The effect of power supply ramp time on SRAM PUFs

Conference Paper · August 2017

DOI: 10.1109/MWSCAS.2017.8053081

CITATION
1

READS
192

3 authors, including:

Abdelrahman T Elshafiey
University of New Mexico
2 PUBLICATIONS **1** CITATION

SEE PROFILE

The Effect of Power Supply Ramp Time on SRAM PUFs

Abdelrahman T. Elshafiey and Payman Zarkesh-Ha
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM, USA
Email: atarief@unm.edu and pzarkesh@unm.edu

Joshua Trujillo
Department of Energy's National Security Campus
Managed by Honeywell
Kansas City, MO, USA
Email: jtrujillo@kcp.com

***Abstract*—In this paper, for the first time, it is demonstrated that the start-up value of an SRAM PUF could be different depending on the SRAM power supply rising time. An analytical model has been developed to determine the range for the power supply ramp time that affects the SRAM PUF start-up value. It has been found that there are two regions of operation. As a result, the generated key could possibly be different from one region to another. An SRAM test chip was designed and fabricated using Tower Jazz's 180 nanometer Silicon Germanium (SiGe) Bipolar/CMOS (BiCMOS) process. Based on our measured data, using the appropriate rising time can decrease the number of flipping bits by 5%. Both simulation and silicon results confirms the analytical model.**

Keywords— Physically Unclonable Functions (PUFs); SRAM; Power supply rising time

I. INTRODUCTION

Physical unclonable functions (PUFs) became a reliable technology for hardware cryptography and key storage. A PUF is a function that generates a set of responses (secrets), when it is stimulated by a set of challenges. It is a physical function because the challenge-response relation is defined by complex properties of a physical material, such as the manufacturing variability of CMOS devices [1]. Its unclonability is attributed to the fact that due to the manufacturing variability that defines the secret, one cannot manufacture two identical chips, even with full knowledge of the chip design. PUF is a promising solution to security issues like intellectual property (IP) protection, device authentication, and user data privacy [1].

SRAM is one of the popular implementations of PUF. SRAM PUF employs an SRAM cell (two cross-coupled inverters), and exploits the random assignment of a stable state from an initial unstable state. The final state of the cell is determined by the random mismatches in the pair of inverters [2]. The mismatches in the SRAM cells produced during fabrication process could vary from cell to cell. We can classify the SRAM cells based on the mismatches degree, non-skewed cells, partially skewed cells and fully skewed cells [3]. All the work in this paper is focused on partially skewed cells, since non-skewed cells always cause bit flipping under any conditions, and fully skewed cells always produce a stable output under normal conditions.

Only few publications have discussed and modeled the start-up value of SRAM PUF. Previous publications focused on the behavior of the start-up value as a function of the supply voltage and temperature only [3]. In this paper, we present an analytical model for the start-up value as a function of the power supply rising time. Thermal sensitivity analysis will be performed on the test chip in future work.

This work was supported primarily by Honeywell corporation, Award No. A15-0517-002.

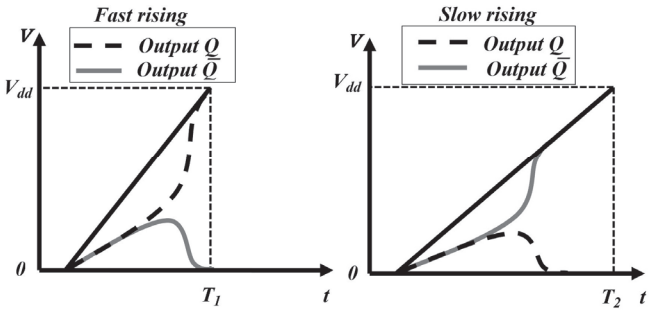

Fig. 1. SRAM response as a function of fast and slow power supply rising time.

The initial concept demonstrated in this paper is shown in Fig. 1, where two different power supply ramps could lead to two different outputs, depending on the cell variations and the power supply ramp time. However, if the variations are very small, or if the cell is symmetrical, the outputs reach a metastable state, this metastable point does not hold for long. Any small deviation from the metastable point is immediately amplified by the positive feedback and the circuit moves away from the metastable point towards one of both stable points. Since electronic circuits are constantly affected by small deviations due to random noise, the non-skewed SRAM cell never stays in its metastable state very long instead it will quickly end up in one of both stable states (randomly) [4].

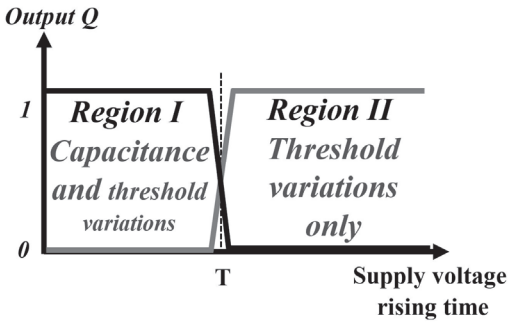

Fig. 2. Two different regions, with different dominant fabrication process variations.

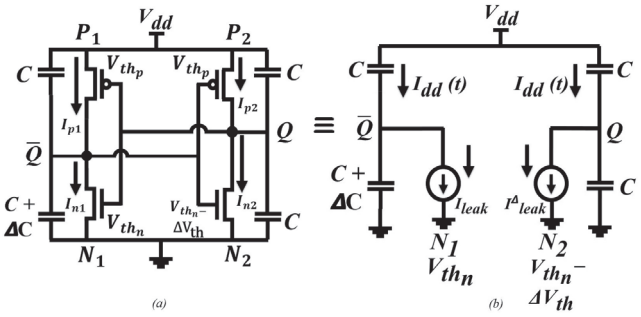

Fig. 3. CMOS SRAM cell including the gate capacitances.

There are several types of mismatch in the cells during the fabrication process, however capacitance and threshold variations are the most crucial mismatches in determining the final SRAM cell state. As shown in Fig. 2, there are two regions of operation. Where T is the power supply rising time, which separate between the two regions. If the rising time of the power supply is faster than time T, the SRAM cell operates in region I, where capacitance and threshold variations decide the final state. If the rising time is slower than T, the SRAM operates in region II, where the threshold variations alone determines the final SRAM state. To understand how capacitance and threshold mismatches play a role in determining the SRAM cell outputs, an analytical model is discussed in section II. Simulation results are presented on an SRAM cell in section III. In Section IV experimental data is shown and discussed. Finally, the paper concludes in section V.

II. Analytical Model

In this section an analytical model is derived for the rising time of the power supply T, which separates the two regions. The model also explain how the rising time of the power supply determine which process variations decide the final state of the output in both regions I and II. The SRAM cell here can be explained using a simple equivalent circuit where the MOSFET effective capacitances are depicted in Fig. 3. Here we assume that all capacitances are equal to C, except the capacitance of N_1 that is equal to $C + \Delta C$, where ΔC is the capacitance variation. Since the rising time of the power supply is gradually increasing, the four transistors do not turn on immediately, and they conduct in the subthreshold region. Also, let's assume that the threshold voltage of transistor N_2 is V_{thn}- ΔV_{th}, where ΔV_{th} is the threshold voltage variation. For simplicity, it can be assumed that the threshold voltage of NMOS transistors is less than that of PMOS, therefore the current of NMOS transistors dominates, and we can ignore the PMOS currents in our analysis.

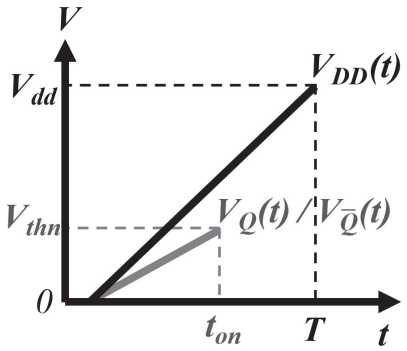

Fig. 4. The outputs of the SRAM increasing with the power supply ramp time at a slower rate.

A. Derivation

The power supply ramp time can be defined by:

$$V_{DD}(t) = V_{dd} * \frac{t}{T} \tag{1}$$

where V_{dd} is the supply voltage, T is the power supply rising time, and t is time. As shown in Fig. 3b, the two outputs Q *and* $\bar{Q}$, build capacitive voltage dividers, and follow $V_{DD}(t)$ based on the ratio between the gate capacitances. At the same time, the outputs Q *and* $\bar{Q}$ are following the supply voltage, the two NMOS transistors are also discharging Q *and* $\bar{Q}$. Therefore, the two outputs increase with V_{dd} at a slower rate. In Fig. 3b, at node Q *and* $\bar{Q}$, there is a superposition between the power supply current, and the NMOS leakage current. To derive an equation for T, we develop a model for V_Q *and* $V_{\bar{Q}}$ as a function of V_{dd}, T and NMOS leakage current.

At node Q, there is a superposition between the current supplied from the power supply, and the current drawn by the NMOS transistor (subthreshold current). $V_1(t)$ is the voltage supplied by $V_{DD}(t)$ at node Q, and is calculated using the voltage division between the two capacitors:

$$V_1(t) = V_{DD}(t) * \frac{C}{2C + \Delta C} \tag{2}$$

Substituting (1) into (2*)*:

$$V_1(t) = V_{dd} * \frac{t}{T} * \frac{C}{2C + \Delta C} \tag{3}$$

$V_2(t)$ is the voltage driven by the NMOS current source at node Q, and is calculated from the current capacitance relation:

$$I_{leak} = C_{equ} * \frac{\Delta V}{\Delta t} \tag{4}$$

where I_{leak} is the NMOS leakage current, C_{equ} is the equivalent capacitance which is C in parallel with $C + \Delta C$, and ΔV is $V_2(t)$.

$$V_2(t) = \frac{I_{leak} * \Delta t}{2C + \Delta C} \tag{5}$$

Therefore, $V_Q(t)$ equals $V_1(t) — V_2(t)$:

$$V_Q(t) = V_{dd} * \frac{t}{T} * \frac{C}{2C + \Delta C} — \frac{I_{leak} * \Delta t}{2C + \Delta C} \tag{6}$$

Repeating the previous steps to find $V_{\overline{Q}}$, considering that both capacitances are equal to C, and there is a variation ΔV_{th} in the threshold of the NMOS transistor, which leads to have I^{Δ}_{leak} instead of I_{leak}.

$$V_{\overline{Q}}(t) = V_{dd} * \frac{t}{T} * \frac{1}{2} - \frac{I^{\Delta}_{leak} * \Delta t}{2C} \quad (7)$$

From Fig. 4, it shows that both $V_Q(t)$ *and* $V_{\overline{Q}}(t)$ are equal, and both reach V_{thn} exactly at the same time t_{on}. At this point the capacitance variation ΔC in $V_Q(t)$, and the threshold voltage variation ΔV_{th} in $V_{\overline{Q}}(t)$ compensate each other and both outputs reach the metastable point described previously. Therefore, we can equate (6) and (7) at time t_{on}, and then get an expression for the power supply rising time T, which is the borderline between region I and II.

$$\frac{V_{dd} * t_{on}}{T} * \frac{C}{2C + \Delta C} - \frac{I_{leak} * t_{on}}{2C + \Delta C} = \frac{V_{dd} * t_{on}}{T * 2} - \frac{I^{\Delta}_{leak} * t_{on}}{2C} \quad (8)$$

After some simplifications the final expression for T is:

$$T = \frac{V_{dd} * \Delta C/2}{[I^{\Delta}_{leak} * (1 + \frac{\Delta C}{2C})] - I_{leak}} \quad (9)$$

Assuming, $\frac{\Delta C}{2C} \ll 1$:

$$T = \frac{V_{dd} * \Delta C/2}{I^{\Delta}_{leak} - I_{leak}} \quad (10)$$

The equations of the leakage currents are:

$$I_{leak} = I_o * e^{\frac{V_{GS} - V_{thn}}{n * V_T}} * (1 - e^{\frac{-V_{DS}}{V_T}}) \quad (11)$$

$$I^{\Delta}_{leak} = I_o * e^{\frac{V_{GS} - V_{thn} + \Delta V_{th}}{n * V_T}} * (1 - e^{\frac{-V_{DS}}{V_T}}) \quad (12)$$

From (11) and (12):

$$I^{\Delta}_{leak} = I_{leak} * e^{\frac{\Delta V_{th}}{n * V_T}} \quad (13)$$

where I_o is the saturation current, V_T is the thermal voltage (25.9 *mv* at room temperature), and n is an empirical parameter (around 1.5).

Finally, equation (10) can be simplified to:

$$T = \frac{C * V_{dd}}{I_{leak}} * \frac{\Delta C/2C}{e^{\frac{\Delta V_{th}}{n * V_T}} - 1} \quad (14)$$

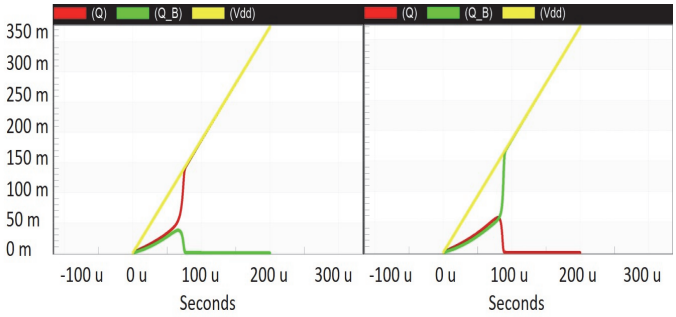

Fig. 6. SRAM cell transient simulation, using two different rising times.

B. Discussion

Equaiton (14) is a simple model that can provide valuable information about the SRAM behaviour under power supply ramp time. For instance, if there is no capacitance variation in the cell, $\Delta C = 0$, the rising time T becomes zero, therefore region I disappears. On the other hand, if there is no threshold voltage variation in the cell, $\Delta V_{th} = 0$, accordingly T goes to infinity, therefore region II disappears. As a result, if the supply ramp is slow enough, the capacitance variations can be neglected and SRAM falls in region II. While in region I, the supply ramp is fast enough, that the capacitance variation cannot be neglected. Note that this model does not determine the output in each region, but rather predicts which variation affects the SRAM output.

III. Simulation Results

TABLE I. Parameters for 180nm SBC18H3 model

Parameter	NMOS	PMOS
Supply voltage V_{dd} (in *V*)	1.8	
Length L (in μm)	0.18	0.18
Width W (in μm)	0.6	0.9
Threshold voltage inverter 1 V_{th} (in *V*)	0.3532	- 0.3958
Threshold voltage inverter *2* V_{th} (in *V*)	0.335	- 0.3958
Threshold voltage variation ΔV_{th} (in *mV*)	17	
Capacitance variation ΔC (*fF*)	7	

A. Setup

The start-up value of an SRAM cell is simulated using SPICE and transistor model. The CMOS parameters used in the simulation are listed in Table. 1. The difference between NMOS V_{thn1} and NMOS V_{thn2} is 5%. A 7 *fF* capacitance is added between node $\overline{Q}$ and ground to simulate the capacitance variations, and to balance the threshold variation at node Q.

B. Results

Since the threshold voltage of N_2 is lower, transistor N_2 discharges the output Q at a faster rate than N_1 discharging $\overline{Q}$.

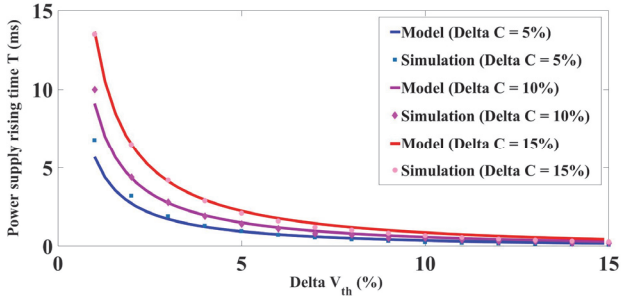

Fig. 7. Threshold voltage variation versus Power supply rising time T, for both the simulation and the model.

However, the device capacitance of N_2 is larger by 17 *fF*, which forces $\overline{Q}$ to increase at a rate slower than Q. This balance in charging and discharging between Q *and* $\overline{Q}$ is separated by the power supply ramp time. After trial and error, it has been found that the boarder time T, between region I and II is around 1.5 *ms* , at which the capacitance variation dominates and $\overline{Q}$ eventually goes to zero. If T is more than 1.5 *ms*, the threshold variation dominates and Q goes to zero. As shown in Fig. 6, on the left figure the rising time is less than 1.5 *ms*, and on the right figure the rising time is more than 1.5 *ms*. In Fig. 7, the rising time T developed from the model is illustrated over the range of different ΔV_{th}, and three different values for ΔC, using the parameters in Table. 1. The SRAM cell is also simulated and measured over the same range of ΔV_{th} and ΔC. It can be clearly seen that the simulation matches the model very well.

IV. Silicon Results

Tower Jazz Semiconductor was chosen to fabricate the PUF test chip in this project. The fabrication process is the 180 nm SiGe BiCMOS technology with 6 metal layers. The top 2 thick metal layers are used for power/ground distributions to minimize the IR drop and switching noise issues. In addition, it offers triple well isolation as well as deep trench isolation for better latch up resiliency. The core power supply is 1.8V, while high voltage power supply for I/O is 3.3V.

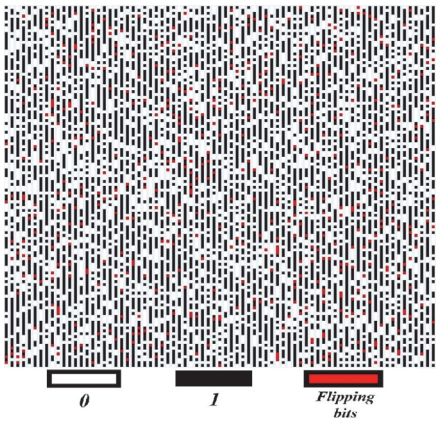

Fig. 8. 125×150 bits SRAM PUF response. Black dots are 1's, white dots are 0's and red dots are flipping bits.

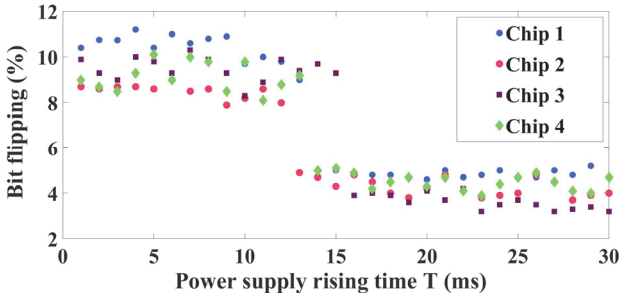

Fig. 9. Extract data from SRAM test chip. Number of flipping bits versus supply voltage rising time.

The map of one of the tests at a power supply rising time of 15 ms is presented in Figure 8. As shown the faulty bits are evenly distributed, which concludes that the chip is well designed and well fabricated.

Figure 9 illustrates the bit flipping percentage versus power supply ramp time in four SRAM test chips, where each chip has been tested 20 consecutive times, over the range of 10 ns to 30 ms power supply rising time. The two regions previously described can be easily identified in this figure with crossover point of about 14ms. The percentages of bit flipping are about 10% in region I and 5% in region II. The higher bit flipping in region I can be explained by the fact that at a rising time less than 14 ms, the capacitance variations start to have an effect. The capacitance variations balances the threshold variations, resulting in more symmetric cells, where the cell output value will be determined by noise rather than device variations.

V. Conclusion

In this paper, the start-up value of an SRAM cell as a function of power supply rising time has been discussed. It has been found that there are two regions of operation. An analytical model was presented, and the power supply rising time T, that separates the two regions, has been developed as a function of threshold and capacitance variations. The simulation results support the analytical model. The extract data from the test chip also show two regions. From measured data, in region I, 90% of the cells are fully skewed, and 10% are non-skewed cells. In region II, 95% of the cells are fully skewed, and only 5% are non- skewed. Therefore, the rising time of the power supply can change the skewness of some cells.

References

[1] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in Proceedings of the 44th annual Design Automation Conference, 2007, pp. 9–14.

[2] Holocomb, D., Burleson, W.: Power-up SRAM State as an Identifying Fingerprint and Source of True Random Numbers. IEEE Transactions on Computers. Vol. 57, No. 11, November (2008).

[3] M. Cortez, A. Dargar and S. Hamdioui, "Modeling SRAM Start-Up Behavior for Physical Unclonable Functions", Int. Symp. *on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT),* 2012.

[4] R. Maes, "Physically Unclonable Functions: Constructions, Properties and Applications", Springer, 2013.