Statistical Analysis and Design of 6T SRAM Cell For Physical Unclonable Function with Dual Application Modes

Le Zhang, Chip-Hong Chang, Zhi Hui Kong, and Chao Qun Liu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

***Abstract*—Apart from performance and power efficiency, security is another critical concern in the modern memory sub-system design. SRAM, which is routinely used as a data preservation component, has now been developed into an effective primitive known as Physical Unclonable Function (PUF) for cryptographic key generation to protect the sensitive local information. Considering the constraints of hardware resource on embedded systems, it is desirable to have an SRAM used both as a regular memory and a PUF to save the overheads of having these two functions implemented independently. Unfortunately, while process variations are the entropy sources for secure key generation, it impacts failure rates in memory-mode operations. This paper presents a statistical analysis on SRAM and provides an insight into how the SRAM cell geometry can be optimized to qualify it for both modes of operation simultaneously.**

I. Introduction

SRAM based Physical Unclonable Function (SPUF) [1] is considered as a promising primitive for security applications such as cryptographic key generation and device identification. Compared to other PUF instances, the mechanism and implementation of SRAM-based PUFs are comparatively straightforward. That is, random bits can be directly produced owing to the random mismatch between the two inverters in an SRAM cell by powering up the array, and the generated bits can then be read through the I/O interface of the SRAM. No other modifications are required on the original circuitry for this functionality. However, to meet the quality specifications for the two different application modes (regular memory and PUF) requires novel design strategies. This is because the design objectives for memory application and PUF application are contradictory to each other by nature. The former needs high tolerance of process variations while the latter actually replies on process variations to strengthen its security. It is because of this conflict that SRAM cells dedicated to PUF function will not be used for regular memory.

To overcome this dilemma, we carry out a statistical study on the SRAM characteristics in the two alternative working modes and analyze how design parameters (e.g., transistor size) affect the qualities. To the best of our knowledge, there is no such kind of study in the existing literature. Statistical analysis based on industrial CMOS model was performed on SRAM cell to evaluate the key performance metrics like failure probability for memory application and reliability for PUF application. Based on the analysis, a method to design SPUF with dual modes of operation is proposed. An illustrative case study on a 45 nm SRAM cell was presented to demonstrate the effectiveness of our design method.

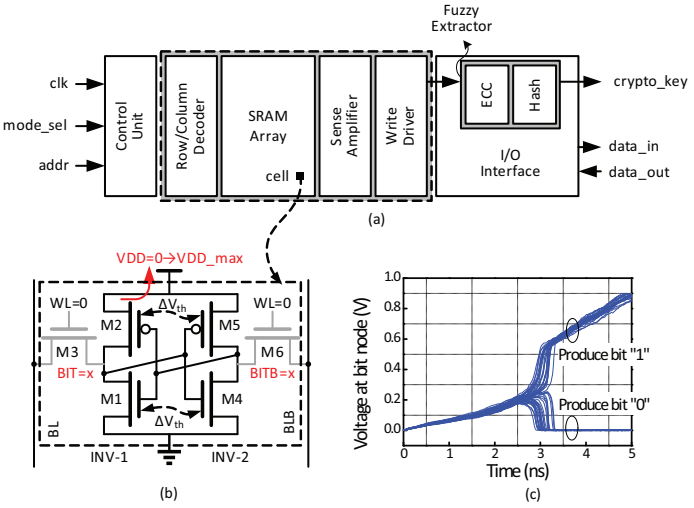
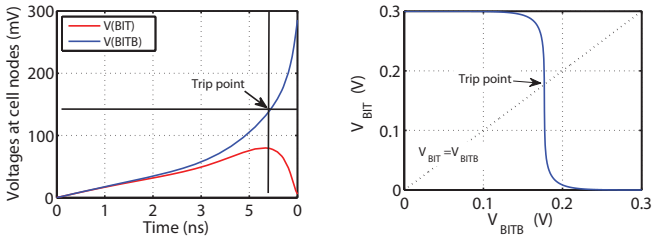


Fig. 1. (a) Architectural diagram of an SPUF; (b) mechanism of SPUF in cell view; (c) voltage signals at BIT and BITB of 50 unique SRAM cells with randomly distributed parameters during power-up process.

The rest of the paper is organized as follows. In Section II, the fundamentals of SRAM-based PUF are reviewed. In Section III-C, circuit-level analysis on SRAM cells are presented. Section IV shows the proposed design method and the simulation results. The paper is concluded in Section V.

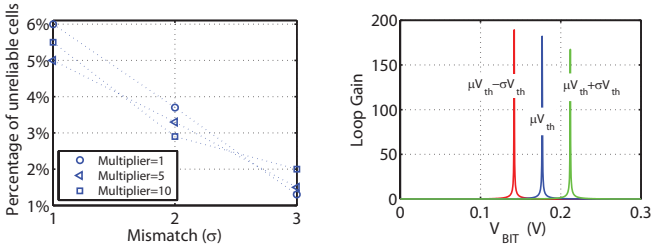
II. Fundamentals of SPUF

An SPUF is essentially a SRAM which can work in alternative modes of memory and PUF. In a good SPUF, 1) the functionality of one application mode does not affect the other; 2) the qualities of both application modes should meet the system-level specifications; 3) no hardware design modifications are needed for the dual-mode applications. The architectural diagram of an SPUF is shown in Fig. 1(a). In the PUF mode, a set of memory addresses (called *challenge* and denoted as C) is used to randomly interrogate a subset of SRAM cells to obtain a binary string (called *response* and denoted as R). The response is usually post-processed by a *fuzzy extractor* to produce the final cryptographic key. The response randomness of the SPUF is ascribed to race condition of the power-up process. During power-up, as the



(a) (b)

Fig. 2. (a) Power-up process of BIT and BITB nodes of an SRAM. The two curves break away when one of them reaches the trip point of the inverter. (b) VTC of an inverter in the sub-threshold region and its trip point.



(a) (b)

Fig. 3. (a) Relationship between the percentage of unreliable bits and variation of V_{th} for different scaling factors $W_{\text{new}}/W_{\text{old}}$, with the transistor length kept at the minimum feature size of the process technology. (b) Relationship between the loop gain and V_{th} variation.

supply voltage ramps up, the two cross-coupled inverters will race to reach their stable and discriminative output values. The final state of the cell is determined by the mismatch of the two inverters. The mechanism is shown in Fig. 1(b). The simulated results of the power-up process on 50 SRAM cells with random mismatch properties are shown in Fig. 1(c).

III. Circuit-level Analysis of SPUF

A. SRAM Power-up behavior

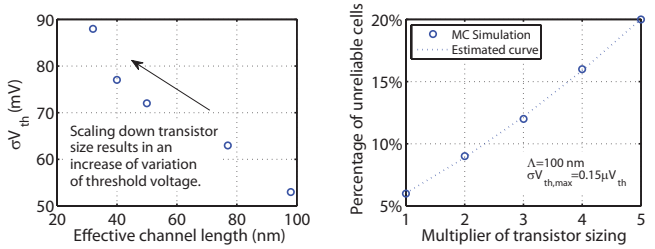
In PUF mode, each SRAM cell produces a random bit in the power-up process. During this process, all the transistors in the cell are in the sub-threshold region. The drain-source current in this region is formulated as

$$I_{SUB} = I_S \exp\left(\frac{V_{gs} - V_{th}}{nV_T}\right)\left[1 - \exp\left(\frac{-V_{ds}}{V_T}\right)\right] \quad (1)$$

where I_S is the saturation current, $n = 1 + C_d/C_{ox}$ is the subthreshold slope factor and $V_T = kT/q$ is the thermal voltage. k, T and q are Boltzmann constant, temperature and elementary charge, respectively. The current flowing through M1 and M4 will gradually pull up the voltage at node BIT and BITB simultaneously. Owing to the mismatch of transistors in the two inverters, one of the two nodes may reach the "trip-point" (V_{trip}) of the inverter more quickly than the other. If the voltage of node BIT of inverter INV-1 climbs to V_{trip} first, the output node of the inverter INV-2, i.e., node BITB, will be pulled down to the ground. As the voltage at BITB descends, transistor M2 will be fully turned on eventually and pulls the voltage at BIT to VDD. This process is shown in Fig. 2(a).

B. Impact factors of PUF quality

SRAM based PUF is featured by its good randomness and uniqueness compared to other PUF instances. This is because the response bits are generated from individual cells with mutually independent stochastic parametric variations. However, the mismatch between the inverters of a SRAM cell is not sufficiently large, which may be easily overwhelmed by noise effects. Therefore, reliability is always the utmost concern in SRAM based PUF design. The reliability is defined as *the probability that the same bit can be generated from the cell at different time points*. Apparently, the reliability of the SRAM-based PUF can be affected by circuit noises. Besides,



(a) (b)

Fig. 4. (a) Changes in σV_{th} with the variation of transistor effective channel length [2]. (b) The relationship between the percentage of unreliable bits and transistor size.

the environmental conditions in which the devices operate will also impact the reliability[1].

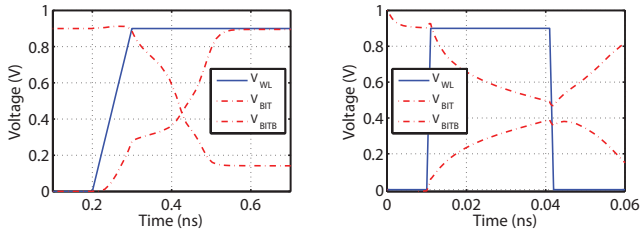
1) Mismatch of INV-1 and INV-2: Intuitively, the larger the mismatch, the higher the reliability. The major contributor of the mismatch between INV-1 and INV-2 is the difference in the transistor threshold voltages. The transistors in power-up process are in sub-threshold region. The current [formulated in Eq. (1)] flowing through the transistor has an exponential relationship with the threshold voltage, so that a small discrepancy of V_{th} between the transistors in INV-1 and INV-2 is sufficient to discriminate the driving current and separate the voltage curves during the power-up process.

2) Loop-gain at trip point: Another factor is the "loop-gain" at the trip point [3]. The larger the gain, the more quickly the voltages at BIT and BITB nodes stabilize at their steady state values. At the same time, it is also more easily for the noise to be amplified during this process to overturn the final output randomly. According to [4], the loop-gain can be formulated as

$$LG(V_{BIT}) = \frac{\partial \text{VTC}_1}{\partial V_{BIT}} \cdot \frac{\partial \text{VTC}_2}{\partial V_{BITB}} \quad (2)$$

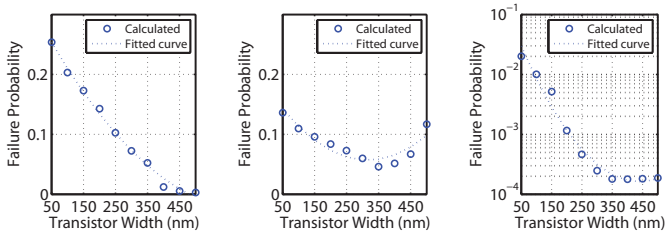
where $V_{BITB} = \text{VTC}(V_{BIT})$. VTC refers to the voltage transfer characteristic function. The VTC of an inverter in sub-

[1]In this paper, we merely focus on the factors related to *SRAM cell properties*. Other factors such as temperature and supply fluctuation may also affect reliability but are not discussed here.



(a) (b)

Fig. 5. (a) Read and (b) write failure effects in an SRAM cell. The original bit preserved in the cell is 1.



(a) (b) (c)

Fig. 6. The relationship between the failure probability and the width of transistor: (a) $M1$, (b) $M2$ (c) $M3$.

threshold region can be formulated as follows [5]

$$V_{BITB} = \frac{nV_T}{2}\left[\ln\frac{I_{S2}}{I_{S1}} + \ln\left(\frac{1-\exp\left(\frac{-VDD-V_{BIT}}{V_T}\right)}{1-\exp\left(-\frac{V_{BIT}}{V_T}\right)}\right)\right] + \frac{VDD}{2} + \frac{V_{th,1}-V_{th,2}}{2} \quad (3)$$

The trip point voltage [as shown in Fig. 2(b)] can be obtained numerically with this formula, and then the loop-gain can be obtained by Eq. (2).

3) Transistor design parameters: From Eq. (1) and Eq. (2), it can be deduced that the transistor size of the SRAM cell does not have a strong influence on either I_{DS} or LG. Hence, variations of the transistor geometric parameters do not contribute much to the noise immunity of SRAM cell operating in PUF mode. Fig. 3(a) shows that sizing the transistors up will not affect the reliability of SRAM based PUF siginificantly. Instead, the variations of V_{th} exhibit an obvious effect on the aforementioned impact factors of PUF reliability. The effect of the variation of V_{th} on the loop-gain is illustrated in Fig. 3(b).

Owing to the correlation of the transistor size to the random dopant density of MOSFET transistors, the intra-die variation of threshold voltage can be related to transistor geometry by the following formula [2]:

$$\sigma V_{th} = \frac{\Lambda \cdot \sigma V_{th,\max}}{\sqrt{(W \cdot L)}} \quad (4)$$

where Λ is a technology-dependent parameter and $\sigma V_{th,\max}$ is the maximum value of σV_{th}. Considering this effect (see Fig. 4(a)), sizing up the transistors will make the mismatch of V_{th} between the two inverters smaller, and consequently lead to lowered reliability of generated response bits. This is verified by the simulation results shown in Fig. 4(b).

C. SRAM Read/write failure effects

In the memory mode, the prime consideration is the failure probability. In this work, we consider two main types of failures, i.e., read failure (RF) and write failure (WF).

- RF occurs when the voltage read at node BIT (or BITB) rises above the trip point of the inverter to cause the data preserved in the SRAM cell to flip accidentally [see Fig. 5(a)]. As shown in Fig. 1(b), when the bitline (BL) is precharged, the access transistor (M6) and the pull-down transistor (M4) act as a voltage divider, and the voltage at node BITB may trigger a flip of the data bit in the SRAM cell.
- WF refers to the effect that when a bit is written into an SRAM cell, the inverters of the SRAM cell fail to respond to the data bit within an acceptable time [see Fig. 5(b)]. Consider the circuit diagram of an SRAM cell shown in Fig. 1(b), bit 0 will be written into the cell which preserves data 1, that is, the voltage at node BIT will be discharged to a low value. If this value is not smaller than the trip point of the inverter INV-2, the write operation will fail and the data bit stored in the SRAM remains unchanged.

The impact of process variations (e.g., V_{th}) will cause certain SRAM cell to fail in operation occasionally. The overall failure probability [$\mathbf{Pr}$(RF,WF)] of an SRAM cell can be estimated by

$$\mathbf{Pr}(\text{RF,WF}) = \mathbf{Pr}(\text{RF}) + \mathbf{Pr}(\text{WF}) - \mathbf{Pr}(\text{RF} \cdot \text{WF}) \quad (5)$$

It is possible to reduce this failure probability by manipulating the SRAM cell transistor sizes. The relationship between $\mathbf{Pr}$(RF,WF) and the sizing of each transistor[2] in the SRAM cell is shown in Fig. 6.

IV. Design Method and Result Evaluation

To design a SPUF with dual application modes, the SRAM cell needs to be well sized so that the quality expectations can be met for both applications. For memory mode, the failure probability should be kept as low as possible. For PUF mode, the randomness and reliability of the generated bits should be made as high as possible. This paper focuses on the reliability since it is a more critical metric for SRAM-based PUF compared to other types of PUF as mentioned earlier. Besides the requirement of failure probability and PUF reliability, area and leakage constraints of the targeted process technology will also be considered.

[2]The readers may refer to [6] to find out more about how geometric parameters of transistors affect the failure probabilities.

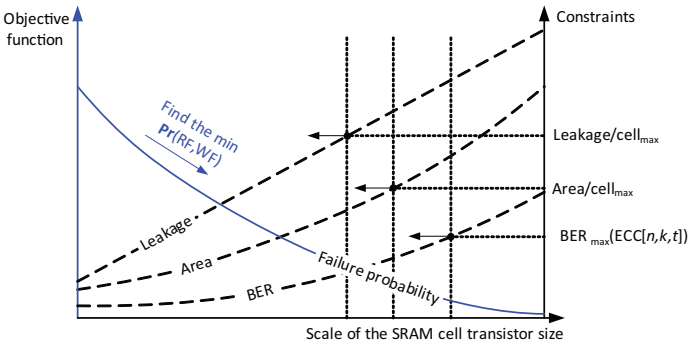


Fig. 7. Design space of an SRAM cell given multiple specifications.

TABLE I
SIMULATION RESULTS @$VDD = 0.9$ V, TEMPERATURE= 25°C

$\mathbf{X}$	$\mathbf{Pr}$(RF,WF)	BER	Area	Leakage
(180, 90, 135)	0.12	1.81×10^{-9}	0.45	144.33
(500, 250, 375)	0.07	5.53×10^{-4}	1.28	270.70
(304, 152, 228)	0.09	7.52×10^{-7}	0.70	166.43

[1]The transistor width is searched in the range of $[45, 500]$ nm. Ratios of transistor widths, i.e., $W_{M1} : W_{M2} : W_{M3}$, for the three designs are kept the same to guarantee that no failure occurs with nominal parameter values in the memory mode.
[2]The units for area and leakage current are μm^2 and nA, respectively.

A. Problem formulation

According to the above analysis, the design space of an SPUF is shown in Fig. 7, and the objective is to find an optimal design point for the following problem.

$$\min_{\mathbf{X}} \mathbf{Pr}(\text{RF,WF}) = f(\mathbf{X})$$

where $\mathbf{X}$ is the SRAM cell parameters (e.g., W). The objective function is subjected to the following constraints:

$$\begin{cases} \text{reliability}(\text{ECC}[n,k,t]) \geq \text{reliability}_{\min} \\ \text{area\_cell} \leq \text{area\_cell}_{\max} \\ \text{leakage\_cell} \leq \text{leakage\_cell}_{\max} \end{cases} \quad (6)$$

where ECC$[n, k, t]$ is the Error Correction Code specified by code parameters n, k and t.

The above objective function and the constraints can be met by the following calculations.

1) Failure probability: The failure probabilities $\mathbf{Pr}$(RF) and $\mathbf{Pr}$(WF) can be estimated efficiently by Monte Carlo simulation with "Mixture Importance Sampling" technique [7].

2) Reliability: The raw reliability (i.e., complementary Bit-Error Rate (BER)) of a response string generated from a PUF can be estimated by the following equation:

$$\text{reliability}_{\text{raw}} = 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{\mathsf{HD}(R, R')}{n} \quad (7)$$

where m is the number of response strings generated to the same challenge, n is the bit-length of the response, and $\mathsf{HD}(.,.)$ computes the Hamming distance of two binary strings.

Considering ECC$[n, k, t]$, the final reliability is

$$\text{reliability}_{\text{final}} = \sum_{i=0}^{t} \binom{n}{t} (1 - \text{reliability}_{\text{raw}})^i (1 - (1 - \text{reliability}_{\text{raw}}))^{n-i} \quad (8)$$

3) Area and leakage: The area of an SRAM cell can be estimated via layout [6], and the leakage can be directly obtained from SPICE simulator.

B. Simulation and discussion

We illustrate our method with three SRAM cell designs with different specifications. The parameters of first design has the minimum values for $\mathbf{X}$. The second design has the lowest $\mathbf{Pr}$(RF,WF) within the search space of $\mathbf{X}$. The third design is optimized under the constraint of reliability (i.e., 10^{-6}). The simulation was carried out by using the HSPICE simulator with CMOS 45 nm bulk Predictive Technology Model. The maximum relative standard deviation of V_{th} is assumed to be 15% of the nominal value and the value of Λ is assumed to be 50 nm. For simplicity, the transistor length is set to 45 nm. The ECC used is BCH-$[127, 15, 27]$. Table I lists the evaluation results. It shows that the optimal design is a tradeoff between the first and second designs, where the failure probability and area are sacrificed for higher reliability. The optimal design reduces the failure probability from the baseline design by 25%, but still maintains the BER within the acceptable range of $\leq 10^{-6}$. The increase in area and leakage are $\sim 56\%$ and $\sim 15\%$, respectively.

V. CONCLUSION

In this paper, a statistical analysis is performed on SRAM for dual application modes. Our analysis shows that the qualities of both memory-mode and PUF-mode are related to the geometric parameters of SRAM cells. The findings offer the researchers and memory circuit designers an idea to explore the design space of SRAM cell sizing to satisfy both requirements of memory and PUF operations.

This work is supported by the Singapore Ministry of Education Academic Research Fund Tier I (MOE 2014-T1-002-141).

REFERENCES

[1] D. Holcomb *et al.*, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Trans. Comput.*, vol. 58, no. 9, pp. 1198–1210, Sept. 2009.
[2] M. Orshansky *et al.*, *Design for manufacturability and statistical design: a constructive approach*. Springer, 2008.
[3] S. Chellappa, A. Dey, and L. T. Clark, "Improved circuits for microchip identification using SRAM mismatch," in *IEEE CICC*, 2011, pp. 1–4.
[4] K. Agarwal and S. Nassif, "Statistical analysis of SRAM cell stability," in *ACM DAC*, 2006, pp. 57–62.
[5] B. H. Calhoun and A. P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 7, pp. 1673–1679, 2006.
[6] S. Mukhopadhyay, H. Mahmoodi, and K. Roy, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled CMOS," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 24, no. 12, pp. 1859–1880, 2005.
[7] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *ACM DAC*, 2006, pp. 69–72.