A Novel Modeling Attack Resistant PUF Design based on Non-linear Voltage Transfer Characteristics

Arunkumar Vijayakumar and Sandip Kundu
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst, USA
{avijayakumar, kundu}@ecs.umass.edu

***Abstract*— Physical Unclonable Function (PUF) circuits are used for chip authentication. PUF designs rely on manufacturing process variations to produce unique response to input challenges. It has been shown that many PUF designs are vulnerable to machine learning (ML) attacks, where a model can be built to predict PUF response to any input after only a few observations. In this work, we propose a ML attack resistant PUF design based on a circuit block to implement a non-linear voltage transfer function. The proposed circuit is simple, exhibits high uniqueness and randomness. Further improvements are proposed to enhance PUF reliability. The proposed circuit was simulated in a 45nm technology process and the results indicate a significant improvement in ML attack resistance in comparison to traditional PUFs. Results on uniqueness and reliability are also presented.**

Keywords*— Physical unclonable function; security; modeling attack

I. Introduction

Digital dependency has become pervasive in our daily lives. As electronic systems and digital agents replace many of the tasks performed previously by humans, security vulnerability of these devices creates financial, personal and social risks. Hardware authentication is a key element of security. At a basic level, hardware security is concerned with establishing authenticity and provenance of ICs.

Physical Unclonable Functions (PUFs) are promising as basic primitives for authentication since they can serve as intrinsically-generated hardware roots-of-trust used in cryptography protocols. Storage (ROM) based alternatives have already been shown to be inadequate due to various attacks including invasive tampering, side channel attacks and other methods for secret key extraction [3].

PUFs are physical structures that map a digital input, also known as challenge to output, also known as response based on exploiting randomness in IC manufacturing process. In addition certain PUF circuits (strong PUFs) provide very large input-output space which makes cloning difficult.

An ideal PUF circuit is expected to exhibit high uniqueness, randomness and reliability. Uniqueness represents how distinctive the PUF responses are across various chips. The circuit structure of the PUF should be such that it enhances the variations of a particular manufacturing process. The PUF responses across various challenges should not be biased towards a particular response. If it is biased, the output would be predictable and hence easier to attack. This characteristic of the PUF indicates the randomness of a PUF. Reliability of the PUF represents the ability of the PUF to produce the same challenge-responses mapping across various environmental variations in voltage and temperature. For example a PUF circuit that is sensitive to temperature variation can produce different responses at different temperature for the same input challenge. This can lead to incorrect authentication failures.

The strength of a PUF is defined by the complexity of modeling input challenge to output response. Model-building attacks using Machine Learning (ML) for PUFs have been studied for over a decade [2][3]. But there are still no strong PUFs that are ML-resilient and remain so over a range of environmental conditions such as voltage and temperature.

In this paper we propose a PUF circuit which relies on a non-linear voltage transfer characteristics of a circuit block to create modeling attack resistance. The circuit is simple and exhibits high uniqueness.

We propose simple circuit changes to improve reliability of the circuit. We achieve this through compensation circuitry for voltage variations.

The proposed circuit was simulated in a 45nm process and we show that the proposed PUF circuits are highly ML resistant and with the reliability modifications, highly robust.

The paper is organized as follows. In Section II we discuss the background and related works. In Section III and IV we present the proposed circuit and technique to enhance the reliability of the circuit. Section V contains the analysis of various metrics and comparison to other PUF circuits. We conclude the paper in Section VI.

II. Background

In this section, we describe authentication in secure applications and elaborate on PUF circuits used in authentication. We further discuss machine learning based attack and related ML resistant PUFs.

A. Secure authentication

Secure electronic systems are widely used for various applications such as smartcards and payment systems. Cryptographic algorithms play a vital role in enabling security of electronic systems. The cryptographic algorithms are dependent on hardware roots-of-trust for authentication. PUF circuits and ROMs are two of the most discussed hardware primitives for authentication. Among them, ROM based

This research has been supported in part by grant 1421352 from the National Science Foundation .

authentication has some documented weaknesses as described below, which is why we focus on PUF circuits in this paper.

B. ROM based authentication

A ROM may be used for storing responses to challenges. Such storages have been shown to be vulnerable to *invasive attacks, side-channel attacks* and other *read-out attack* [3]. Since ROMs have finite storage capacity, they are also vulnerable to *cloning* and *replay* attacks if used directly. Increasing tamper resistance is possible but it also increases the area and power of the circuit. Therefore, we focus on PUF circuits.

C. PUF based authentication

PUF circuits provide an alternative technique for secure authentication. PUF circuits exploit variations in manufacturing process to create responses to challenges that are *unique* to each chip. Examples of manufacturing variations include variations in transistor channel lengths, widths, threshold voltage and interconnect resistance, capacitance and inductance. Such variations are captured as analog or digital information to generate responses to challenges.

The primary advantage of PUF circuits is the fact that reproducing variations is statistically difficult. This is due to the fact that many such variations are random and beyond the influence of manufacturers. Invasive techniques to extract any information from PUF circuit can destroy or alter the circuit characteristics. In addition, layout technique such as extra metallization and routing of interconnects can be employed to further increase the resistance to such *invasive attacks* [1].

PUF circuits can be widely classified as *weak PUF* and *strong PUF* based on their physical implementation [2]. Weak PUFs such as SRAM PUFs have limited challenge-response pairs (CRPs). Due to the limited CRPs, weak PUFs are susceptible to *cloning attacks* and *replay attacks* if used directly. Hence it is more suitable for key generation where the key should not leave the device. In contrast, strong PUFs have exponential number of CRPs which makes cloning and replay attacks impractical. Hence we focus on strong PUFs in this work.

D. Strong PUF circuits

Creating statistical delay variations that are unique to each chip was one of the earliest proposed PUF concept [1]. In such a PUF, delay components are connected in accordance to a challenge applied, creating challenge-response mapping. As the delay components are statistical in nature, each chip produces a unique challenge-response space. This circuit is known as delay based arbiter PUF [1]. Similarly PUF circuits based on variation in static voltage and current signals have also been proposed [4][5]. Further details on creating a secure system using PUF circuits can be found in previous publications [11]

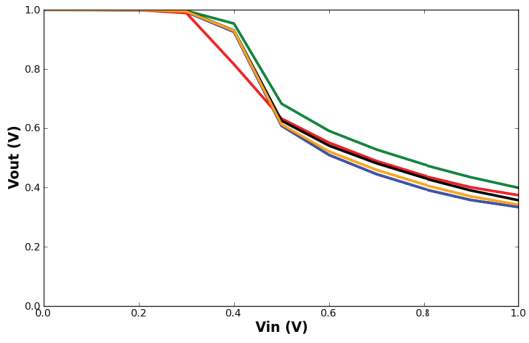

Fig. 1. Non-linear Voltage Transfer Characteristic under process variation

E. Modeling attacks

As mentioned above it is practically impossible to manufacture two exact PUF circuits. Despite this fact, machine learning based modeling attacks have exposed the vulnerability of PUF circuits [2][3]. A machine learning model trained with a certain number of responses from PUF circuits, can predict the future PUF response with high degree of success. Arbiter PUFs were initially shown to be vulnerable to ML attacks [3]. Digital modifications were proposed to increase the machine learning resistance but machine learning techniques such as Support Vector Machine (SVM), Logistic Regression and Evolutionary Strategies have been used to mount attacks with increasing success [2].

F. Machine learning resistant PUFs

Kalyanaraman *et al.* have proposed a machine learning resistant PUF based on non-linear operation of leakage current of MOSFETs [4]. Their proposed circuit relies on differences between two arrays of transistors which are in sub-threshold region to generate responses. Leakage current's exponential dependence on supply voltage and temperature is well known. Hence these circuits have reliability issues with variations in temperature or supply voltage. Kumar *et al.* have presented a circuit that relies on non-linear current mirrors to generate machine learning resistant PUF [5]. The current sources used in the simulation were assumed to be ideal current sources which in practical circuit can experience voltage and temperature variations. So the impact of using ideal current sources for simulation on the reliability metric is not clear.

This motivates the need for further investigation into design of modeling attack tolerant PUF circuits that are also robust with respect to variations in environmental conditions.

III. Proposed PUF Circuit

In this section, we describe the proposed circuit and discuss its machine learning resistance.

A. Main idea

The traditional delay based PUFs rely on delays of two paths to create challenge response set [1]. The total delay in each path is sum of delays of each delay element. The delay of an element does not directly affect the delay of any other element in the path. This linear delay model of the PUF circuit makes it vulnerable to machine learning attacks. In our case, we aim at creating a PUF by cascading circuit blocks which have a non-linear Voltage Transfer characteristics (VTC). The basic idea is that, as the input and output of each block are voltage signals and as each block has non-linear VTC, cascading them creates a complex input-output mapping.

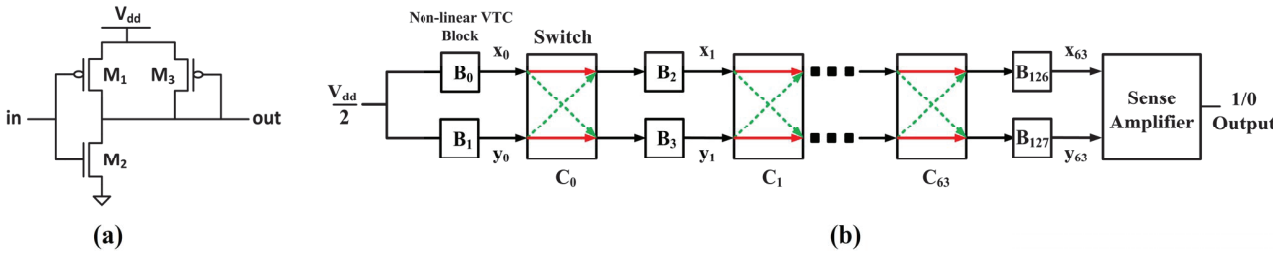

Fig. 2. Proposed circuit. (a) Non-linear VTC block and (b) Complete circuit diagram of a 64-bit PUF

For example a non-linear VTC for the basic block is shown in Fig. 1. The multiple plotlines in the figure represent VTC under different process variation corners. The VTC shown in Fig.1 can be realized by a simple 3-transistor circuit shown in Fig. 2(a). For our discussion let us assume that the supply rails are at V_{dd} and 0V. In the circuit, the transistors M_1 and M_2 act similar to an inverter. The transistor M_3 acts as a feedback transistor whose gate is connected to the node *out*. The PMOS M_3 ensures that the VTC curve does not saturate to 0V when input voltage nears supply voltage V_{dd}. If the VTC curve saturates to 0V, the VTC would be similar to an inverter and cascading multiple blocks would saturate the final output to either 0V or V_{dd}. For example, as input voltage tends to V_{dd}, the output would decreases towards 0V due to the inverter transistors M_1 and M_2 but the current through PMOS M_3 increases as the output voltage tends towards 0V, thereby increasing the output voltage. This VTC is similar to VTC of pseudo-NMOS circuit but using the feedback transistor M3 along with inverter gives better control of slope of the curve.

The VTC of the circuit is sensitive to process variations occurring in the transistors. For example, the variation of the VTC curve under different process variation instance is shown in Fig. 1. Both the slope and shape of the curve vary with process variation. When such blocks are cascaded, final output becomes highly sensitive to the process variation in each block.

The complete circuit of the proposed PUF is shown in Fig. 2(b). The circuit presented has a 64-bit input challenge and a single bit output. Each block consists of the three transistor circuit shown in Fig. 2(a). The outputs of a pair of such blocks are connected to a 2-input switch. For example for stage *i*, the output of the pair of blocks are x_i and y_i. Depending on the challenge input C_i, the outputs x_i and y_i are connected to inputs of the blocks in stage *i+1*. The switches are created by simple transmission gate based circuit. Thus by cascading these blocks a PUF circuit with input challenge bits of any length can be created. The input node for the blocks in first stage is connected to *Vdd/2*. Such input signal can be easily created with voltage divider circuit. Variation in creating the initial input signal does not affect the normal operation of the PUF circuit. The differential output of the blocks in last stage is measured to create a single bit output signal. A voltage sense amplifier is used to determine the final output. For example if the differential output at final stage is positive the output of sense amplifier resolves to logic 1 and vice verse.

As expressing the output in closed form equation is tough, we present a graphical illustration of the PUF characteristic. Consider the Fig. 3, in which the differential output x_i-y_i is plotted for each stage *i* for a 64 stage PUF circuit (the plotlines are displayed as continuous lines for better readability). The differential output at each stage evolves in a complex manner varying each stage. The two different plotlines represent two different process variation instances for the same challenge input. The dissimilarity between the two plots represents the dependence of the PUF circuit on process variation. As a result the sense amplifier creates a response of logic 0 for one process and logic 1 for another. The complex evolution of this differential signal increases the machine learning resistance and sensitivity of process variation of the circuit.

B. *Experimenal Settings*

In this section we describe the core experimental settings used in this paper. The circuit simulation platform is 45nm predictive technology model [8]. The process variation is modeled as threshold voltage variation with a normal distribution consistent with ITRS [9]. The circuits were operated at nominal supply voltage of 1V and temperature of 25 C.

C. *Machine Learning Resistance*

Several machine learning techniques such as logistic regression, evolutionary techniques and support vector machines (SVM) have been used to attack PUFs [2][3]. In this paper we use SVM due to their favorable property in modeling non-linear problems. Support vector machines are non-probabilistic, linear classification technique for binary classification problems. *SVMlight* machine learning tool was used in our experiments [7]. The SVM tools rely on choosing appropriate kernels depending on the problem. In our case we have chosen radial basis function RBF kernel as it is more suited to model non-linear problems [4].

In order to model the PUF circuit for SVM machine learning, the parity vectors may have to be derived [3]. As the switch selection architecture of our circuit is similar to a traditional arbiter PUF, the parity vector derivation remains similar as in arbiter PUF. The mapping of sample space to vector space have been derived in detail in previous publications and are omitted her for sake of brevity [3]. The machine learning resistance of the proposed circuit is shown in Fig. 4. The reduction in prediction error with the number of training samples is shown. The training set (challenges and collected responses) were chosen randomly. To estimate the prediction error, a set of 50,000 randomly chosen challenges were used. For comparison, prediction error of a 64-bit arbiter PUF is also plotted [1]. From the figure it is evident that the proposed PUF is orders of magnitude more resistant to

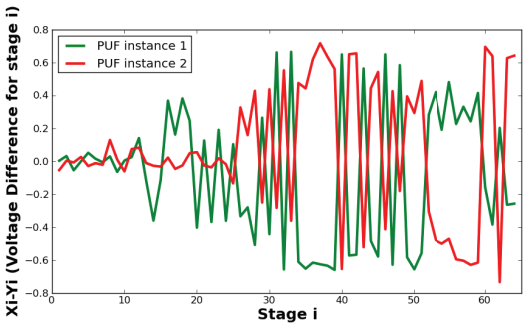

Fig. 3. Voltage difference of VTC blocks at each stage in a 64-bit PUF

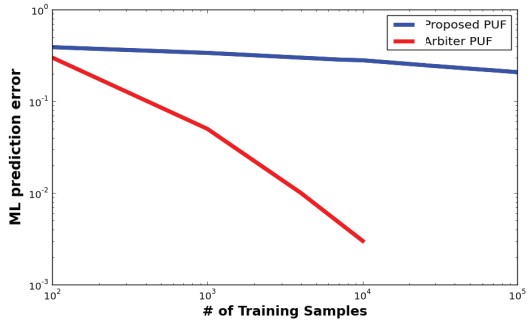

Fig. 4. Machine learning resistance of proposed PUF and arbiter PUF [1]

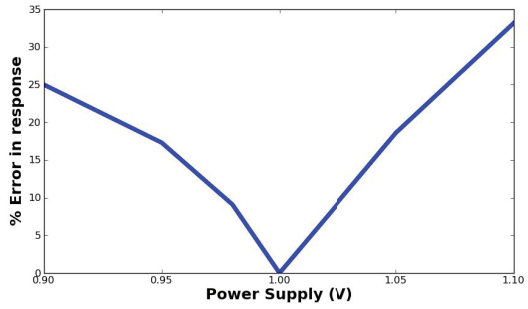

Fig.5.Response errors due to voltage variation

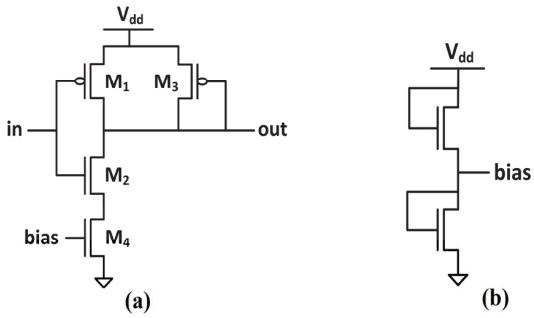

Fig.6. (a) Circuit modification to enhance reliability and (b) bias circuit

modeling attack than delay based arbiter PUF. Even with a training set of 100,000 samples the prediction error is as high as 20.8 %.

From the above results and discussion, the PUF circuit displays excellent improvement in machine learning resistance which was the primary design motivation. Other metrics to assess uniqueness and reliability are presented in next sections.

IV. Reliability Enhancement

In this section we evaluate the reliability of the proposed circuit and present a reliability enhancement technique to compensate for the response errors due to supply voltage variations.

A. Reliability evaluation

PUF circuits are expected to provide a stable response for the challenges over a range of temperature and supply voltage variation. If the circuit is sensitive to temperature or voltage noise it can result in authentication failure. The circuit proposed above relies on non-linear VTC block that are cascaded. As the shape of VTC of circuits is sensitive to supply voltage, any supply variation can impact the circuits operation. Also due to the non-linearity described above, the voltage variation can reduce the reliability of the circuit considerably. For example, the percentage of errors with voltage variation is shown in Fig. 5. To generate the data, a random set of 100 challenges were used and their responses were collected over the range of supply voltage of +/- 10% variation. The responses were then compared to the ideal responses which were characterized at supply of 1V and at 25C. As shown in the figure, the reliability of the circuit with supply voltage variation is low. The error is as high as 33% for supply variation of 10% of nominal supply voltage. A simple circuit change to correct for the noise is discussed below.

B. Reliability enhanced circuit

Consider the circuit shown in Fig. 6(a). This circuit is similar to the circuit in Fig. 2(a) except the extra footer transistor M_4. The gate of the NMOS M_4 is connected to a bias signal which is linear function of V_{dd}. The bias signal can be easily generated by a simple resistive divider shown in Fig. 6(b). In original circuit whenever there is a drop in supply voltage, the output voltage also reduces (in comparison to output under nominal V_{dd}). In order to compensate for the drop at the output, the footer transistor is added. As the bias signal reduces with V_{dd}, it increases the resistance of transistor M_4 thereby stabilizes the output of the block through negative feedback. The bias generation circuit is sized to reduce the impact of process variation. The bias generation circuit can be common to all the blocks or can be spread out throughout each block. A centralized bias generator would reduce the efficiency to correct for high frequency supply noise but would be easier to control the process variation impact. Also it will result in lower area and power. We name the modified circuit as reliability enhanced/compensated circuit.

The percentage error with temperature and voltage variation for an instance of the reliability enhanced circuit is shown in Fig. 7. The reliability has significantly improved with a maximum error rate of only 4% even for voltage variation as high as 10% and industrial temperature range of 0 to 85C. Even though we have not compensated for temperature variation explicitly, the circuit inherently exhibits tolerance. In comparison, arbiter PUF has been demonstrated with error of 4.8% over half the temperate range of our simulation [1]. More analysis of reliability with Intra Hamming distance metric is

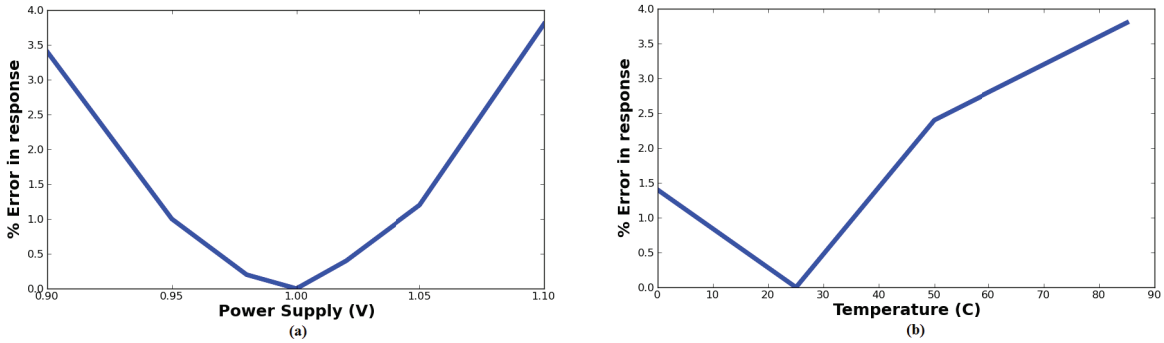

Fig. 7. Reliability of modified circuit: (a) Response errors with supply voltage variation and (b) Response errors with temperature changes

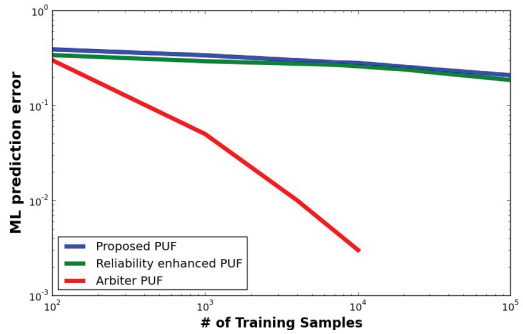

Fig. 8. Machine learning resistance of Reliability enhanced PUF

presented in next section. Thus simple feedback changes in circuit can be used to enhance the reliability. In Fig. 8, the machine learning resistance of the reliability-enhanced circuit is shown along with the unreliable circuit and arbiter PUF circuit. The machine learning resistance of the reliability-enhanced circuit has reduced negligibly and is significantly better than arbiter PUF.

V. Analysis of PUF Properties

A. PUF Metrics

In previous sections, the machine learning resistance and reliability were analyzed. In this section we present results of other PUF metrics for the proposed circuit. The reliability-enhanced circuit is used for all the results presented here on. Uniqueness, uniformity and reliability metrics are plotted in Fig. 9 and the average values are tabulated in Table 1.

1) Uniqueness:

A PUF design should create different response in each chip instance. This property is known as uniqueness and we use inter-class Hamming Distance (HD) as a metric to assess the uniqueness. Inter-class HD is defined as [6][5] :

$$d_{inter} = \frac{2}{m(m-1)} \sum_{p=1}^{m-1} \sum_{q=p+1}^{m} \frac{Hamming\ Distance(R_p, R_q)}{k} \quad (1)$$

where m and k are the number of PUF instances and number of challenge bits used respectively. R_p and R_q represent responses from a pair of PUF instances. 100 different PUF instances with 10000 randomly chosen challenge bits were used to evaluate the uniqueness.

2) Uniformity:

Uniformity measures the ratio of zeros (or ones) to total bits measured for multiple challenges. Bias towards one or zero reduces the randomness and makes the output predictable and easier to attack. Ideally number for zeros and ones should be equal, with ideal uniformity of 0.5. We evaluate the uniformity with 10,000 different challenges over 100 different PUF process instances. The histogram and results are presented in Fig. 9 and Table 1 respectively.

3) Reliability:

In previous section only one instance of PUF circuit was used to evaluate the reliability of temperature and supply voltage variation. Here we analyze the reliability for the compensated circuit in detail. The PUF circuit was first characterized at supply voltage of 1 V and temperate of 25 C. The responses for 100 random bit sequences were collected. 25 different operating conditions were simulated by varying the supply voltage between 0.9 to 1.1 volt and temperature in the set {0, 25, 50, 75, and 85} and the responses were collected. This process was repeated for 100 different PUF instances. The reliability is calculated with the metric intra class Hamming distance which is defined as [6][5]:

$$d_{intra} = \frac{1}{s} \sum_{j=1}^{s} \frac{Hamming\ Distance(R_i, R'_{i,j})}{k} \quad (2)$$

where s is the number of environmental corners considered (25 in our case) and k is the number of challenges (k=100 in our case). R_i is the k-bit response at ideal operating condition and R_{ij} is the k-bit response for corner j. For the wide operating condition, the circuit has an excellent intra-class Hamming distance of 0.021 (ideal value is 0.0).

The distribution and ideal values for the above metrics discussed are plotted in Fig. 9 and Table.1 respectively. From the distribution and table we can conclude that the PUF circuit exhibits high uniqueness and reliability. In our experiments we

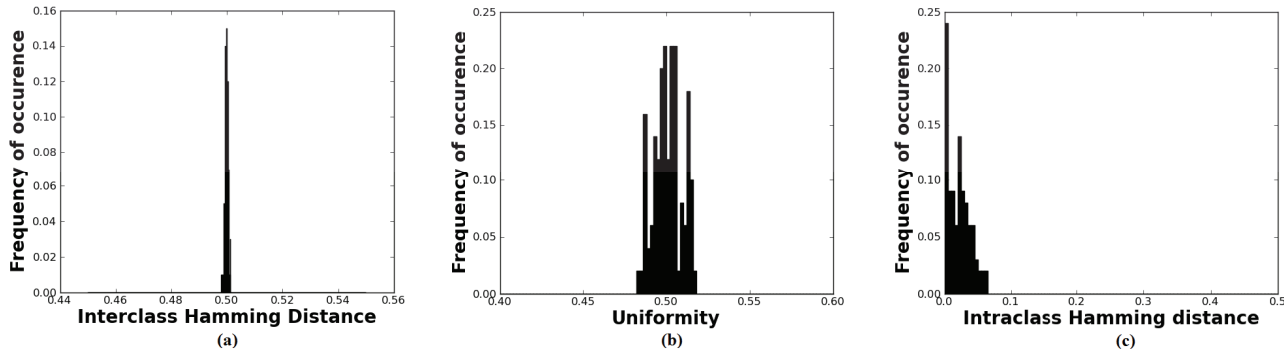

Fig. 9. Histogram of PUF metrics: (a) Uniqueness, (b) Uniformity and (c) Reliability

TABLE I. PUF METRICS

Metric	Ideal value	Mean value for proposed circuit
Uniformity	0.5	0.501
Inter-class HD	0.5	0.498
Intra-class HD	0.0	0.021

have assumed that the sense amplifier's process variation is minimal. High bias in sense amplifier reduces the uniformity but special comparator circuit based on offset cancellation to tackle this issue has already been used [10].

B. Other metrics

In this section we present other metrics evaluated on the proposed circuit. The machine learning resistance of the circuit was discussed in previous sections. With 100,000 CRPs for training, the prediction error is as high as 21% for the 64-bit proposed PUF. This is comparable to previously published PUF which uses 80 stages and has 30 % error rate for 100,000 CRPs [5]. The proposed circuit consumes 100 µW of dynamic power operating at a frequency of 100 MHz. Similar to [5], this circuit also consumes static current. Power gating can be employed to minimize static current as authentication is infrequent. Accurate area estimation based in silicon implementation is part of proposed future work. However, for reference, each stage of the proposed circuit features 20 transistors in contrast to 32 transistors in [5].

VI. CONCLUSION

Physical Unclonable Function (PUF) circuits enable chip authentication. PUF circuits been shown to be vulnerable to modeling attack using machine learning techniques. Previously researchers have proposed machine learning tolerant PUF designs that have certain weaknesses, such as assuming existence of ideal current sources or ideal operating conditions motivating the need for further investigations. We propose a novel PUF circuit which relies on a non-linear voltage transfer characteristic to improve machine learning attack resistance. To improve reliability against power supply noise, we further propose a simple circuit alteration that is shown to be highly effective. Simulation results indicate excellent PUF properties including uniqueness, reliability and high tolerance against machine learning attacks.

REFERENCES

[1] J. W. Lee *et al.*, "A Technique to build a Secret Key in Integrated Circuits for Identification and Authentication Applications," in Proc. Symposium on VLSI Circuits, 2004, pp. 176–179.

[2] U. Ruhrmair *et al.*, "Modeling attacks on Physical Unclonable Functions," in ACM conference on Computer and communications security, 2010.

[3] D. Lim, "Extracting Secret Keys from Integrated Circuits," Master's thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2004.

[4] M. Kalyanaraman, M. Orshansky, "Novel strong PUF based on nonlinearity of MOSFET subthreshold operation, " in IEEE Hardware Oriented Security and Trust, 2013.

[5] Kumar, R.; Burleson, W., "On design of a highly secure PUF based on non-linear current mirrors," Hardware-Oriented Security and Trust (HOST), 2014 IEEE International Symposium on, 2014.

[6] A. Maiti, *et al*, "A Systematic Method to Evaluate and Compare the Performance of Physical Unclonable Functions", IACR ePrint 2011/657, November 2011.

[7] T. Joachims, "Making large scale SVM learning practical," 1999.[Online]. Available: http://svmlight.joachims.org

[8] Y. Cao, "Predictive technology model." [Online]. Available: http://ptm.asu.edu/

[9] ITRS, "International technology roadmap for semiconductors."[Online]. Available: http://public.itrs.net

[10] Yeung, J.; Mahmoodi, H., "Robust Sense Amplifier Design under Random Dopant Fluctuations in Nano-Scale CMOS Technologies," SOC Conference, 2006 IEEE International , Sept. 2006

[11] Gassend, et al, "Controlled physical random functions," Computer Security Applications Conference, 2002. Proceedings. 18th Annual , 2002