logic," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 10, pp. 1508–1519, Oct. 2009.

[11] N. Zheng and P. Mazumder, "Modeling and mitigation of static noise margin variation in subthreshold SRAM cells," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 10, pp. 2726–2736, Oct. 2017.

[12] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, "A high-density subthreshold SRAM with data-independent bitline leakage and virtual ground replica scheme," in *Proc. IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2007, pp. 330–606.

[13] B. H. Calhoun and A. P. Chandrakasan, "A 256-kb 65-nm sub-threshold SRAM design for ultra-low-voltage operation," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

[14] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV robust Schmitt trigger based subthreshold SRAM," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2303–2313, Oct. 2007.

[15] I. J. Chang, J.-J. Kim, S. P. Park, and K. Roy, "A 32 kb 10T sub-threshold SRAM array with bit-interleaving and differential read scheme in 90 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 2, pp. 650–658, Feb. 2009.

[16] B. Vanhoof and W. Dehaene, "SRAM with stability monitoring and body bias tuning for biomedical applications," *IEEE Solid-State Circuits Lett.*, vol. 5, pp. 29–32, 2022.

[17] B. Vanhoof and W. Dehaene, "A 1 MHz 256 kb ultra low power memory macro for biomedical recording applications in 22 nm FD-SOI using FECC to enable data retention down to 170 mV supply voltage," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 71, no. 1, pp. 299–305, Jan. 2024.

[18] E. Rezaei, M. Donato, W. R. Patterson, A. Zaslavsky, and R. I. Bahar, "Fundamental thermal limits on data retention in low-voltage CMOS latches and SRAM," *IEEE Trans. Device Mater. Rel.*, vol. 20, no. 3, pp. 488–497, Sep. 2020.

[19] M. Donato, R. I. Bahar, W. R. Patterson, and A. Zaslavsky, "A sub-threshold noise transient simulator based on integrated random telegraph and thermal noise modeling," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 3, pp. 643–656, Mar. 2018.

[20] P. Jannaty, F. C. Sabou, R. I. Bahar, J. Mundy, W. R. Patterson, and A. Zaslavsky, "Full two-dimensional Markov chain analysis of thermal soft errors in subthreshold nanoscale CMOS devices," *IEEE Trans. Device Mater. Rel.*, vol. 11, no. 1, pp. 50–59, Mar. 2011.

[21] P. Jannaty et al., "Shot-noise-induced failure in nanoscale flip-flops—Part I: Numerical framework," *IEEE Trans. Electron Devices*, vol. 59, no. 3, pp. 800–806, Jan. 2012.

[22] N. Freitas, K. Proesmans, and M. Esposito, "Reliability and entropy production in nonequilibrium electronic memories," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 105, no. 3, Mar. 2022, Art. no. 034107.

[23] L. B. Kish, "End of Moore's law: Thermal (noise) death of integration in micro and nano electronics," *Phys. Lett. A*, vol. 305, nos. 3–4, pp. 144–149, Dec. 2002.

[24] F. Veirano, F. Silveira, and L. Naviner, "Minimum operating voltage due to intrinsic noise in subthreshold digital logic in nanoscale CMOS," *J. Low Power Electron.*, vol. 12, no. 1, pp. 74–81, Mar. 2016.

[25] L. Van Brandt and J.-C. Delvenne, "Predicting state transitions in autonomous nonlinear bistable systems with hidden stochasticity," *IEEE Control Syst. Lett.*, vol. 8, pp. 850–855, 2024.

[26] A. G. Nobile, L. M. Ricciardi, and L. Sacerdote, "Exponential trends of Ornstein–Uhlenbeck first-passage-time densities," *J. Appl. Probab.*, vol. 22, no. 2, pp. 360–369, Jun. 1985.

[27] M. Alioto, "Understanding DC behavior of subthreshold CMOS logic through closed-form analysis," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 7, pp. 1597–1607, Jul. 2010.

[28] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SSC-22, no. 5, pp. 748–754, Oct. 1987.

[29] J. Lohstroh, "Static and dynamic noise margins of logic circuits," *IEEE J. Solid-State Circuits*, vol. SSC-14, no. 3, pp. 591–598, Jun. 1979.

[30] P. Bolcato and R. Poujois, "A new approach for noise simulation in transient analysis," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, vol. 2, May 1992, pp. 887–890.

[31] B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky, "Analytical modeling of SRAM dynamic stability," in *Proc. IEEE/ACM Int. Conf. Comput. Aided Design*, Nov. 2006, pp. 315–322.

[32] W. Dong, P. Li, and G. M. Huang, "SRAM dynamic stability: Theory, variability and analysis," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2008, pp. 378–385.

[33] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill Series in Electrical Engineering). New York, NY, USA: McGraw-Hill, 1991.

[34] L. Van Brandt and J.-C. Delvenne, "Noise–dissipation relation for non-linear electronic circuits," *Appl. Phys. Lett.*, vol. 122, no. 26, Jun. 2023, Art. no. 263507.

[35] A. Demir, E. W. Y. Liu, and A. L. Sangiovanni-Vincentelli, "Time-domain non-Monte Carlo noise simulation for nonlinear dynamic circuits with arbitrary excitations," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 15, no. 5, pp. 493–505, May 1996.

[36] M. Bonnin, F. L. Traversa, and F. Bonani, "Colored noise in oscillators. Phase-amplitude analysis and a method to avoid the Itô-Stratonovich dilemma," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 10, pp. 3917–3927, Oct. 2019.

[37] N. G. Van Kampen, "Itô versus Stratonovich," *J. Stat. Phys.*, vol. 24, pp. 175–187, Jan. 1981.

[38] R. Winkler, "Stochastic differential algebraic equations of index 1 and applications in circuit simulation," *J. Comput. Appl. Math.*, vol. 163, no. 2, pp. 435–463, Feb. 2004.

[39] N. Berglund, "Kramers' law: Validity, derivations and generalisations," 2011, *arXiv:1106.5799*.

[40] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Field*, vol. 42. Cham, Switzerland: Springer, 2013.

[41] N. J. Kasdin, "Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation," *Proc. IEEE*, vol. 83, no. 5, pp. 802–827, May 1995.

[42] D. Krapf, E. Marinari, R. Metzler, G. Oshanin, X. Xu, and A. Squarcini, "Power spectral density of a single Brownian trajectory: What one can and cannot learn from it," *New J. Phys.*, vol. 20, no. 2, Feb. 2018, Art. no. 023029.

[43] A. A. Abidi, "Phase noise and jitter in CMOS ring oscillators," *IEEE J. Solid-State Circuits*, vol. 41, no. 8, pp. 1803–1816, Aug. 2006.

[44] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Rev.*, vol. 43, no. 3, pp. 525–546, Jan. 2001.

[45] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, Apr. 1940.

[46] G. Wirth, "Time-dependent random threshold voltage variation due to random telegraph noise," *IEEE Trans. Electron Devices*, vol. 68, no. 1, pp. 17–23, Jan. 2021.

[47] M. B. da Silva, G. I. Wirth, H. P. Tuinhout, A. Zegers-van Duijnhoven, and A. J. Scholten, "Random telegraph noise in analog CMOS circuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 70, no. 6, pp. 2229–2242, Jun. 2023.

Authorized licensed use limited to: Univ Catholique de Louvain/UCL. Downloaded on March 10,2025 at 09:15:44 UTC from IEEE Xplore. Restrictions apply.