Title

Understanding SRAM stability via bifurcation analysis: Analytical models and scaling trends

Permalink

https://escholarship.org/uc/item/2dh05299

Authors

Ho, Yenpo
Huang, Garng M
Li, Peng

Publication Date

2014

Understanding SRAM Stability via Bifurcation Analysis: Analytical Models and Scaling Trends

YENPO HO, GARNG M. HUANG, and PENG LI, Texas A&M University

In the past decades, aggressive scaling of transistor feature size has been a primary force driving higher Static Random Access Memory (SRAM) integration density. Due to technology scaling, nanometer SRAM designs become increasingly vulnerable to stability challenges. The traditional way of analyzing stability is through the use of Static Noise Margins (SNMs). SNMs are not capable of capturing the key nonlinear dynamics associated with memory operations, leading to imprecise characterization of stability. This work rigorously develops dynamic stability concepts and, more importantly, captures them in physically based analytical models. By leveraging nonlinear stability theory, we develop analytical models that characterize the minimum required amplitude and duration of injected current noises that can flip the SRAM state. These models, which are parameterized in key design, technology, and operating condition parameters, provide important design insights and offer a basis for predicting scaling trends of SRAM dynamic stability.

Categories and Subject Descriptors: B.7.2 [**Integrated Circuits**]: Design Aid

General Terms: Design, Theory, Verification

Additional Key Words and Phrases: Analytical model, critical current, critical time, dynamic noise margin, static noise margin, static random access memory, voltage transfer curves

ACM Reference Format:
Yenpo Ho, Garng M. Huang, and Peng Li. 2014. Understanding SRAM stability via bifurcation analysis: Analytical models and scaling trends. ACM Trans. Des. Autom. Electron. Syst. 19, 4, Article 41 (August 2014), 25 pages.
DOI: http://dx.doi.org/10.1145/2647957

1. INTRODUCTION

SRAM provides indispensable on-chip data storage for an extremely wide variety of electronic applications including microprocessor, ASICs, FPGAs, and DSPs. In today's chip designs, the silicon area occupied by SRAM-based caches dominates over other logic devices, which may constitute more than 70% of chip area. In the past decades, aggressive scaling of transistor feature size has been a primary force driving higher SRAM integration density [Edenfeld et al. 2004; Angelov and Hristov 2004]. On the other hand, the supply voltage is scaled down to meet device reliability constraints and to reduce power consumption. However, the stability margin of SRAM has been significantly degraded by such aggressive scaling. As a result, nanometer SRAM designs are getting increasingly susceptible to various noise problems and there is a growing concern on readability and writeability. Increasing process variation also has a dramatic impact on the stability of highly scaled SRAM designs.

Beginning with Section 2, we first start with the background on SRAM operations and stability issues. In Section 3, we discuss the bifurcation study to demonstrate the SRAM stability issues. We show that three equilibria are located in three different regions.

This material is based upon work supported by the National Science Foundation under grant no. 0917204.
Authors' addresses: Y. Ho (corresponding author), G. M. Huang, and P. Li, Texas A&M University, College Station, TX 77843; email: ypho@neo.tamu.edu.

DOI: http://dx.doi.org/10.1145/2647957

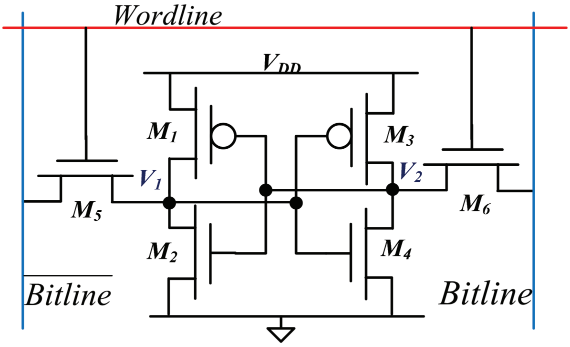


Fig. 1. A 6-T SRAM cell.

Then we show the equilibria are two stable equilibria and a saddle (or metastable point). From there, we show that the saddle-node bifurcation will happen at a certain injected current magnitude called critical current or I_C. From the phase portrait analysis, when injected current amplitude reaches I_C, we observed that two equilibria collide and result in a saddle-node bifurcation. The collision location is called the bifurcation point. When this happens, the two colliding equilibria disappear and only the other remaining stable equilibrium point will survive. The cell state will traverse to that equilibrium point and causes state flip.

In Section 4, we introduce the SRAM circuit and its corresponding nonlinear differential equations based on the Shichman-Hodges model. Next, in Section 5, we introduce region analysis to derive the stability margin analytically for an SRAM. We partition the state space into regions. The equilibrium point locations in terms of a noise injection and system parameters are derived. Furthermore, we focus on the region of bifurcation; we analytically derive the bifurcation point and I_C. However, the outcome of the analytical solution on the bifurcation point and I_C is very complicated. For that, we elaborate on the numerical property and propose a new method to derive an analytical solution for I_C that greatly simplifies the equation but keeps the accuracy.

In Section 6, we further derive the analytical formula for critical time (T_C). We show that a perturbed transient state trajectory will pass the stability boundary (called separatrix) resulting in the state flip when the injected current has higher magnitude than I_C. For a perfectly symmetric SRAM, the stability boundary is a 45° line that passes through the origin. However, the injected current greater than I_C does not necessarily imply that the cell will flip its state [Dong et al. 2008; Zhang et al. 2010]. The current must be greater than I_c for a certain period of time (defined as critical time or T_C) to cross the separatrix. Once the state of the cell crosses the separatrix, the state will flip and even the noise disappears. However, it is still not clear how the SRAM parameters physically influence the phenomena observed from phase portrait analysis. Accordingly, we resort to analytical form solutions to find the relations. Lastly, in Section 7, we investigate the I_C and T_C dependency on technology parameters for design insights.

2. BACKGROUND

Consider the widely adopted six-transistor SRAM cell design shown in Figure 1. In this section, focusing on the 6-T SRAM cell, we describe the traditional static SRAM margins and show their limitations by discussing the dynamic nature of the standby, read, and write operations of the 6-T cell.

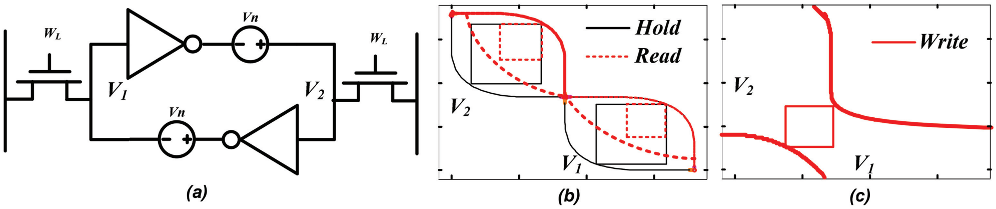


Fig. 2. (a) Characterization of the tranditional SNMs; (b) SNM in standby; (c) SNM in write.

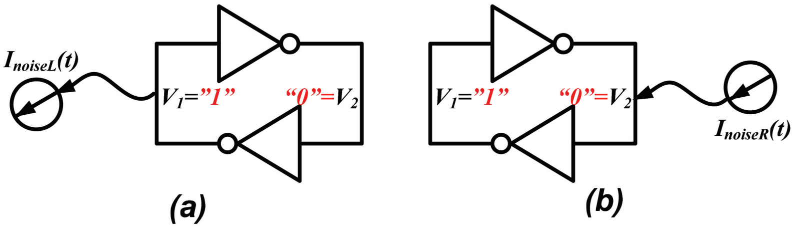


Fig. 3. An SRAM state flip caused by: (a) a current going away; (b) a current injection in standby mode.

2.1. Traditional Static Noise Margin (SNM)

The traditional static noise margin analysis characterizes the robustness of an SRAM cell by using two voltage sources as shown in Figure 2(a). Conventional SNMs measure the largest differential voltage noise that can be tolerated at the two storage nodes [Seevinck et al. 1987; List 1986]. In standby, as shown in Figure 2(b), the SNM is determined as the side of largest square that can be inscribed between the mirrored DC voltage transfer curves (VTCs) of the cross-coupled inverters. The SNM in read can be defined similarly by including the two access transistors as part of the inverter pair VTCs. The SNM in read represents the largest DC voltage perturbation that can be tolerated without a state flip. During write, the SNM is found by inscribing the largest square in between the two VTCs as shown in Figure 2(c).

An SNM metric describes the maximum voltage (or current) perturbation the SRAM circuit can tolerate without resulting in a state flip. However, such a measure is intrinsically unable to characterize the dynamic process that leads to state flips, which is critical for understanding the complete stability picture. In this article, stability will be defined by examining both the magnitude and duration of the injected current noise required to flip the SRAM state. As such, our new stability margin concepts fundamentally capture the temporal aspects of the state flip and provide immediate design insights for enhancing dynamic stability.

Clearly, as SNMs are characterized by finding the largest static voltage noise that can be tolerated in standby, read, or write, they are not positioned in capturing the essential dynamic properties of these operations, as further discussed in the following sections.

2.2. Noise Injection in the Standby Mode

In the standby mode, state flips may occur if certain coupling noise, in the form of a noisy current, strikes one of the bit-lines. This noise injection process is illustrated in Figure 3, where it is assumed that nodal voltages V_1 and V_2 correspond to logic "1" and "0", respectively. The same process has been analyzed to study the SRAM's immunity to single even upsets [Garg et al. 2008, 2006; May and Woods 1979; Pickle and Blandford 1981; Massengill et al. 1993].

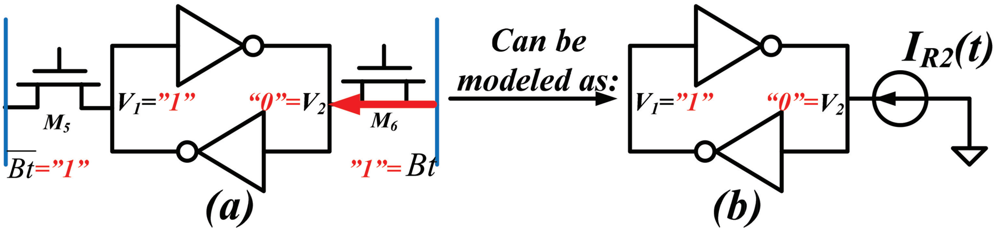


Fig. 4. (a) Noise injection during a read operation; (b) its equivalent model.

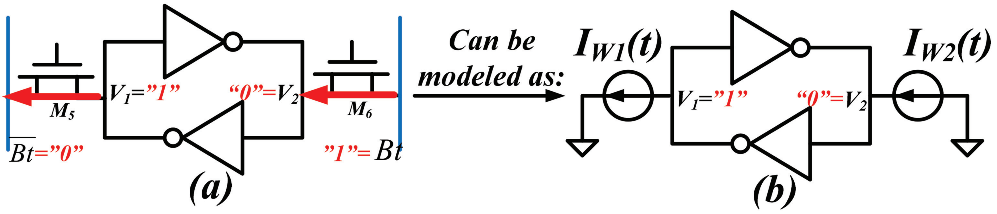


Fig. 5. (a) Noise injection during a write operation; (b) its equivalent model.

Generally, the following scenarios can cause the SRAM state flip: a noise current going away from the high-voltage node (Figure 3(a)), a noise current going into a low-voltage node (Figure 3(b)), or both. For the purpose of characterization, the noise currents may be described by a simple square pulse with certain amplitude (In) and duration (Tn),

$$I_{noiseL,noiseR}(t) = \begin{cases} I_n & 0 \le t \le T_n \\ 0 & otherwise \end{cases}. \tag{1}$$

Naturally, to characterize the noise immunity of the cell, one may seek to find the minimum noise current amplitude and/or duration that renders a state flip. Clearly, the traditional SNM is unable to capture the dynamic characteristics of this process.

2.3. Current Injection in Read/Write Mode

Consider the read operation shown in Figure 4(a). In this case, pass transistor M6 is on and channels a charging current into node V_2. In other words, the NMOS transistor in the inverter at the bottom pulls the bit-line voltage down. The resulting voltage difference between the two bit-lines may be detected by a sense amplifier in order to read out the stored value.

From a stability point of view, it can be seen that the access transistor on the right injects a current into one of the critical storage nodes of the SRAM cell, as illustrated in Figure 4(b). It is the magnitude and duration of this injected current that determine the stability of the read operation. The successful or stable read occurs if the current passing through M6 does not flip the state.

On the other hand, the write operation, shown in Figure 5(a), injects current into the SRAM cell at one side and extracts current at the other. The simulated pass transistor currents are shown in Figure 5(b), where the currents through access transistors M5 and M6 are denoted as I_5 and I_6, respectively.

The foregoing discussion clearly demonstrates the dynamic nature of basic SRAM operations. The stability in each operation mode critically depends on the way noisy currents are injected into the cell. To derive a measure of dynamic stability, it is instructive to examine the stability of the cell under stereotyped noise current injections, that is, idealistic current pulses with a fixed amplitude (I_{R2}, I_{W1}, I_{W2}) and finite duration

(T_{R2}, T_{W1}, T_{W2}), namely,

$$I_{R2,W1,W2}(t) = \begin{cases} I_{R2,W1,W2} & 0 \leq t \leq T_{R2,W1,W2} \\ 0 & Otherwise \end{cases} . \tag{2}$$

In the rest of the article, we focus on studying the dynamic conditions that lead to a state flip under a single current pulse injection, which corresponds to the situation shown in Figures 3(b) and 4(b). This basic approach may be extended in a straightforward way to cases of multiple noisy current injections.

3. SRAM STABILITY AND BIFURCATION STUDY

As a first step towards our goal, we derive a lower bound for the amplitude of the injected current pulse that flips the state. Considering a current injected into V_2 node as shown in Figure 3(b) and Figure 4(b), there exists a threshold on I_n magnitude that can flip the cell state. This threshold current amplitude is referred to as the *critical current* (I_C). As shown later, the significance of I_C is that any injected current pulse with amplitude less than I_C would not be able to flip the state, regardless of its duration. On the other hand, for a current pulse with the amplitude above I_C, the occurrence of a state flip is conditional on the duration of the current pulse. The critical current I_C is derived based on the concept of *saddle-node bifurcation* from the nonlinear system theory [Khalil 2002].

Generally, an SRAM cell has three equilibrium points. Among the three equilibria, two are stable and one is unstable, which is called the *saddle*. Consider again Figure 4(b), assuming a DC current (I_{n2}) is injected to the V_2 node as shown in (2), varying the magnitude of the injected DC current I_{n2} will change the SRAM equilibria. To see how these equilibria will change as functions of I_{n2}, we show a sequence of butterfly curves on the 2D space defined by the voltages V_1 and V_2 obtained from transistor-level circuit simulation for the SRAM cell in Figure 6(a). The equilibria correspond to the intersections between the voltage transfer curves of the back-to-back connected inverter pairs. At $I_{n2} = 0$, the equilibria are labeled as “1”, “2”, and “3”. Among these, “1” and “3” are stable equilibria and “2” is the unstable saddle.

The stability boundary, also called separatrix [Huang et al. 2007; Dong et al. 2008], is important for understanding the transition between two stable equilibria, hence the dynamic stability of the cell. On the 2D state space defined by V_1 and V_2, the separatrix splits the regions of attraction of the two stable equilibria. For the symmetric 6-T cell we study here, the separatrix is the 45^o line passing through the saddle. Starting from any initial state above the separatrix, the SRAM state will eventually go to the stable equilibrium “1”. Similarly, the state will be driven towards the other stable equilibrium “3” if starting from a point below the separatrix.

The preceding discussion assumes that there exist two stable equilibria and hence the separatrix. The dynamic property of the cell will change with injected current. As the magnitude of I_{n2} increases to 150uA, the three equilibra change their location as shown in Figure 6(b). The saddle (marked as “2”) and the stable equilibrium point (marked as “3”) come closer to each other. In Figure 6(c), the saddle collapses with the stable equilibrium. The collapse results in saddle-node bifurcation [Khalil 2002]. The location where the bifurcation happens is called the *bifurcation point*, denoted by (V_{1B}, V_{2B}). In Figure 6(d), the injected current increases to $I_{n2} = 200uA$, yielding only one equilibrium point (marked as “1”) in the entire state space. Starting from any point in the state space, the SRAM state will eventually go to this remaining stable equilibrium.

As shown in Figure 6(c), the occurrence of saddle-node bifurcation marks a critical structural change of the dynamic property of the SRAM cell. When the injected current I_{n2} is above 192uA, there is only one stable equilibrium. When the injected current I_{n2}

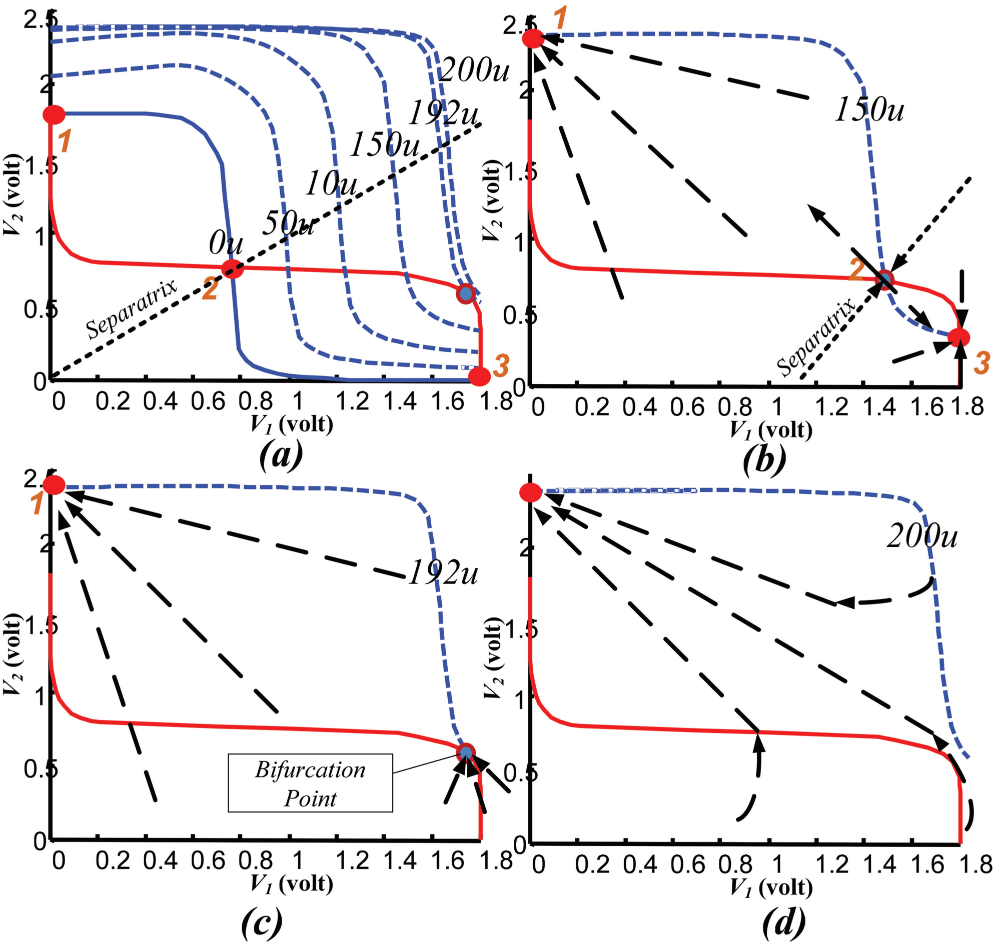


Fig. 6. (a) Illustration of saddle-node bifurcation as I_{n2} increases from zero to 200uA; (b) SRAM butterfly curve when $I_{n2} = 150uA$; (c) SRAM butterfly curve when $I_{n2} = 192uA$; (d) SRAM butterfly curve when $I_{n2} = 200uA$.

is less than 192uA, there are still two stable equilibria although their precise locations may be altered by the current. To possibly flip the state, say, in the standby or read operation, the injected DC (constant) current must be above 192uA such that the starting stable equilibrium collapses with the saddle and hence disappears, and then the state is attracted to the remaining stable equilibrium.

In conclusion, the critical current (I_C) we are seeking would be 192uA in this example. As Figure 6(c) shows, the voltage transfer curves of the two inverters in the cell become tangent to each other at the bifurcation point. Evidently, two curves that are tangent to each other also have the same slope at that tangent point. It can be shown that the Jacobian matrix corresponding to the differential equation of the SRAM cell becomes singular at this point [Lohstroh et al. 1983; Seevinck 1980]. This theoretical result is leveraged to develop a model for I_C in our work.

4. BASIC TRANSISTOR-LEVEL AND DYNAMICAL MODELING FOR SRAMS

Before deriving the proposed models for dynamic stability, we first discuss the basic transistor-level modeling of SRAMs and how a cell can be modeled as a dynamic system.

Table I. Basic Transistor Drain Current Equations

	NMOS	PMOS
Cut-off	$V_{gs} < V_{thn}$ $I_{ds} = 0$	$V_{sg} < \|V_{thp}\|$ $I_{sd} = 0$
Linear	$V_{gs} > V_{thn}$ $V_{ds} < V_{gs} - V_{thn}$ $I_{ds} =$ $K_n(2(V_{gs} - V_{thn})V_{ds} - V_{ds}^2)$	$V_{sg} > \|V_{thp}\|$ $V_{sd} < V_{sg} - \|V_{thp}\|$ $I_{sd} = K_p(2(V_{sg} - \|V_{thp}\|)V_{sd} - V_{sd}^2)$
Saturate	$V_{gs} > V_{thn}$ $V_{ds} > V_{gs} - V_{thn}$ $I_{ds} = K_n(V_{gs} - V_{thn}^2)$	$V_{sg} > \|V_{thp}\|$ $V_{sd} > V_{sg} - \|V_{thp}\|$ $I_{sd} = K_p(V_{sg} - \|V_{thp}\|)^2$

To accurately account for transistor behaviors, sophisticated device models, for example, BSIM3/4 models [Hu 2010; Liu and Hu 1998, 2011; Morshed et al. 2010; Cheng et al. 1997], are usually adopted. These device models, however, make it impossible to derive closed-form design models and prevent development of useful design insights. Instead, in this article, we adopt the popular simple Shichman-Hodges (level-1) transistor models [Nassif 2006; Shichman and Hodges 1968] for developing the targeted dynamic stability models. This choice, nevertheless, allows us to rather accurately predict the scaling trends of dynamic stability as will be shown by the experimental results.

A circuit may be described using a modified nodal analysis formulation in the time domain

$$\dot{Q}(x) = F(x) + u, \tag{3}$$

where $u \in R^N$ is the input, $x \in R^N$ are the state variables, F describes the resistive devices of the circuit, and Q are the capacitive devices of the circuit. For the SRAM cell in Figure 1, for simplicity, we only consider two state variables, voltage (V_1) and its complement (V_2). The circuit equations for the SRAM cell are

$$\begin{cases} C_{11}(V_1, V_2) \cdot \dot{V}_1 + C_{12}(V_1, V_2) \cdot \dot{V}_2 = f_1(V_1, V_2) + u_1 \\ C_{21}(V_1, V_2) \cdot \dot{V}_1 + C_{22}(V_1, V_2) \cdot \dot{V}_2 = f_2(V_1, V_2) + u_2 \end{cases}, \tag{4}$$

where the Cs are the capacitances associated with the two storage nodes, f_1 and f_2 represent the currents of the transistors in the two cross-coupled inverters, and u_1 and u_2 represent additional currents injected to the two storage nodes. We assume the coupling effect between V_1 and V_2 is small, thus C_{12} and C_{21} are neglected. Note that physically C_{11} and C_{22} are mostly contributed by gate and drain parasitic capacitances at V_1 and V_2 nodes. For simplicity, we use circuit simulation to extract averaged small-signal capacitance values C_1 and C_2 by averaging C_{11} and C_{22} over a range of operating points, and finally arrive at

$$\begin{cases} C_1 \cdot \dot{V}_1 = f_1(V_1, V_2) + u_1 \\ C_2 \cdot \dot{V}_2 = f_2(V_1, V_2) + u_2 \end{cases}. \tag{5}$$

In (5), $f_1(\cdot)$ and $f_2(\cdot)$ are determined by the drain currents of the transistors, which are modeled using the level-1 device equations in Table I.

One of the key difficulties in deriving analytical dynamic stability models lies in the fact that different equations are typically used for determining drain currents in the cut-off, linear, and saturation regions. To resolve this problem, we adopt the equivalent Shichman-Hodges representation of the drain currents shown in Table II [Ho 2008]. We further define the S-function

$$S(x) = \begin{cases} 0 & X \leq 0 \\ X & X > 0 \end{cases}. \tag{6}$$

Table II. Shichman-Hodges Representation

	NMOS	PMOS
Cutoff	$I_{ds} = 0$	$I_{sd} = 0$
Linear	$I_{ds} = K_n((V_{gs} - V_{thn})^2$ $-(V_{gd} - V_{thn})^2)$	$I_{sd} = K_p((V_{sg} - \lvert V_{thp}\rvert)^2$ $-(V_{dg} - \lvert V_{thp}\rvert))^2$
Saturate	$I_{ds} = K_n(V_{gs} - V_{thn})^2$	$I_{sd} = K_p(V_{sg} - \lvert V_{thp}\rvert)^2$

Using $S(x)$ and Table II, it is not difficult to derive the following equations for the drain currents for NMOS and PMOS transistors:

$$I_{dsn} = K_n \cdot (S^2(V_{gs} - V_{thn}) - S^2(V_{gd} - V_{thn})), \tag{7}$$

$$I_{sdp} = K_p \cdot (S^2(V_{sg} - |V_{thp}|) - S^2(V_{dg} - |V_{thp}|)). \tag{8}$$

Note (7) and (8) are valid for all regions of operation. This constitutes an important step towards deriving the proposed analytical dynamic stability models. Furthermore, note that the threshold voltage of typical enhancement mode PMOS transistors is negative. For simplicity of presentation, with some abuse of notation, throughout the rest of the article we use a variable such as V_{thp} to indicate the absolute value of the threshold voltage of a PMOS transistor, which is positive. As such, the dynamic equations for SRAM become

$$\begin{aligned} C_1\dot{V}_1 &= K_1[S^2(V_{dd} - V_2 - V_{th1}) - S^2(V_1 - V_2 - V_{th1})] \\ &\quad -K_2[S^2(V_2 - V_{th2}) - S^2(V_2 - V_1 - V_{th2})] - I_{n1}, \end{aligned} \tag{9}$$

$$\begin{aligned} C_2\dot{V}_2 &= K_3[S^2(V_{dd} - V_1 - V_{th3}) - S^2(V_2 - V_1 - V_{th3})] \\ &\quad -K_4[S^2(V_1 -, V_{th4}) - S^2(V_1 - V_2 - V_{th4})] + I_{n2}, \end{aligned} \tag{10}$$

where I_{n1} and I_{n2} represent the injected DC currents. For instance, I_{n2} can be used to model the injected noisy current in Figure 3 or the read current through M_6 in Figure 4(a). The parameters K_1 to K_4 are the MOS device parameters of transistor M_1 to M_4

$$K_{1,2,3,4} = \frac{1}{2}\mu_{n,p} \cdot C_{OX} \cdot W_{1,2,3,4}/L_{1,2,3,4}, \tag{11}$$

where $\mu_{n,p}$ is the carrier mobility (μ_n or μ_p), C_{OX} is the per-unit area gate capacitance, and $W_{1,2,3,4}$ and $L_{1,2,3,4}$ are the effective channel width and length of the transistor, respectively. Eqs. (9) and (10), which are the dynamic equations of the SRAM, have already integrated operation states of each transistor.

The voltage transfer curve, called nullcline, is the set of points satisfied $dV_1/dt = 0$ (V_1-nullcline) in (9) or $dV_2/dt = 0$ (V_2-nullcline) in (10). According to nonlinear theory, equilibrium points are found by solving functions $dV_1/dt = 0$ and $dV_2/dt = 0$. In other words, the points of intersection on the V_1-nullcline and V_2-nullcline are exactly the equilibrium points.

5. ANALYTICAL MODEL FOR THE CRITICAL CURRENT (I_C)

The critical current is highly related to the bifurcation point since it causes equilibra to collapse. That is, the critical current can be found once the bifurcation point is known. In order to have analytical form expression for the bifurcation point and critical current, we introduce *region analysis* [Ho 2008]. Each region in this analysis corresponds to one particular combination of transistor regions of operation (states) (e.g., *M1: Linear; M2: Cut-off; M3: Cut-off; M4: Linear*). Through the region analysis, the transistor states at

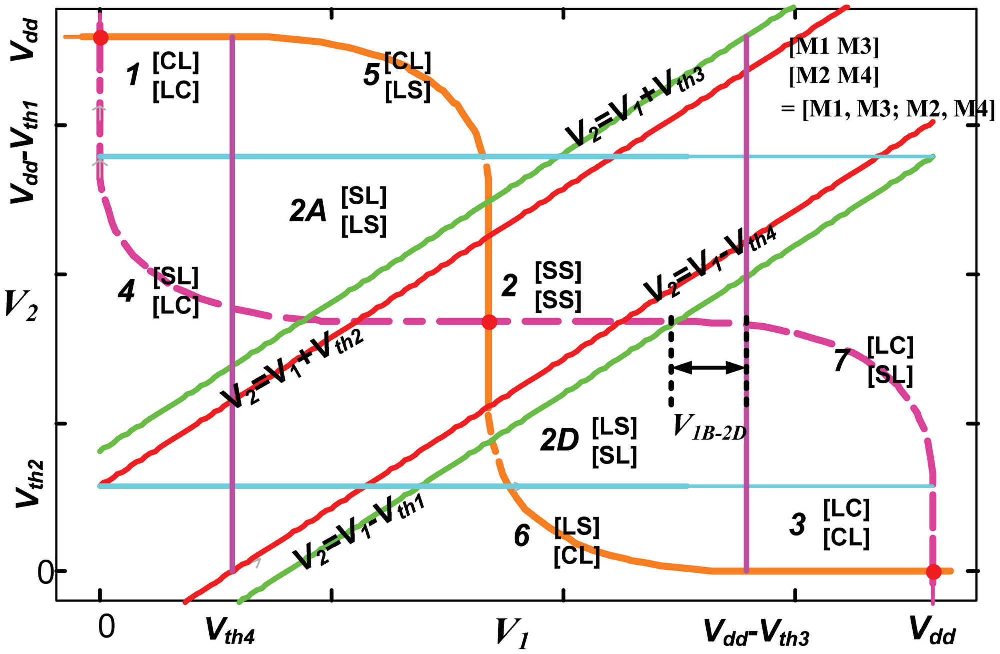


Fig. 7. The nullclines and region formation of an SRAM. The V_{1B-2D} represents all the possilbe ranges represents for V_1 to have bifurcation in region 2D [C = Cut-off; L = Linear; S = Saturation].

the bifurcation can be determined and critical current can therefore be expressed in terms of system parameters.

The V_1 and V_2 voltages physically swing between zero to V_{dd}. This creates a state space. The entire state space can be partitioned into many small disjoint small areas. Each small area is a *region*. The lines separating the state space are based on the transistor threshold voltages. In other words, every region has its corresponding dynamic equation based on (9) and (10), and certain *S(·)* terms are on or off in that particular region. Using region 7 in Figure 7 as an example, the transistor state combination [L,C;S,L] reads *M1 = Linear, M2 = Saturation, M3 = Cut-off*, and *M4 = Linear*. Every point in this region has such a state combination and the corresponding dynamic equations are

$$\begin{cases} C_1\dot{V}_1 = f_1(V_1, V_2) - I_{n1} \\ C_2\dot{V}_2 = f_2(V_1, V_2) + I_{n2}, \end{cases} \tag{12}$$

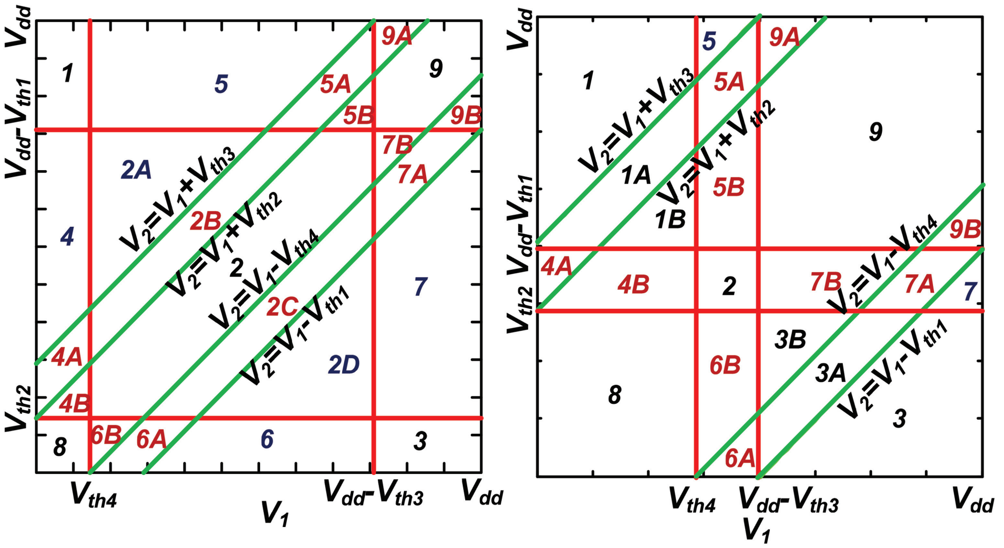
and

$$\begin{cases} f_1(V_1, V_2) = I_{sdp1}{}^{LIN} - I_{dsn2}{}^{SAT} \\ f_2(V_1, V_2) = I_{sdp3}{}^{CUT} - I_{dsn4}{}^{LIN}, \end{cases} \tag{13}$$

where $I_{sdp1}{}^{LIN}$ and $I_{dsn2}{}^{SAT}$ are

$$I_{sdp1}{}^{LIN} = K_1[(V_{dd} - V_2 - V_{th1})^2 - (V_1 - V_2 - V_{th1})^2]; I_{dsn2}{}^{SAT} = K_2(V_2 - V_{th2})^2.$$
$$I_{sdp3}{}^{CUT} = 0; I_{dsn4}{}^{LIN} = K_4[(V_1 - V_{th4})^2 - (V_1 - V_2 - V_{th4})^2] \tag{14}$$

Changing the threshold voltages or V_{dd} would shift the region lines and change the number of regions. As an example shown in Figure 8, the state space would change from (a) to (b) by decreasing V_{dd}. As we can see regions 2A, 2B, 2C, and 2D no longer



(a) *(b)*

Fig. 8. (a) An example of assigned regions for an SRAM; (b) the assigned regions when V_{dd} is reduced.

exist. This means the transistor combinations that correspond to those regions cannot happen under low V_{dd}. Further decrease of V_{dd} can make region 2 disappear. When that happens, the output of one of the inverters will be floating.

The equilibria of an SRAM cell, denoted as (V_{1e}, V_{2e}), are the solutions found by solving $dV_1/dt = 0$ and $dV_2/dt = 0$. When $I_{n1} = I_{n2} = 0$, region 1 and region 3 each have a stable equilibrium strictly at $(V_{dd}, 0)$ and $(0, V_{dd})$, and the saddle can fall onto one of the regions: 2A, 2B, 2, 2C, or 2D. For convenience, we assume that the SRAM cell is symmetric, that is, the two inverters in the cell are identical. For a symmetrical SRAM design, it can be shown that the saddle is located in region 2. Let the location of the saddle denoted by $(V_{1saddle}, V_{2saddle})$ be

$$\left(\frac{\sqrt{K_3}(V_{dd} - V_{th3}) + \sqrt{K_4}V_{th4}}{\sqrt{K_3} + \sqrt{K_4}}, \frac{\sqrt{K_1}(V_{dd} - V_{th1}) + \sqrt{K_2}V_{th2}}{\sqrt{K_1} + \sqrt{K_2}}\right). \tag{15}$$

5.1. The Regions of Bifurcation

The region of bifurcation is the region where bifurcation happens, in other words, the region of bifurcation contains the bifurcation point. The region of bifurcation will be selected from *the candidate regions for bifurcation* and the candidate regions can be mathematically determined by the numbers of equilibrium points.

5.1.1. The Candidate Regions for Bifurcation. Not all the regions in the phase portrait can happen to have bifurcation. Some regions can never have a bifurcation point for all possible parameter sets. The candidate regions for bifurcation are those regions that have bifurcation, and only one region in the candidate regions is the region of bifurcation.

Every region can be classified as having 2, 1, or 0 equilibrium points (e.p.). Those regions can only have 0 e.p. and the mathematical equations in that region cannot have any equilibrium solutions. For those regions that have 1 e.p., there would be mostly one equilibrium solution that can satisfy the region equations. The equilibrium

Table III. Summary of Regions of Bifurcation

	INJECTION CONDITION	THE CANDIDATE REGIONS FOR BIFURCATION	THE REGION OF BIFURCATION BASED ON SYMMETRICAL DESIGN
Single-Side	$\mathbf{I}_{N1} = \mathbf{0}, \mathbf{I}_{N2} > \mathbf{0}$	**2D and 7**	**7**
	$I_{N1} = 0, I_{N2} < 0$	*2A and 4*	4
	$I_{N1} > 0, I_{N2} = 0$	*2D and 6*	6
	$I_{N1} < 0, I_{N2} = 0$	*2A and 5*	5
Double-Side	$I_{N1} > 0, I_{N2} > 0$	*2D, 3, 6, and 7*	*2D*
	$I_{N1} < 0, I_{N2} < 0$	*1, 2A, 4 and 5*	*2A*
	$I_{N1} < 0, I_{N2} > 0$	*2A, 2D, 5 and 7*	*5* ($\lvert I_{N1}\rvert > \lvert I_{N2}\rvert$), *7* ($\lvert I_{N1}\rvert < \lvert I_{N2}\rvert$)
	$I_{N1} > 0, I_{N2} < 0$	*2A, 2D, 4 and 6*	*4* ($\lvert I_{N1}\rvert < \lvert I_{N2}\rvert$), *6* ($\lvert I_{N1}\rvert > \lvert I_{N2}\rvert$)

solution can therefore be symbolically examined. Those regions can have more than one equilibrium point and are the regions of bifurcation; they are in the category of having 2 equilibrium points. The equation complexity reaches forth-order polynomial form for the regions of bifurcation.

Table III summarizes the region of bifurcation where Figure 4(b) illustrates one-sided current injection and Figure 5(b) illustrates the double-sided current injection scenario. Based on the result from Table III, we see that the region of bifurcation would not happen in the "*strap-regions*", meaning regions 2, 2B, 2C, 4A, 4B, 5A, 5B, 6A, 6B, 7A, 7B, 8, 9, 9A, and 9B combined in Figure 8.

5.1.2. Choose the Region of Bifurcation from the Candidate Regions. The region analysis eliminates all the impossible regions for bifurcation. However, it does not give the specific one region of bifurcation. Judgment based on transistor knowledge needs to be made to pick the region of bifurcation from the candidate regions. Using the first case as an example, we select region 7 instead of region 2D. Since transistor M3 can only conduct negligible drain current in region 2D, we assume it is in cut-off. By selecting region 7 as the region of bifurcation, we are taking the chance that M3 is in cut-off when bifurcation happens. In addition, this assumption is valid because the bifurcation point is likely to be at the curviest point of the transfer curve and the curviest point is usually in region 7. Here we complete the column for the region of bifurcation assuming the SRAM is symmetrically designed.

Due to the limited space, we will mainly focus on SRAM stability under the case of ($I_{n1} = 0$ and $I_{n2} > 0$); the other cases can be followed in a similar manner.

5.2. The Analytical Formula for Critical Current

The analytical expression of critical current (I_C) involves solving for the bifurcation point in the region of bifurcation. Let the notation f and g be: $f = dV_1/dt$ and $g = dV_2/dt$. As illustrated in Section 3, the system instability happens when the equilibria collapse. It is proven that the Jacobian matrix of (12) becomes a singular matrix at bifurcation point. [Loshtroh et al. 1983] Therefore, the following formulae can be established.

$$\begin{cases} f = f_1(V_1, V_2) - I_{n1} = 0 \\ g = f_2(V_1, V_2) + I_{n2} = 0 \end{cases} \tag{16}$$

and

$$h = (\partial f/\partial V_1) \cdot (\partial g/\partial V_2) - (\partial f/\partial V_2) \cdot (\partial g/\partial V_1) = 0. \tag{17}$$

Let $(V_{1B}, V_{2B}, I_{C1}, I_{C2})$ be the solution of $(V_1, V_2, I_{n1}, I_{n2})$ that satisfies the previous f- g-, and h-functions, where (V_{1B}, V_{2B}) is the bifurcation point and (I_{C1}, I_{C2}) are the critical currents. In the case of SRAM, there can be many sets, of $(V_{1B}, V_{2B}, I_{C1}, I_{C2})$ for one system parameter, but only one set of (I_{C1}, I_{C2}) will correspond to one bifurcation point

(V_{1B}, V_{2B}) and vice versa. We will demonstrate the simplest case assuming $I_{n1} = 0$ and $I_{n2} > 0$, and the other cases can be followed in a similar manner. The critical current (I_C) means the I_{n2} magnitude level that causes the bifurcation. Then, (16) and (17) become

$$\begin{cases} f(V_{1B}, V_{2B}, I_C) = f_1(V_{1B}, V_{2B}) = 0 \\ g(V_{1B}, V_{2B}, I_C) = f_2(V_{1B}, V_{2B}) + I_C = 0 \\ h(V_{1B}, V_{2B}, I_C) = ((\partial f/\partial V_1) \cdot (\partial g/\partial V_2) - (\partial f/\partial V_2) \cdot (\partial g/\partial V_1))|_{V_{1B}, V_{2B}} = 0 \end{cases}. \tag{18}$$

The problem becomes that of solving three equations for three variables (V_{1B}, V_{2B}, I_C).

Following are the summarized steps to solve (V_{1B}, V_{2B}, I_C):

(1) Determine the transistor states at the bifurcation point.
(2) Formulate continuous f-, g-, and h-functions based on the transistor states from step 1, where f and g the are the differential equations for the region of bifurcation and h is given in (17).
(3) Solve $f = 0$ and $h = 0$ for (V_{1B}, V_{2B}) since f and h are independent of I_C.
(4) Once the analytical form of (V_{1B}, V_{2B}) is known, solve $g = 0$ for I_C.

The preceding steps are applicable to any transistor models including the BSIM4 model. However, obtaining an analytical solution with complex transistor models is quite difficult. Hence, we use the simple L-1 model to demonstrate.

Solving for (V_{1B}, V_{2B}) and I_C symbolically in the L-1 model is involved. For the case in Figure 4(b), $I_{n1} = 0$ and $I_{n2} = I_C$, the simplest analytical formula for I_C without any approximation is

$$I_C = K_4[(V_{1B} - V_{th4})^2 - (V_{1B} - V_{2B} - V_{th4})^2], \tag{19}$$

where V_{1B} and V_{2B} are the bifurcation point and can be expressed as follows.

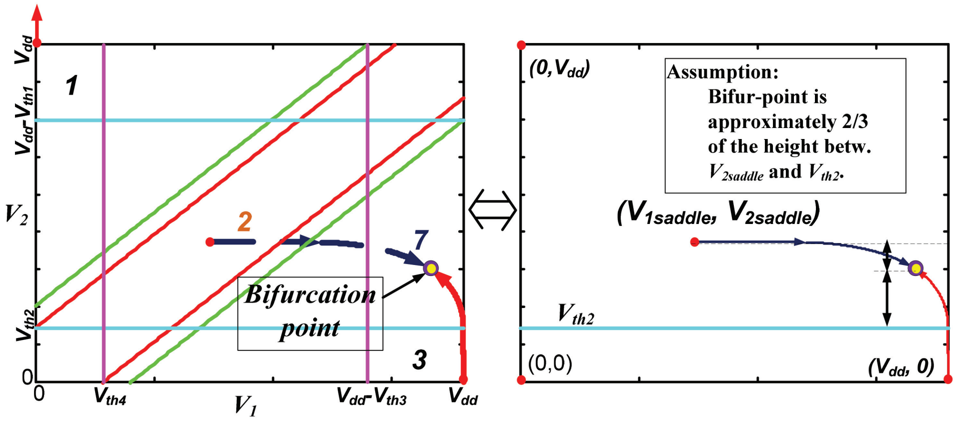
$$V_{1B} = V_{2B} + V_{th1} + \sqrt{(V_{dd} - V_{2B} - V_{th1})^2 - \frac{K_2}{K_1}(V_{2B} - V_{th2})^2} \tag{20}$$

$$\begin{aligned} V_{2B} = {} & \frac{K_1(V_{dd} + V_{1B} - V_{th1} - V_{th4}) - K_2 \cdot V_{th2}}{2(K_1 - K_2)} \\ & - \sqrt{\left(\frac{K_1(V_{dd} + V_{1B} - V_{th1} - V_{th4}) - K_2 \cdot V_{th2}}{2(K_1 - K_2)}\right)^2 - \frac{K_1(V_{1B} - V_{th1})(V_{1B} - V_{th4})}{K_1 - K_2}}. \end{aligned} \tag{21}$$

As can be seen, V_{1B} and V_{2B} are cross-coupled. Solving them would involve a forth-order polynomial, with polynomial roots having more than 10 symbolic terms.

Because the bifurcation point is always found in between $V_{2saddle}$ and V_{th2} as illustrated in Figure 9(a), we simplified the expression for V_{2B} by approximating V_{2B} as a weighted sum of $V_{2saddle}$ and V_{th2} as: $w \cdot (V_{saddle} - V_{th2})$. The weighting factor w is chosen by averaging over more than 30 different parameter settings. It was observed that the weight factor for the exact value of V_{2B} is within 8% of $w = 2/3$ as illustrated in Figure 9(b). With that, we have

$$V_{2B} = w \cdot (V_{2saddle} - V_{th2}) + V_{th2} \quad w = 2/3, \tag{22}$$



(a) *(b)*

Fig. 9. (a) The plot of SRAM equilibrium points as I_{n2} changes. Increasing the magnitude of I_{n2} will make the saddle (in region 2 when $I_{n2} = 0$) collapse with the stable node (in region 3 when $I_{n2} = 0$) and resulting saddle node bifurcation in region 7; (b) illustration showing that the bifurcation point is approximately 2/3 of the height between $V_{2saddle}$ and V_{th2} on the same phase portrait.

where $V_{2saddle}$ is

$$V_{2saddle} = \frac{\sqrt{K_1}(V_{dd} - V_{th1}) + \sqrt{K_2}V_{th2}}{\sqrt{K_1} + \sqrt{K_2}}. \tag{23}$$

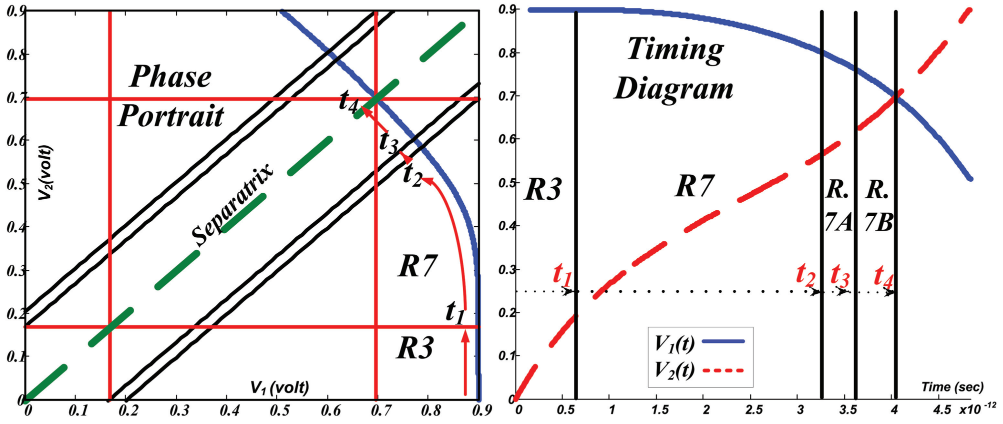
Therefore, the critical current, I_C can be expressed in terms of system parameters by plugging the V_{1B} and V_{2B} expressions given in (20) and (22) into (19).

6. ANALYTICAL MODEL FOR THE CRITICAL TIME

The SRAM cell will flip if the cell state crosses the *stability boundary*. During the operation of the SRAM cell, if a stable state is perturbed across that boundary, a state flipping will result. For a perfectly symmetric SRAM cell, the stability boundary can be simply defined by passing a 45° line through the origin on the phase portrait of the SRAM cell. The stability boundary of the SRAM is also called *separatrix* because the stability boundary separates two stability regions [Ho 2008; Zhang et al. 2006; Wang et al. 2008; Jahinuzzaman et al. 2009; Song et al. 2013]. If the injected noise current is higher than the critical current, the state of the cell will drive from the initial stability and eventually go across the separatrix. The time it takes from the initial state to go into the separatrix is called *critical time* (T_C). After the trajectory across the separatrix, the cell state will fall into the stability region of the other stable equilibrium and result in a state flip.

An example is demonstrated in Figure 10. Assume the SRAM cell is symmetric; the separatrix is the 45° line passing through the origin. In Figure 10(a), the cell state initially starts at (0.9, 0) in region 3 (R3) at time 0. It enters region 7 (R7) at time t_1 and enters region 7A (R7A) at time t_2. The state will eventually reach the separatrix in region 7B (R7B) at time t_4. Once the state passes the separatrix, the state can never be recovered even if the noise injections disappear. The total time taken for a state to reach the separatrix is the critical time, which is t_4 in this case.

Figure 10(b) shows the timing diagram for that cell state. The state transits through many regions to flip the state. The rigorous way to find the critical time is to separately find the time spent in each region then sum each together. However, this results in symbolic expressions that are very cumbersome. The way we simplify the analytical



(a) (b)

Fig. 10. (a) The simulated phase portrait of a 65nm SRAM based on the S-H model. It shows a cell state crosses the separatrix (45° line through the origin) and flips to the other side; (b) the timing diagram of the cell state.

formula is based on the observation that the vector field strength around the bifurcation point is weak so that the trajectory takes up more time in the region of bifurcation. In this regard, it is efficient to focus on the time spent in the region of bifurcation to arrive at a simple but physically meaningful expression for the critical time. In other words, we find the expression of the time spent in the region of bifurcation to be the critical time analytical formula. As demonstrated in Figure 10(b), the trajectory spends most of the time in the region of bifurcation, region 7 (R7), as opposed to any other regions. The analytical formula for T_C is to solve the nonlinear ODE corresponding to the transistor combination in region 7 (R7) that is mentioned in (12).

However, solving the cross-coupled nonlinear ODE in (12) is cumbersome. Mathematically, there is no good technique to directly solve this type of ODE. The way we bypass the nonlinearity is to linearize the ODE at the bifurcation point. In this regard, we preserve the characteristics of the cell state trajectory around the bifurcation point and simplify the complexity of the equation at the same time. By doing that, the system can be modeled using two cross-coupled linear ODEs as shown. we have

$$\begin{cases} C_1 \cdot \dot{V}_1(t) = g_{11}(V_1 - V_{1B}) + g_{12}(V_2 - V_{2B}) - (I_{n1}(t) - f_1(V_{1B}, V_{2B})) \\ C_2 \cdot \dot{V}_2(t) = g_{21}(V_1 - V_{1B}) + g_{22}(V_2 - V_{2B}) + (I_{n2}(t) + f_2(V_{1B}, V_{2B})) \end{cases} \tag{24}$$

or

$$\begin{cases} \dot{V}_1(t) = a_1 \cdot V_1(t) + b_1 \cdot V_2(t) + I_1(t) \\ \dot{V}_2(t) = a_2 \cdot V_1(t) + b_2 \cdot V_2(t) + I_2(t) \end{cases}, \tag{25}$$

where

$$\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} \partial f_1/\partial V_1 & \partial f_1/\partial V_2 \\ \partial f_2/\partial V_1 & \partial f_2/\partial V_2 \end{bmatrix}\Bigg|_{\substack{V1 = V1B \\ V2 = V2B}}, \tag{26}$$

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, \tag{27}$$

$$\begin{bmatrix} I_1(t) \\ I_2(t) \end{bmatrix} = \begin{bmatrix} -(a_1 \cdot V_{1B} + b_1 \cdot V_{2B}) - (I_{n1}(t) - I_{C1})/C_1 \\ -(a_2 \cdot V_{1B} + b_2 \cdot V_{2B}) + (I_{n2}(t) - I_{C2})/C_2 \end{bmatrix}, \tag{28}$$

and

$$\begin{aligned} I_{C1} &= f_1(V_{1B}, V_{2B}) \\ I_{C2} &= -f_2(V_{1B}, V_{2B}). \end{aligned} \tag{29}$$

The coefficients ($g_{11}, g_{12}\ldots$, etc.) are functions of system parameters, and f are the functions from (12). Since the Jacobian matrix (27) is singular at the bifurcation point, the eigenvalues will be 0 and λ, and λ is a negative value. The singular Jacobian matrix means to determine zero,

$$a_1 b_2 - a_2 b_1 = 0. \tag{30}$$

Solving (25) yields the following general solution using the Laplace transform.

$$\begin{cases} V_1(t) = \left(\frac{b_2}{\lambda} + \frac{a_1}{\lambda}e^{\lambda\cdot t}\right)\cdot V_1(0) + \left(-\frac{b_1}{\lambda} + \frac{b_1}{\lambda}e^{\lambda\cdot t}\right)\cdot V_2(0) \\ \qquad + \left(\frac{b_2}{\lambda} + \frac{a_1}{\lambda}e^{\lambda\cdot t}\right) * I_1(t) + \left(-\frac{b_1}{\lambda} + \frac{b_1}{\lambda}e^{\lambda\cdot t}\right) * I_2(t) \\ V_2(t) = \left(-\frac{a_2}{\lambda} + \frac{a_2}{\lambda}e^{\lambda\cdot t}\right)\cdot V_1(0) + \left(\frac{a_1}{\lambda} + \frac{b_2}{\lambda}e^{\lambda\cdot t}\right)\cdot V_2(0) \\ \qquad + \left(-\frac{a_2}{\lambda} + \frac{a_2}{\lambda}e^{\lambda\cdot t}\right) * I_1(t) + \left(\frac{a_1}{\lambda} + \frac{b_2}{\lambda}e^{\lambda\cdot t}\right) * I_2(t) \end{cases}, \tag{31}$$

where

$$\lambda = a_1 + b_2. \tag{32}$$

The $(V_1(0), V_2(0))$ is the initial condition and $(*)$ is the convolution integral. In our case, we treat the injected current as constant. Thus, the expression becomes

$$\begin{cases} V_1(t) = A_{P1} + B_{P1}\cdot e^{\lambda t} + C_{P1}\cdot t \\ V_2(t) = A_{P2} + B_{P2}\cdot e^{\lambda t} + C_{P2}\cdot t \end{cases}, \tag{33}$$

where

$$C_{P1} = -\frac{1}{\lambda}\left(\frac{b_1(I_{n2} - I_{C2})}{C_2} + \frac{b_2(I_{n1} - I_{C1})}{C_1}\right), B_{P1} = \frac{\dot{V}_1(0) - C_{p1}}{\lambda}, A_{P1} = V_1(0) - B_{P1}, \tag{34}$$

$$C_{P2} = \frac{1}{\lambda}\left(\frac{a_1(I_{n2} - I_{C2})}{C_2} + \frac{a_2(I_{n1} - I_{C1})}{C_1}\right), B_{P2} = \frac{\dot{V}_2(0) - C_{p2}}{\lambda}, A_{P2} = V_2(0) - B_{P2}, \tag{35}$$

$(\dot{V}_1(0), \dot{V}_2(0))$ is acquired by evaluating (25) at $t = 0$. The trajectory in (33) will cross the separatrix at

$$V_1(T_C) = V_2(T_C) \tag{36}$$

since the separatrix is a 45° line through the origin.

We assume the exponential terms in (33) become negligible by the time the state trajectory reaches the separatrix due to the exponential decay, so the formula for the critical time T_C is

$$T_C = \frac{A_{P1} - A_{P2}}{C_{P2} - C_{P1}}. \tag{37}$$

And it leads to

$$T_C = \frac{(-\lambda\cdot(V_1(0) - V_2(0)) + (\dot{V}_1(0) - \dot{V}_2(0)) - (C_{p1} - C_{p2}))}{\lambda(C_{p1} - C_{p2})}. \tag{38}$$

We eliminate $(\dot{V}_1(0) - \dot{V}_2(0))$ and $(C_{P1} - C_{P2})$ on the numerator because together they are close to cancelling each other and become insignificant. That simplifies the equation to

$$T_C = \frac{(C_2 \cdot g_{11} + C_1 \cdot g_{22}) \cdot (V_1(0) - V_2(0))}{(g_{11} + g_{12}) \cdot (I_{n2} - I_{C2}) + (g_{21} + g_{22}) \cdot (I_{n1} - I_{C1})}. \tag{39}$$

For the single-sided current injection case, we repeat the process from (24) to (39) without considering the I_{n1} injected current and the result will be the same as dropping the term $(I_{n1} - I_{C1})$. Moreover, the critical time formula is derived as follows after evaluating (39) at the initial condition $(V_1(0) = V_{dd}, V_2(0) = 0)$ as

$$T_C = \frac{(C_2 \cdot g_{11} + C_1 \cdot g_{22}) \cdot V_{dd}}{(g_{11} + g_{12}) \cdot (I_{n2} - I_{C2})} \tag{40}$$

or

$$T_C = \frac{(C_2 \cdot K_1(V_{1B} - V_{2B} - V_{th1}) + C_1 \cdot K_4(V_{1B} - V_{2B} - V_{th4}))}{(K_1(V_{dd} - V_{2B} - V_{th1}) + K_2(V_{2B} - V_{th2}))} \cdot \left(\frac{V_{dd}}{I_{n2} - I_{C2}}\right), \tag{41}$$

where the coefficients ($g_{11}, g_{12} \ldots$, etc.) are acquired from evaluating (26) in the region of bifurcation. If full symmetry is assumed, the capacitances at each internal storage node are the same, namely $C = C_1 = C_{2,}$ and

$$T_C = C \cdot \left(\frac{V_{dd}}{I_{n2} - I_C}\right) \cdot \left(\frac{K_1(V_{1B} - V_{2B} - V_{th1}) + K_4(V_{1B} - V_{2B} - V_{th4})}{K_1(V_{dd} - V_{2B} - V_{th1}) + K_2(V_{2B} - V_{th2})}\right). \tag{42}$$

Furthermore, if taking the exponential terms in (33) into account, the critical time equation can be simpler if we substitute the exponential term ($e^{\lambda t}$) by its Taylor expansion $1 + \lambda t$. The formula becomes

$$T_C = \frac{(A_{P1} - A_{P2}) + (B_{P1} - B_{P2})}{(C_{P2} - C_{P1}) + \lambda \cdot (B_{P2} - B_{P1})} = \frac{V_1(0) - V_2(0)}{\dot{V}_1(0) - \dot{V}_2(0)} = \frac{V_1(0) - V_2(0)}{\frac{I_{n2} - I_{C2}}{C_2} + \frac{I_{n1} - I_{C1}}{C_1}}, \tag{43}$$

and equivalently

$$T_C = C_2 \cdot \frac{V_{dd}}{I_{n2} - I_{C2}} \tag{44}$$

after evaluating (43) at the initial condition $(V_1(0) = V_{dd}, V_2(0) = 0)$ and dropping the $(I_{n1} - I_{C1})$ term. Eq. (44) is a good approximation if the injected current magnitude (I_{n2}) is more than five times its critical current (I_{C2}). If I_{n2} goes beyond eight times of I_{C2}, the formula can be shown as $T_C = C_{\dot{2}} V_{dd}/I_{n2}$, which is the same formula shown in [Zhang et al. 2006].

The following are the summarized steps to solve critical time (T_C):

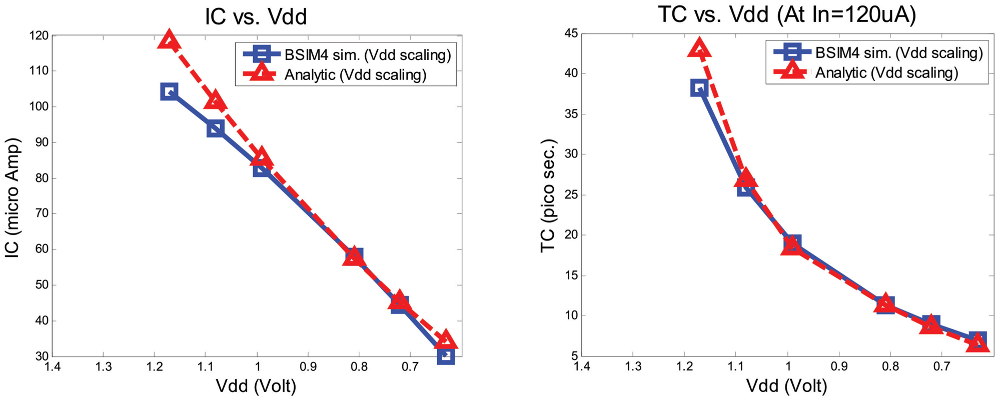
---

(1) Solve the critical current and bifurcation point formula (V_{1B}, V_{2B}, I_C).
(2) Formulate the linearized ODE at the bifurcation point.
(3) Solve the general solution and in particular the solution for the linearized ODE.
(4) Find the critical time T_C by solving $V_1(T_C) = V_2(T_C)$ for symmetrical designs.

In summary, the simplification was made to the dynamic system formulation at the bifurcation point (linearization) to obtain two linear ODEs, from which an analytic solution was found for the critical time (the time from the initial state to the stability boundary).

Table IV. The Parameter Values Used in Shichman-Hodges Model on a 65nm SRAM ($V_{DD} = 0.9$)

	NMOS	PMOS
W	*130nm*	*84.5nm*
K	*2.46e-4* (A/V^2)	*5.33e-5* (A/V^2)
V_{th}	*0.24 (V)*	*0.24 (V)*



(a) ***(b)***

Fig. 11. (a) Critical current (I_C); (b) critical time (T_C) vs. V_{dd}.

7. SIMULATION

The analytical models developed in the previous sections provide insightful understanding of the SRAM dynamic stability. In this section, we use industrial commercial software (Cadence) to investigate the SRAM dependency on design parameters and the scaling trends of the SRAM dynamic stability.

The setup for this section will be comparing our analytical models with SPICE simulation using Cadence Spectre with a 65nm BSIM4 PTM (Predictive Technology Model) [Cao 2012]. We performed least-square fitting with the BSIM4 model data to derive the level-1 device model parameters. Table IV shows the fitted level-1 device model parameter values used for our nominal values.

7.1. Dependency on Design/Technology Parameters

The designers can simply plug in the key design and technology parameters to predict important dynamic properties of a targeted SRAM design. By leveraging these analytical models, we systematically study the dependencies of the critical current and critical time have on design and device parameters on supply voltage and transistor width.

7.1.1. Dependency on Supply Voltage (Vdd). The effect of supply voltage scaling is shown in Figure 11(a). The analytical formula on critical current is shown to have a square dependency on V_{dd}. As can be seen, our analytical model matches well with SPICE simulation.

The scaling of the critical time as a function of V_{dd} is shown in Figure 11(b). Interestingly, in the considered range of supply voltage, T_C shows an approximate quadratic dependency on V_{dd}. From a design perspective, this clearly shows that supply voltage is a critical design knob for controlling SRAM dynamic stability.

7.1.2. Dependencies on Transistor Width (Wn/Wp). The NMOS/PMOS transistor width is embedded in variable Kn/Kp as described in (11). Thus, the change of K reflects the

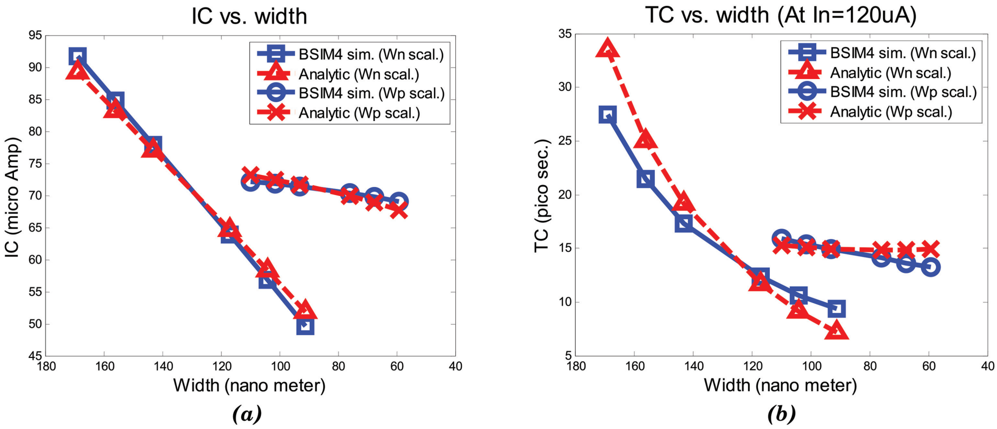


Fig. 12. (a) Critical curent (I_C); (b) critical time (T_C) vs. NMOS/PMOS width scaling.

dependency on the width. The critical current is in the form of drain current (I_{ds}) of the NMOS transistor *M4* in the linear region. In the read operation, the NMOS transistor *M4* plays a role of draining out the current from the access transistor, thus pulling down the bit-line. As such, the injected current may be sufficiently large to offset the pull-down current of *M4* in order to create a state flip. This makes the critical current have high dependency on K_n. Knowing this dependency on K_n also allows us to straightforwardly determine the effects of the parameters on which K_n depends. For example, increasing the NMOS channel width will increase K_n by the same factor. This will proportionally increase the critical current.

As shown in Figure 12(a), the dependency on the PMOS width is rather weak compared to the dependency on NMOS width, and smaller Wp leads to a reduced I_C.

The critical time has a stronger dependency on W_n than on W_p as shown in Figure 12(b). Practically speaking, increasing the channel width of the pull-down NMOS transistors or decreasing V_{thn} will increase the pull-down strength and make T_C longer. As the cell state approaches the stability boundary, the NMOS transistor (such as M4) acts to pull the cell state back and slows down the state flipping process. As a result, a stronger pull-down NMOS transistor will increase the time needed to flip the state.

7.2. Process Variations

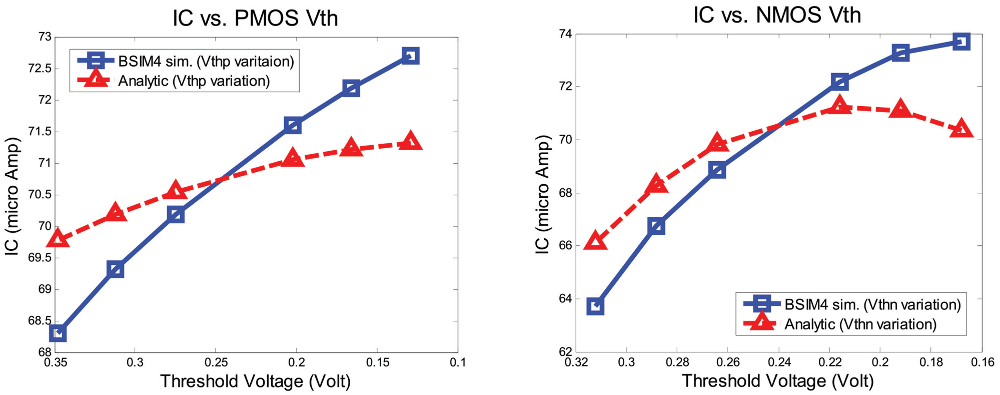
Process variations can be classified into two main categories: intra-die variation (local variation) and inter-die variation (global variation). In this section, we will analyze the impacts of inter-die and intra-die variations and compare the results of the analytical model and BSIM4 model.

7.2.1. Intra-Die Variations (Local Variations). *Intra-die variations (local variations)* are variations within a single chip and can lead to a mismatch across different transistors in the SRAM cell. To analyze intra-die variability, we first label the transistor individually (M1 to M4) according to Figure 1 and apply perturbation to each of the four transistor Vth's and widths. Table V shows some simulation results. For this mismatch analysis, the widths are varied by ±10% and Vths are varied by ±30%.

The first case (no. 1) in Table V is the nominal/symmetrical case. The case no. 2 and no. 3 are the cases in which the transistor widths have been perturbed off from the nominal value. From the data in case nos. 2 and 3, I_C does not get affected by perturbing W_3 in both the analytic and BSIM4 model. As Section 5 mentioned, the

Table V. Few Cases to Demonstrate Inter-Die Variation (VDD = 0.9; TC is evaluated at In = 120μA)

	W_1/	W_3/	V_{th1}/	V_{th3}/	*Analytic*		*BSIM4*	
	W_2	W_4	V_{th2}	V_{th4}	I_C	T_C	I_C	T_C
#	*(nm)*	*(nm)*	*(V)*	*(V)*	*(uA)*	*(ps)*	*(uA)*	*(ps)*
1	*84.5/130*	*84.5/130*	*0.24/0.24*	*0.24/0.24*	*70.46*	*14.15*	*70.94*	*14.48*
2	*92.5/117*	*84.5/143*	*0.24/0.24*	*0.24/0.24*	*78.93*	*18.27*	*79.46*	*18.20*
3	*92.95/117*	*101.4/143*	*0.24/0.24*	*0.24/0.24*	*78.93*	*18.27*	*79.46*	*18.66*
4	*84.5/130*	*84.5/130*	*0.21/0.25*	*0.25/0.21*	*76.89*	*16.26*	*76.44*	*16.33*
5	*92.95/117*	*76.05/143*	*0.21/0.25*	*0.25/0.21*	*86.14*	*21.67*	*85.49*	*20.81*
6	*92.95/117*	*76.05/143*	*0.18/0.3*	*0.3/0.18*	*99.5*	*36.76*	*95.95*	*29.08*
7	*76.05/143*	*92.95/117*	*0.3/0.18*	*0.18/0.3*	*47.21*	*9.24*	*48.26*	*9.87*



(a) ***(b)***

Fig. 13. Critical curent (I_C) vs. (a) PMOS; (b) NMOS threshold voltages.

SRAM reaches instability (or the moment bifurcation happens) with the following transistor combination: (*M1:Linear; M2: Saturate; M3:Cut-off; M4:Linear*). Since the transistor M3 is in cut-off mode, changing M3 parameters (V_{th3} and W_3) does not affect the I_C. Because we simplified the analytic model by ignoring the switching time other than the bifurcation region, there is no change on critical time (T_C) on the analytic model but a small change on BSIM4 critical time is observed.

The case no. 6 shows the best case, meaning that the perturbations give the highest critical current (I_C) out of all combinations of transistor width variation of ±10% and Vths variation of ±30%. If we want I_C to be higher, we need to make M1 and M4 stronger and M2 and M3 weaker. To maintain an SRAM ($V_1 = V_{dd}$, $V_2 = 0$) not flipping its state, we want strong M1 and week M2 to maintain the high voltage at V_1 as well as weak M3 and strong M4 to drain the voltage at V_2. Therefore, the best case would be: size up W_1 and W_4; size down W_2 and W_3; decrease V_{th1} and V_{th4}; increase V_{th2} and V_{th3}. Likewise, to have the worst case wherein perturbations give the lowest critical current (I_C), we must do the best-case sizing in the opposite way. The case no. 7 shows the worst case.

7.2.2. Inter-Die Variations (Global Variations). *Inter-die variations* cause global cross-chip variations. We examine the impacts of inter-die variations by varying NMOS/PMOS threshold voltages (V_{thn} and V_{thp}) and transistor lengths (L_N and L_P) for different transistors in the same way, and compare the results from the analytic model- and BSIM4-based transistor-level simulation. Overall, the analytic model may not perfectly fit the BSIM4-based simulation, but our analytic model shows a reasonable trend for predicting the device variations.

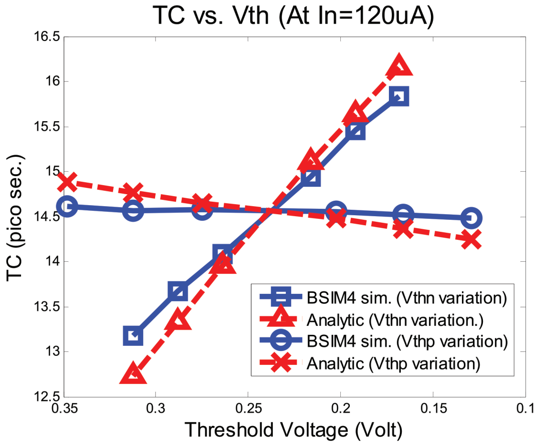


Fig. 14. Critical time (T_C) vs. NMOS/PMOS threshold voltages.

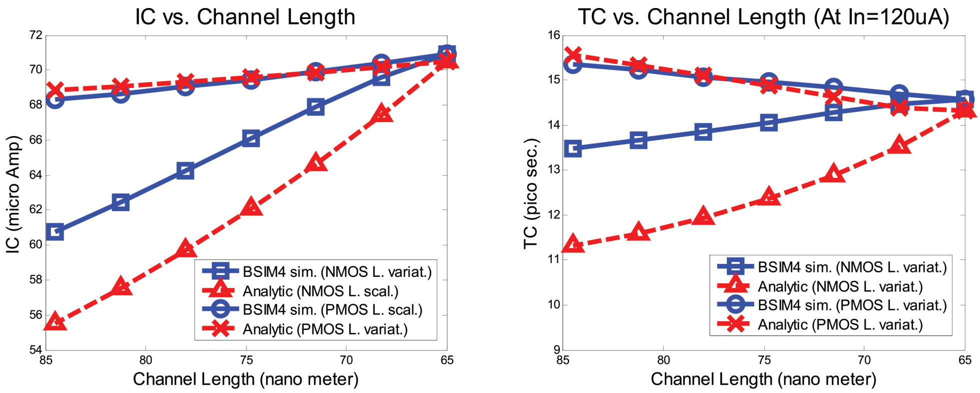
—*Dependencies on Transistor Threshold Voltages.* The dependencies on the PMOS and NMOS threshold voltages are shown in Figure 13(a) and Figure 13(b). As we can see, the critical current I_C has an approximate linear relationship with V_{thp} around the nominal parameter values, but the relationship with V_{thn} is more nonlinear when compared with V_{thp}.

The critical time is more sensitive to the change of V_{thn} than V_{thp}. An interesting phenomenon we observed is that changing V_{thp} does not significantly affect T_C as shown in Figure 14. When V_{thp} is varying ±30%, T_C is changing less than ±1%. Based on the SRAM state trajectory shown in Figure 10(a), the SRAM state transition from initial state to the stability boundary passes the following regions: region 3 → region 7 → region 7A → region 7B, and the corresponding transistor combinations are: R3 (*M1:Linear; M2:Cut-off; M3:Cut-off; M4:Linear*) → R7 (*M1:Linear; M2:Saturate; M3:Cut-off; M4:Linear*) → R7A (*M1:Linear; M2:Saturate; M3:Cut-off; M4:Saturate*) → R7B (*M1:Saturate; M2:Saturate; M3:Cut-off; M4:Saturate*). Notice that transistor M3 is in cut-off mode throughout the transition. Since only V_{th1}, and V_{th3} are PMOS threshold voltages, the change of T_C is caused by $V_{th1,}$, not V_{th3}. From the derived critical time equation in (37), the V_{th1} appears in both the numerator and the denominator. Since other parameters such as K and I_C have low sensitivity to V_{th1}, increasing V_{th1} would not influence T_C too much.

—*Dependencies on the Channel Length.* The dependencies on the NMOS/PMOS channel length are shown in Figure 15(a). The critical current seems to be increased when the transistor lengths decrease. It is also observed that I_C is more sensitive to NMOS channel length variation than on that of PMOS. On the other hand, as shown in Figure 15(b), the variation of NMOS and PMOS channel length seems to have the same influence on critical time.

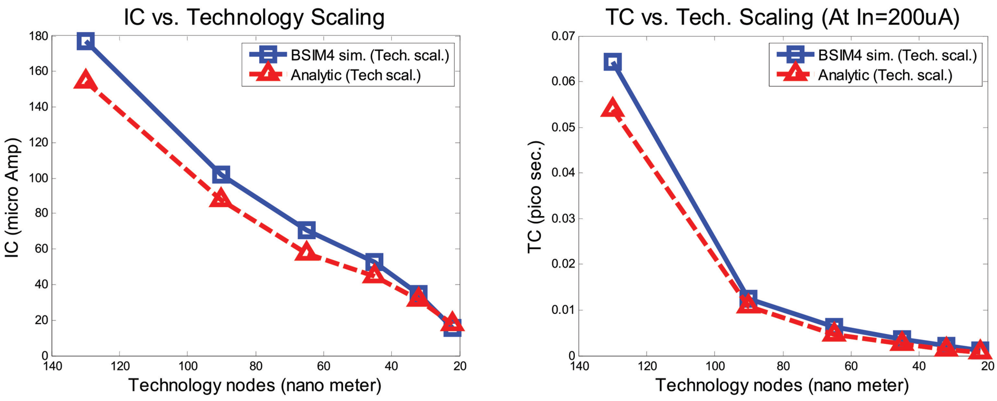
7.3. Technology Trend Scaling

Lastly, we study how the critical current and critical time change with technology scaling. We examined I_C and T_C in a wide range of technology nodes from 130nm down to 22nm by using Predictive Technology Models (PTMs) [Cao 2012], and we plot the results using SPICE simulation against the derived formula as shown in Figure 16. We performed least-square fitting with the BSIM4 model data to derive the level-1 device model parameters; we adopted suggested V_{dd} levels and transistor sizes from Ramesh et al. [2011], Arnaud et al. [2003], Utsumi et al. [2005], Toh et al. [2010], Chang et al.



(a) ***(b)***

Fig. 15. (a) Critical current (I_C); (b) critical time (T_C) vs. NMOS/PMOS channel lengths.



(a) ***(b)***

Fig. 16. (a) Critical current (I_C); (b) critical time (T_C) vs. various technology nodes (from 130nm to 22nm).

[2005], Wang et al. [2009], and Haran et al. [2008]. Table VI summarizes the fitted results. The capacitance C is calculated by adding all the coupling capacitance at the storing node. The K_N, K_P and supply voltage V_{dd} noticeably get smaller as the transistor size reduces from one technology node to the next. Based on the derived analytical formula, we expect a decrease in critical current as the transistor size goes smaller. As we can see from Figure 16, in reference to SPICE simulation using the BSIM4 device models, the derived analytical formulas are able to predict the trends of scaling.

7.4. The Computational Efficiency

We use a transistor-level circuit simulator, in this case Cadence Spectre, to find both I_C and T_C as follows: I_C is found by incrementing the injected current until an SRAM state flip results, while T_C is acquired by doing a transient simulation. On average, it takes the Cadence Spectre simulator 0.777 seconds to compute the critical current with a nano-amp precision. In addition, the average runtime for the critical time is 48 milliseconds. In comparison, for our C-based analytical models, the average runtime for I_C is 0.25 microseconds and 0.02 microseconds for T_C. As a result, the overall runtime speedup of our models over transistor-level circuit simulation is about 6 orders of magnitude.

Table VI. The Least-Square Fitted Parameter Values across the Technology Nodes (130nm–22nm)

nm	$W_N/W_P(nm)$	$V_{dd}(V)$	$V_{thn}(V)$	$V_{thp}(V)$	$K_N(A/V^2)$	$K_P(A/V^2)$	$C(F)$
130	*260/169*	*1.3*	*0.2*	*0.2*	*2.48e-4*	*6.53e-5*	*1.0e-15*
90	*180/117*	*1*	*0.2*	*0.2*	*2.43e-4*	*5.43e-5*	*6.0e-16*
65	*130/84.5*	*0.9*	*0.18*	*0.18*	*1.95e-4*	*4.59e-5*	*3.7e-16*
45	*90/58.5*	*0.9*	*0.18*	*0.18*	*1.54e-4*	*3.27e-5*	*2.2e-16*
32	*64/41.6*	*0.85*	*0.16*	*0.16*	*1.23e-4*	*2.76e-5*	*1.3e-16*
22	*44/28.6*	*0.8*	*0.16*	*0.16*	*7.48e-5*	*1.95e-5*	*7.3e-17*

Table VII. Summary on the Sensitivity of the System Parameters

	V_{dd}	V_{thn}	V_{thp}	K_N	K_P
I_C	*Very Strong*	*Weak*	*Very Weak*	*Strong*	*Weak*
T_C	*Very Strong*	*Weak*	*Very Weak*	*Strong*	*Weak*

7.5. Summary

In summary, the dependencies of critical time and critical current on several key design and technology parameters are evaluated. We also examine the effect of temperature and process variation effect on I_C and T_C. Furthermore, we studied the I_C and T_C dependencies on the system parameters shown in the equations in the Appendix. The simplification is done by keeping the targeted parameter as a variable while plugging nominal values of the other parameters into the equation. This provides us an immediate understanding of the parametric dependency of the targeted parameter. A short summary and key observation on sensitivity of system parameters with respect to global variation are as follows.

(1) Both I_C and T_C have very high dependency on V_{dd}. They grow approximately quadratically with V_{dd}.
(2) Both I_C and T_C also have high dependency on K_n. I_C tends to increase linearly with K_n, but T_C increases more rapidly with K_n.
(3) Both I_C and T_C have low dependency on the rest of the parameters.
(4) Both I_C and T_C increase as K_n and K_p increase but decrease as V_{thn} and V_{thp} increase.
(5) I_C does not depend on C, but T_C is highly dependent on the capacitance at stored nodes.

The critical time is approximately proportional to $1/(I_n\text{-}I_C)$. Clearly, a current injection must be greater than I_C in order to flip the state. Intuitively, a larger injection would make the cell to flip its state faster and the time to flip the state is inversely proportional to the difference between the amplitude of the current noise and I_C.

Furthermore, we rank the sensitivity of the system parameters (V_{dd}, V_{thn}/V_{thp} and K_n/K_p) as summarized in Table VII. T_C and I_C both depend on the same sets of device parameters such as transistor threshold voltages, which create correlation between the two.

In addition, we generate 500 samples on a level-1 MOSFET transistor model and on the derived analytical formula as shown in Figure 17. In the same figure, the data marked with a triangle are some design points simulated using the BSIM4 model (42 samples). In each of the 500 random samples, the system parameter values (V_{dd}, V_{thn}, V_{thp}, W_n, W_p, L_n, and L_p) are independently generated. Each system parameter follows the uniform distribution within ±30% of its nominal value. Based on the observation in Figure 17, critical time and critical current are highly correlated. By comparing the results with the design choice in BSIM4 marked in triangle in Figure 17 (42 samples; 6 samples for each parameter at ±10%, ±20, and ±30% with others remain at the nominal value; there are 7 parameters, so $6 \times 7 = 42$), the 500 analytical

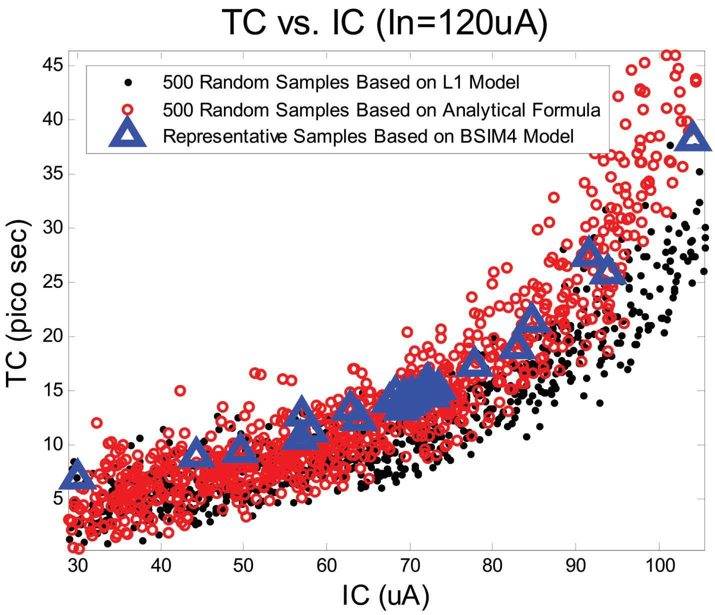


Fig. 17. Critical current (T_C) vs. critical current (I_C) (500 samples).

data fit the BSIM4 data well. Therefore, the analytical formula is able to capture the IC-TC relationship in BSIM4.

8. CONCLUSION

In conclusion, this article explores an analytical approach to the evaluation of dynamic stability analysis for SRAMs. The concepts of critical current and critical time, based on theoretically rigorous stability analysis of the dynamic behaviors of SRAM cells, provide physical characterizations of SRAM stability. While simple device models have been employed to derive the analytical dynamic stability models, our experimental results show that the models can provide good prediction of various parametric dependencies of dynamic stability and its technology scaling trends. Furthermore, the analytic requires less computational power. Compared with the transistor-level simulation, the derived analytic provides a speedup of six orders of magnitude. Lastly, the derived analytical models are also able to provide useful design insights and aid the designers to perform SRAM design optimization while considering the key dynamic stability property.

REFERENCES

G. Angelov and M. Hristov. 2004. SPICE modeling of mosfets in deep submicron. In *Proceedings of the 27th IEEE International Spring Seminar on Electronics Technology: Meeting the Challenges of Electronics Technology Progress*. Vol. 2, 257–262.

F. Arnaud, F. Boeuf, F. Salvetti, D. Enoble, F. Acquant, C. Regnier, et al. 2003. A functional 0.69 μm2 embedded 6t-sram bit cell for 65 nm cmos platform. In *Proceedings of the IEEE Symposium on VLSI Technology Digest of Technical Papers*. 65–66.

K. Cao. 2012. http://ptm.asu.edu/latest.html.

L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, et al. 2005. Stable sram cell design for the 32 nm node and beyond. In *Proceedings of the IEEE Symposium on VLSI Technology Digest of Technical Papers*. 128–129.

Y. Cheng, K. Imai, M. Jeng, Z. Liu, K. Chen, and C. Hu. 1997. Modelling temperature effects of quarter micrometre mosfets in bsim3v3 for circuit simulation. *Semiconductor Sci. Technol.* 12, 11.

W. Dong, P. Li, and G. M. Huang. 2008. SRAM dynamic stability: Theory, variability and analysis. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD'08)*. 378–385.

D. Edenfeld, A. B. Kahng, M. Rodgers, and Y. Zorian. 2004. 2003 technology roadmap for semiconductors. *Comput.* 37, 1, 47–56.

R. Garg, N. Jayakumar, S. P. Khatri, and G. Choi. 2006. A design approach for radiation-hard digital electronics. In *Proceedings of the 43rd Annual ACM Design Automation Conference*. 773–778.

R. Garg, P. Li, and S. P. Khatri. 2008. Modeling dynamic stability of srams in the presence of single event upsets (seus). In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'08)*. 1788–1791.

B. S. Haran, A. Kumar, L. Adam, J. Chang, V. Basker, S. Kanakasabapathy, et al. 2008. 22 nm technology compatible fully functional 0.1 μm 2 6t-sram cell. In *Proceedings of the IEEE International Electronic Devices Meeting (IEDM'08)*. 1–4.

Y. Ho. 2008. Dynamic stability margin analysis on sram. http://repository.tamu.edu/bitstream/handle/1969.1/ETD-TAMU-2722/HO-THESIS.pdf?sequence=1.

C. Hu. 2010. *Modern Semiconductor Devices for Integrated Circuits*. Vol. 1. Prentice Hall, Upper Saddle River, NJ.

G. M. Huang, W. Dong, Y. Ho, and P. Li. 2007. Tracing sram separatrix for dynamic noise margin analysis under device mismatch. In *Proceedings of the IEEE International Behavioral Modeling and Simulation Workshop (BMAS'07)*. 6–10.

S. M. Jahinuzzaman, M. Sharifkhani, and M. Sachdev. 2009. An analytical model for soft error critical charge of nanometric srams. *IEEE Trans. VLSI. Syst.* 17, 9, 1187–1195.

H. K. Khalil. 2002. *Nonlinear Systems*. Vol, 3, Prentice Hall, Upper Saddle River, NJ.

F. J. List. 1986. The static noise margin of sram cells. In *Proceedings of the 12th IEEE European Solid-State Circuits Conference (ESSCIRC'86)*. 16–18.

W. Liu and C. Hu. 1998. BSIM3v3 mosfet model. *Int. J. High Speed Electron. Syst.* 9, 3, 671–701.

W. Liu and C. Hu. 2011. *BSIM4 and MOSFET Modeling for IC Simulation*. World Scientific, Singapore.

J. Lohstroh, E. Seevinck, and J. D. Groot. 1983. Worst-case static noise margin criteria for logic circuits and their mathematical equivalence. *IEEE J. Solid-State Circ.* 18, 6, 803–807.

L. W. Massengill, M. L. Alles, and S. E. Kerns. 1993. SEU error rates in advanced digital cmos. In *Proceedings of the 2nd IEEE European Conference on Radiation and its Effects on Components and Systems (RADECS'93)*. 546–553.

T. C. May and M. H. Woods. 1979. Alpha-particle-induced soft errors in dynamic memories. *IEEE Trans. Electron. Devices,* 26, 1, 2–9.

T. H. Morshed, D. D. Lu, W. M. Yang, M. V. Dunga, X. Xi, et al. 2010. BSIM4v4.7 mosfet model. http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf.

S. Nassif. 2006. C-∞ shichman hodges model. http://ece.tamu.edu/~huang/files/materials606/sani.pdf.

J. C. Pickel and J. T. Blandford. 1981. CMOS ram cosmic-ray-induced-error-rate analysis. *IEEE Trans. Nuclear Sci.* 28, 6, 3962–3967.

A. Ramesh, S.-Y. Park, and P. R. Berger. 2011. 90nm 32 × 32 bit tunneling sram memory array with 0.5ns write access time, 1ns read access time and 0.5v operation. *IEEE Trans. Circ. Syst.* 58, 10, 2432–2445.

E. Seevinck. 1980. Deriving stability criteria for nonlinear circuits with application to worst-case noise margin of i2l. *IEEE Electron. Lett.* 16, 23, 867–869.

E. Seevinck, F. J. List, and J. Lohstroh. 1987. Static-noise margin analysis of mos sram cells. *IEEE J. Solid-State Circ.* 22, 5, 748–754.

H. Shichman and D. A. Hodges. 1968. Modeling and simulation of insulated-gate field-effect transistor switching circuits. *IEEE J. Solid-State Circ.* 3, 3, 285–289.

Y. Song, H. Yu, S. M. Pudukotai-Dinakarrao, and G. Shi. 2013. SRAM dynamic stability verification by reachability analysis with consideration of threshold voltage variation. In *Proceedings of the ACM International Symposium on Physical Design*. 43–49.

S. O. Toh, Z. Guo, and B. Nikolic. 2010. Dynamic sram stability characterization in 45nm cmos. In *Proceedings of the IEEE Symposium on VLSI Circuits (VLSIC'10)*. 35–36.

K. Utsumi, E. M. Morifuji, K. S. Aota, T. Yoshida, K. Honda, et al. 2005. A 65nm low power cmos platform with 0.495μm2 sram for digital processing and mobile applications. In *Proceedings of the IEEE Symposium on VLSI Technology Digest of Technical Papers*. 216–217.

Y. Wang, U. Bhattacharya, F. Hamzaoglu, P. Y. Ng, L. Wei, et al. 2009. A 4.0 ghz 291mb voltage-scalable sram design in 32nm high-κ metal-gate cmos with integrated power management. In *Proceedings of the IEEE International Solid-State Circuits Conference-Digest of Technical Papers (ISSCC'09)*. 456–457.

J. Wang, S. Nalam, and B. H. Calhoun. 2008. Analyzing static and dynamic write margin for nanometer srams. In *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED'08)*. 129–134.

B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky. 2006. Analytical modeling of sram dynamic stability. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD'06)*. 315–322.

Y. Zhang, P. Li, and G. M. Huang. 2010. Separatrices in high-dimensional state space: System-theoretical tangent computation and application to sram dynamic stability analysis. In *Proceedings of the 47th ACM/IEEE Design Automation Conference (DAC'10)*. 567–572.

Received October 2013; revised May 2014; accepted May 2014