A Large Scale Characterization of RO-PUF

Abhranil Maiti, Jeff Casarona, Luke McHale, Patrick Schaumont
Electrical and Computer Engineering Department
Virginia Tech
Blacksburg, VA, USA
email : { abhranil, jeffC, lam07, schaum }@vt.edu

***Abstract*—To validate the effectiveness of a Physical Unclonable Function (PUF), it needs to be characterized over a large population of chips. Though simulation methods can provide approximate results, an on-chip experiment produces more accurate result. In this paper, we characterize a PUF based on ring oscillator (RO) using a significantly large population of 125 FPGAs. We analyze the experimental data using a ring oscillator loop delay model, and quantify the quality factors of a PUF such as uniqueness and reliability. The RO-PUF shows an average inter-die Hamming distance of 47.31%, and an average intra-die Hamming distance of 0.86% at normal operating condition. Additionally, we intend to make this large RO frequency dataset available publicly for the research community.**

I. INTRODUCTION

An on-chip Physical Unclonable Function (PUF) provides hardware trustworthiness. It maps a set of inputs/challenges to a set of outputs/responses based on the random behavior of the logics and interconnects due to manufacturing process variation. A PUF can produce chip-unique signature based on the static process variation imprint that varies randomly from one chip to the other. It is an emerging secure primitive that can solve several security issues such as Intellectual Property (IP) protection, cryptographic key generation, and chip authentication.

However, generating a PUF signature that is truly random in nature as well as stable over a wide range of operating condition is a challenge. For example –

- Factors such as systematic process variation can lead to bit-aliasing. In bit-aliasing, different chips produce nearly identical (if not completely) PUF signatures that result in false positives in chip authentication.
- Dynamic variation like varying ambient temperature, thermal noise, supply voltage variation can make the PUF output unstable.

Like many PUF techniques proposed so far, a ring oscillator (RO) PUF is also affected by these issues. In this PUF, frequency variability in a group of identically laid-out ROs is exploited to map a challenge to a response [1]. The oscillation frequency of an RO loop varies with changing operating conditions. This affects the reliability of the response bits. Moreover, though the distribution of the RO frequencies is expected to be random, correlation among them exists [2]. In general, correlated or systematic process variation may lead to non-uniform PUF response if not taken care of [3,4].

Knowledge about the circuit-level behavior such as process variation pattern, variation of circuit parameters (delay, threshold voltage) over changing operating conditions will help designers to mitigate these issues. The study of on-chip variability is also helpful for general circuit design. Though simulation models can be used for this purpose, they are less accurate than on-chip measurements. This is more so because of the fact that with the increasing density of silicon devices, the nature of circuit variation is becoming more complicated.

However, to produce significant information about on-chip variation, a large scale experiment is required. In this paper, we present chip variation data using a large group of a commercial 90nm FPGA chip. The dataset is collected in terms of ring oscillator frequencies over a sample size of 125 FPGA chips. We analyze the dataset to show how the variation affects the functionality of an RO-PUF. To our knowledge, no PUF experiment of this size has been reported so far in the related literatures. The main contributions of this paper are –

- Measurement of a significantly large set of chips to characterize the effect of on-chip variation on RO-PUF.
- Analysis of the dataset in order to show the effect of circuit-level variation on RO-PUF. We describe a delay model of the RO loop, and show how the uniqueness and the reliability of the RO-PUF are influenced by the variation of the oscillation-loop delay.

Based on the experimental results, the RO-PUF shows an average inter-die Hamming distance of 47.31%, and an average intra-die Hamming distance of 0.86% at normal operating condition. The Hamming weight of the response bit string lies between 46% and 56%. We plan to make our dataset publicly available for the researchers. Currently, we are not aware of any such experiment whose dataset is publicly available. This is a unique effort in this respect.

The rest of the paper is organized as follows. In section 2, we discuss the background of an RO-PUF. The experimental set up is described in section 3. Section 4 presents the

978-1-4244-7812-5/10/$26.00 ©2010 IEEE

experimental results and its analysis. We also show why a large scale experiment provides better accuracy. We conclude the paper in section 5.

II. Background

In an RO-PUF, a pair of RO frequencies f_a and f_b $(a \neq b)$ are quantized through a simple comparison process to produce a response bit *r*. ($r = 1$ *if* $f_a > f_b$, $r = 0$ *otherwise*). We first discuss a delay model of an oscillator-loop, and explain how the variation in loop delay may affect the functionality of the RO-PUF. Next, we discuss the factors that determine the quality of a PUF in terms of unique identification of chips, and reliability of the response bits. In the result section, we characterize the RO-PUF by analyzing the effect of loop-delay variation on the quality factors.

A. Delay Model of RO-PUF

For a pair of ROs, *a* and *b*, let us define the following delay terms –

$$d_a = d_{AVG} + d_{PVa} + d_{NOISEa} \qquad (1)$$

$$d_b = d_{AVG} + d_{PVb} + d_{NOISEb} \qquad (2)$$

where d_{AVG} is the average delay of the RO. This quantity is assumed to be same across all the ROs on a die. d_{PV} is the delay component due to the process variation. This is static for an RO, and assumed to be constant over time (we neglect ageing effect here). This term may vary from one RO to the other. d_{NOISE} is the delay component due to the noise factor. It is a dynamic component, and changes over time. Both d_{PV} and d_{NOISE} represents delay variation, and can have positive or negative sign in the equations (1) and (2).

In a simple comparison method of quantization, d_a and d_b are compared, and the sign of the quantity $d_a - d_b$ determines the output.

$$d_a - d_b = (d_{PVa} - d_{PVb}) + (d_{NOISEa} - d_{NOISEb}) = \Delta d_{PV} + \Delta d_{NOISE} \qquad (3)$$

During the PUF enrollment, assuming $\Delta d_{NOISE} = 0$, the sign of $d_a - d_b$ is the sign of Δd_{PV} producing a stable reference response bit. At runtime, if Δd_{NOISE} changes in such a way that the sign of $d_a - d_b$ becomes different from that of the reference case, the response bit flips. This results in unstable response bits. For a stable operation of the PUF, Δd_{PV} should be maximized and Δd_{NOISE} should be minimized. Using the experimental data, we will derive these quantities.

B. Quality Factors of RO-PUF

In this section, we define the factors that determine the quality of the PUF, namely, uniqueness and reliability.

1) Uniqueness: The uniqueness of a PUF shows how unique are the signatures generated by the PUF from different chips. Average inter-die Hamming Distance (HD) of the PUF signatures is an estimate of the uniqueness property. With two different chips, *u* and *v*, having *n*-bit responses R_u and R_v respectively, the average inter-die HD for a group of *m* chips is defined as –

$$\frac{2}{m(m-1)} \sum_{u=1}^{m-1} \sum_{v=u+1}^{m} \frac{HD(R_u, R_v)}{n} \times 100\%$$

The above expression includes all possible pair-wise HDs among m chips. For a truly random PUF output, it should be close to 50%.

To check if the PUF response is uniformly distributed or not, we calculate percentage Hamming Weight (HW) of the response bit string. For an *n*-bit response from a chip i, it is defined as $\left(\frac{1}{n}\sum_{t=1}^{n} r_{i,t}\right) \times 100\%$ where $r_{i,t}$ is the *t*-th binary bit in an *n*-bit response of the chip i. For a uniformly distributed response, it should be 50% of the total number of response bits.

Furthermore, bit-aliasing may occur across the responses from different chips. For example, if the *t*-th bit in an *n*-bit response has same binary value across all the chips in comparison, the inter-die HD for the *t*-th bit will be zero. In that case, even if the HW of the response is 50%, the inter-die HD may yield a low value. To check the bit-aliasing, we calculate percentage HW for each of the bit positions in the response across all the chips. For the *t*-th bit, it is defined as $\left(\frac{1}{m}\sum_{i=1}^{m} r_{i,t}\right) \times 100\%$ where $r_{i,t}$ is the *t*-th binary response bit in an *n*-bit response of the chip *i* in a group of *m* chips. A value close to 50% confirms lower rate of bit aliasing.

2) Reliability: Reliability quantifies the change in PUF outputs over varying operating conditions. An *n*-bit reference response (R_i) is extracted from the chip *i* at the normal operating condition. The same *n*-bit response is extracted at a different operating condition (different ambient temperature or different supply voltage) with a value R'_i. *x* samples of R'_i is taken for each of the operating conditions. The reliability is estimated as the average intra-die Hamming distance i.e. *HD(R , R')* over *x* samples. For the chip *i*, it is defined as –

$$\frac{1}{x} \sum_{y=1}^{x} \frac{HD(R_i, R'_{i,y})}{n} \times 100\%$$

where $R'_{i,y}$ is the y-th sample of R'_i. The responses being compared are produced from the same chip. Hence, we call it intra-die HD. A *lower* value of average intra-die HD results in a more reliable PUF response.

We also estimate the total number of distinct response bits that flipped at least once in the sample measurements. This gives us an estimate about the worst-case reliability value. In our result section, we present the reliability data for the RO-PUF for a range of varying temperature and supply voltage.

III. Experimental setup

This section describes the experimental set up used for the data collection. The main emphasis in designing the experimental setup was to collect as much useful data as possible while keeping the runtime of the experiment within a fairly low limit. This was required because we measured the FPGAs of a large group of graduate and undergraduate students in our university. The complete experimental set up is shown in Fig 1.

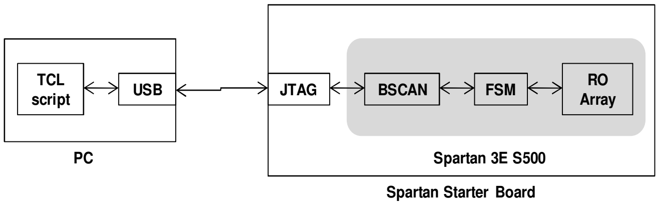

Figure 1. Experimental Setup.

The hardware portion of the set up is implemented on a Spartan 3E starter board, and the software part is implemented in a PC. A laptop PC is used to make the measurement setup portable.

A. *Hardware*

An array of 512 ROs has been implemented on the Spartan3E S500 FPGA. The ROs are placed in a 16×32 array in the middle of the FPGA fabric. Individual ROs have five inverting stages implemented using Look Up Tables (LUTs). One of the inverting stages is a 2-input NAND gate with one of the input used an enable signal for the RO oscillation (Fig 2). One complete RO loop is implemented inside a Configurable Logic Block (CLB) to maintain local routing among the inverting stages.

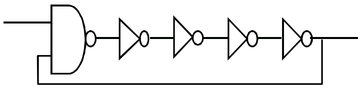

Figure 2. Basic five-stage ring oscillator loop.

To identically configure all the ROs, they are created as a hard macro, and instantiated multiple times during synthesis. During the measurement, a single RO is activated at a time using a decoder and the enable signal. The output of an RO is selected using a multiplexer, and fed to the clock input of a 32-bit counter that counts the RO oscillation. During the period when the enable signal is high, in another 32-bit counter, we simultaneously count the clock periods of an on-board 50MHz crystal oscillator as a reference. Using the total number of cycle counts in the reference clock counter, we calibrate the frequency of each RO as $((x\times50)/y)$ MHz where x is the cycle counts of the RO, and y is the cycle counts of the reference clock. For validation purpose, we measured few of the ROs using oscilloscope, and the results matched very closely.

The counter outputs are sent back to the PC using the Boundary Scan (BSCAN) module of the FPGA through the JTAG pins. A shift register is used to serialize the data being transferred through the BSCAN module. A simple finite state machine (FSM) controls enabling/disabling the ROs, loading the shift register, and clearing the counters. The FSM is controlled by the software implemented in the PC.

The hardware is designed using Xilinx ISE 11.1, and validated through simulation using Modelsim XE tool.

B. *Software*

In the PC, a TCL script runs to interact with the hardware. It configures the JTAG channel using USB port, and sends instructions to the FSM for enabling/disabling ROs, reading the counter values, and resetting the counters. The enabling period of the ROs can be dynamically changed using the script. The script stores the frequency values in a data-file for subsequent analysis.

During the measurement, the bit-stream is downloaded onto the FPGA, and the TCL scripts send its command to the hardware for measurement. To measure 100 samples of RO frequencies for an array of 512 ROs, the total period of time spent is less than 2 minutes.

C. *Large scale data collection*

In our university, most of the students in Computer Engineering major are required to buy a Spartan 3E Starter Board for course-works and projects. According to our estimate, a group of nearly 200 students use the board in a semester. We planned to measure the FPGAs on their boards, and use the data for our analysis. We publicized the experiment plan through a webpage and through emails with the technical details and the purpose of the experiment. In eight different sessions (each two hours long), we were able to measure 125 FPGAs in the computer engineering lab of our department where students do their project works. This experiment made use of the large amount of resources that would have been unused otherwise. We plan to share this data with the research community through the internet. We believe this dataset can be significantly useful in different research directions. For example –

- Ring oscillator data are used in modeling process variation [2, 5]. However, unavailability of larger sample is an issue. Sedcole et. al. mentioned the limitation of smaller sample size in their variability study [2]. Our dataset can be used for studying several on-chip variability issues such as systematic variation, layout–dependent circuit behavior. We plan to study the effect of spatial dependency of RO frequencies on PUF.
- Different methods of PUF entropy extraction can be studied based on this dataset. For example, using the frequency difference of RO pairs, rather than the sign of their difference, may increase the entropy extracted per RO pair. Using our dataset, this can be evaluated over a large sample of chips.
- Since each RO is measured 100 times, noise effects can be studied as well, and, for example, used to investigate new PUF error correction methods.
- We note that all of the existing, experimental PUF researches tend to use small population of chips [1, 6, 7], or older generations of chips [7]. Our dataset based on a 90 nm FPGA will help the researches significantly.

The WWW location of our dataset is -
http://rijndael.ece.vt.edu/variability/main.html.

IV. Result

In this section, we analyze the data collected from the experiment using 125 FPGAs. 100 frequency samples for each of the 512 ROs per FPGAs have been used in the analysis. These samples are taken at normal operating condition. The average cycle counts in the counter measuring the RO oscillation is 256080. The average cycle counts in the counter measuring the 50MHz clock source is 62426. Since the accuracy of the 50 MHz crystal oscillator is ±50 ppm (manufacturer's datasheet), the accuracy of the RO frequencies is (256080/62426)× (±50ppm) ≈ ±205ppm. The resolution of measurement is around $\log_2(256080) \approx 18$ bits.

A. Delay variation of the RO loop

First, we show how the average RO frequencies of each of the FPGAs are distributed. The average frequency of the *i*-th FPGA *($1 \leq i \leq 125$)* , is measured as

$$F_i = \frac{1}{512}\sum_{j=1}^{512} f_{i,j} \tag{4}$$

where individual RO frequencies, $f_{i,j}$, are measured as

$$f_{i,j} = \frac{1}{100}\sum_{k=1}^{100} f'_{i,j,k} \tag{5}$$

where $f'_{i,j,k}$ is the *k*-th frequency sample of the *j*-th RO in the *i*-th FPGA. F is the estimate of d_{AVG} in equation (1) and (2) whereas f is the estimate of the quantity $d_{AVG} + d_{PV}$. Fig.3 shows the distribution of the average frequencies of individual FPGAs. The average of the distribution is 205.1 MHz which we call the global average.

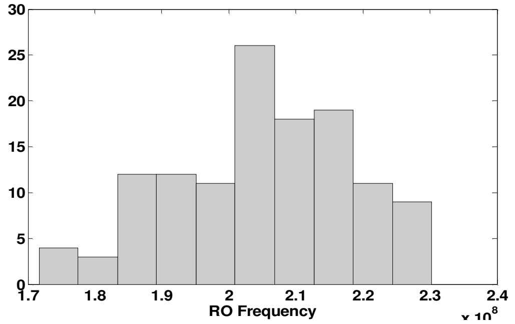

Figure 3. Distribution of average RO frequency of individual FPGAs .

The slowest FPGA has a frequency of 171.66 MHz while the fastest has a frequency of 230.24 MHz. There is no sample having a frequency beyond 230 MHz. This is due to the speed binning of the chips. The speed grade of the measured FPGAs is 4 which specifies 0.52 ns setup time for a register (manufacturer's datasheet). On the other hand, those FPGAs having frequencies around 170 MHz may have defects due to the reasons such as short-circuit, ageing etc. The standard deviation of F_i among all 125 chips is 13.54 MHz which is 6.61% of the global average. This represents the inter-die variation in terms of the average RO frequency of an FPGA.

Now, we present the data to show the intra-die variation in terms of the RO frequency. We calculate the standard deviation of the 512 RO frequencies in an FPGA. For the *i-th* FPGA, it is derived as follows –

$$\sigma_{PVi} = \sqrt{\frac{1}{511}\sum_{j=1}^{512}(f_{i,j} - F_i)^2} \tag{6}$$

Fig 4. shows the distribution of the normalized quantity *(σ_i/F_i) × 100 %*. This is an estimate of the static variation d_{PV}. Hence, we term it as σ_{PVi}.

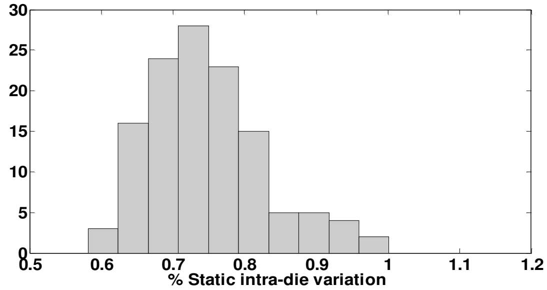

Figure 4. Distribution of static intra-die variation .

The average static intra-die variation is 0.75% with the maximum and the minimum of 1% and 0.58% respectively. For an RO-PUF, this quantity should be as large as possible in order to have a higher reliability of PUF response bits. This is because higher static intra-die variation will increase the value of Δd_{PV} in equation (3), thus increasing the reliability as Δd_{NOISE} needs to have higher magnitude to flip a response bit.

To estimate the value of d_{NOISE}, we first separately calculate the standard deviation of each RO frequency over 100 samples as follows –

$$\sigma_{i,j} = \sqrt{\frac{1}{99}\sum_{k=1}^{100}(f'_{i,j,k} - f_{i,j})^2} \tag{6}$$

Then we define σ_{NOISEi} for each FPGA to estimate d_{NOISE} in a normalized form as shown in equation (7) –

$$\sigma_{NOISEi} = \frac{1}{512}\sum_{j=1}^{512}\left(\frac{\sigma_{i,j}}{f_{i,j}} \times 100\%\right) \tag{7}$$

Fig. 5 shows the distribution of σ_{NOISEi}. It can be noticed that most of the FPGAs are centered on a value of 0.025% with few outliers. These outliers are expected to generate relatively higher number of unstable responses.

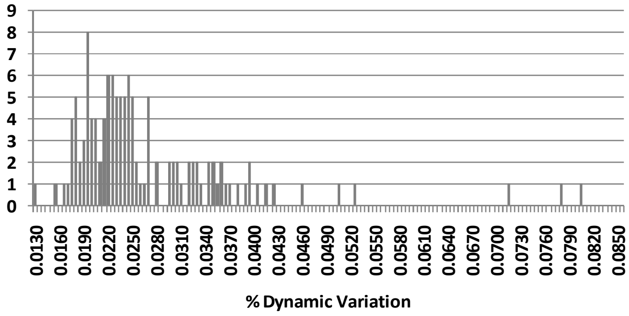

Figure 5. Distribution of dynamic variation .

The average static intra-die variation of 0.75% is significantly higher than the average dynamic variation of 0.025%. This shows that the RO-PUF is expected to be highly reliable at the normal operating condition.

B. *Why is a large scale experiment better?*

We show that the above quantities are better in accuracy compared to a smaller dataset. As a test case, we calculate the global average frequency using a group of 16 FPGAs. Using a sliding window method, we formed several groups of 16 FPGAs (1 to 16, 2 to 17 and so on) in an arbitrary order. Fig. 6 shows the plot of the scaled down averages with respect to the global average of 205.1 MHz using 125 samples. We note that the variation of smaller groups is as high as 6% over the entire set of 125 FPGAs. In other words, the variation due to a smaller population is comparable in magnitude with the inter-die variation of 6%. Clearly, a larger dataset is important to minimize this estimation error.

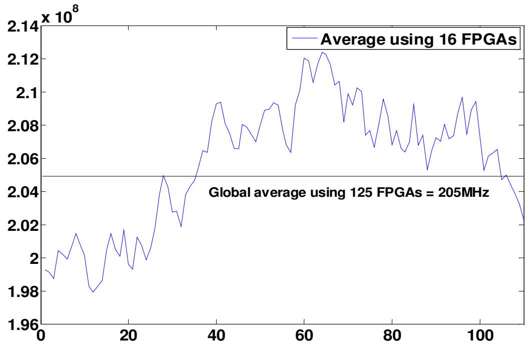

Figure 6. Comparison of measurement using large and smaller dataset.

Additionally, we checked the intra-die variation, and found similar magnitude for the measurement error.

C. *RO-PUF quality factors*

The uniqueness and the reliability of the RO-PUF is estimated based on a 511-bit key extracted from the array of 512 ROs. These 511 bits are extracted by comparing adjacent pair of ROs in the array. We used average frequency of ROs, $f_{i,j}$ (eqn(5)), for this process.

1) Uniqueness: Fig. 7 shows the distribution of the inter-die Hamming distance among the sample FPGAs. The average is 47.31% with a maximum value of 56.36% and a minimum of 38.98%.

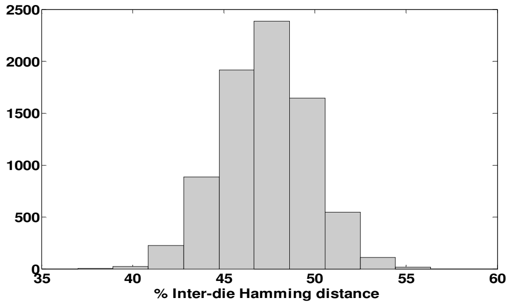

Figure 7. Distribution of the inter-die Hamming distance

Though the PUF signature is significantly long, the inter-die HD deviates from 50% value by nearly 3%. This might be due to non-uniform distribution of ‘0’/’1’ bits in the PUF response or due to bit-aliasing. Fig. 8, shows the percentage HW in the responses taken from all 125 FPGAs. The average value is 50.72%. The maximum value is 56.94% whereas the minimum value is 45.98%.

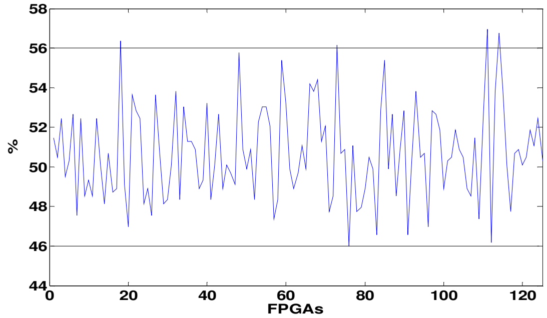

Figure 8. % Hamming weight in the PUF response.

This shows that the response bits are fairly evenly distributed among ‘0’ and ‘1’. As an estimate of bit-aliasing, Fig. 9 shows the percentage HW for all 511 bit positions across all the 125 FPGA chips. Even though, the average value is 50.72%, there are occurrences of high value of bit-aliasing at regular interval of the bit positions (the spikes in Fig 9). The maximum value is 96.8% which means almost 98 FPGAs out of 125 produced the bit ‘1’ for that particular bit position. This may cause the reduction in the inter-die HD.

Since the pair of ROs being compared in the response evaluation are selected from physically adjacent location the FPGA fabric, systematic process variation may be ruled out as the cause of this regular bit-aliasing pattern. Upon closer examination of the RO array, we found that these bits are located at two of the four boundaries of the RO-array. Moreover, these two boundaries are closer to the BRAM. One of the possible reasons might be the variation in power distribution near the BRAMs. However, it requires further investigation to reach a conclusion.

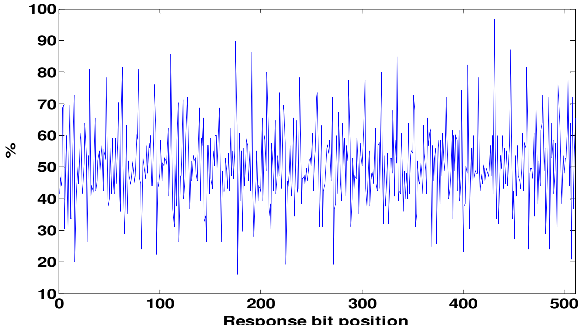

Figure 9. Bit-aliasing as percentage Hamming weight.

In any case, it is clear that there is a significant occurrence of bit-aliasing at regular interval. This type of critical observation is difficult to be made in a smaller sample size.

2) Reliability: For reliability analysis, all the 511 bits of the reference response R_i are derived using the average RO frequencies, $f_{i,j}$ *($1 \leq j \leq 512$). R_i’* is derived using $f_{i,j,k}$*’ ($1 \leq k \leq 100$).* Fig. 10 shows the distribution of the intra-chip Hamming distance among the sample FPGAs. The average is 0.86%

with a maximum value of 1.39% and a minimum of 0.38%. The low value of intra-chip HD shows that the PUF is highly stable at normal operating condition. This observation is in accordance with the result in the previous section where it was observed that the average static intra-die variation is significantly larger than the average dynamic variation predicting higher reliability in the PUF response.

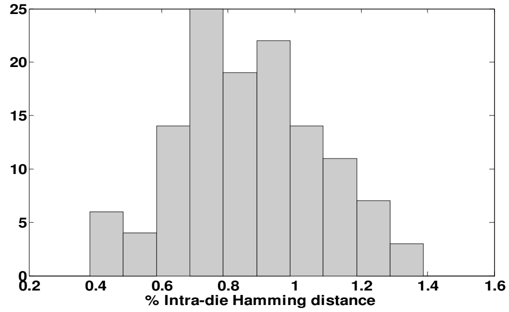

Figure 10. Distribution of intra-die Hamming Distance.

We also evaluated the number of distinct response bits that flipped at least once over the sample measurements. We call it distinct because an unstable bit may not deviate from its reference for all the sample measurements. For example, suppose in one sample measurement of a PUF, bit a, b and c out of a 100-bit long response flipped. In another sample measurement of the same PUF, bit b, c and d flipped. In both the cases, the intra-die HD is 3%. However, the total percentage of distinct unstable bit is 4% (includes a, b, c and d). This is useful as an estimate of the upper bound of the unstable bits in the PUF response. Fig. 11 shows the distribution of the percentage distinct unstable bits for the whole population of the FPGAs at normal operating condition. The average is 5.36% that is roughly 27 bits out of 511 bits. The maximum value is 7.63% while the minimum is 3.13%.

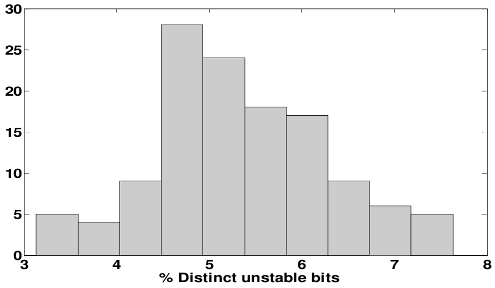

Figure 11. Distinct unstable bits at normal condition.

We also evaluated the intra-die HD at different temperatures and supply voltages for a group of *five* FPGA chips. Fig.12 shows the result. The voltage variation results in lower reliability than the temperature variation. At 0.96V, the intra-chip HD is as high as 15%. The reliability remains fairly consistent for the temperature variation. In Table 1, we show the total number of distinct unstable bits in the PUF response for both temperature and the supply voltage variation for five FPGA samples.

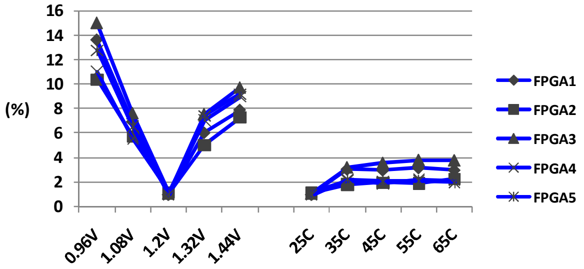

Figure 12. Intra-die HD at different temperature and core supply voltage .

TABLE I. TOTAL NUMBER OF DISTINCT UNSTABLE BITS

Voltge Variation (0.96V – 1.44V)					Temperature Variation (35C-65C)				
143	128	167	141	150	52	39	51	47	43

The overall observation is that voltage variation causes more unstable bits in the PUF response than the temperature variation.

V. CONCLUSION

In this paper, we characterized a ring-oscillator PUF over a significantly large population of FPGAs. The results shows that PUF output signatures are fairly uniformly distributed with high rate of uniqueness in terms of inter-die Hamming distance. The high ratio of static variation to the dynamic variation conforms to high reliability of the PUF output in terms of intra-chip Hamming distance. In future, we plan to continue this work in terms of expanding the scale of the experiment. Further analysis of the dataset along with testing factors such as ageing effect of the chip is also part of our future work, as well as updating our public database as we make progress in our measurement efforts.

ACKNOWLEDGMENT

This work has been supported in part by National Science Foundation (NSF) grant no 0855095.

REFERENCES

[1] G.E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," Proc. 44th Design Automation Conf. (DAC 07), ACM Press, pp. 9-14.

[2] P. Sedcole and P. Y. K. Cheung, “Within-die delay variability in 90nm FPGAs and beyond,” in Proc. IEEE International Conference on Field-Programmable Technology, Jun. 2006,pp. 97–104..

[3] Y. Su, J. Holleman, and B. Otis, "A 1.6pJ/bit 96% Stable Chip ID Generating Circuit Using Process Variations," Proc. IEEE Int'l Solid-State Circuits Conf. (ISSCC 07), IEEE Press, 2007, pp. 406-407, 611.

[4] Daniel E. Holcomb, Wayne P. Burleson, Kevin Fu, "Power-Up SRAM State as an Identifying Fingerprint and Source of True Random Numbers," IEEE Trans on Computers, vol. 58, no. 9, pp. 1198-1210.

[5] H.Onodera, “Variability: Modeling and its impact on design,” IEICE Trans. Electron, E89-C, p.342 (2006)..

[6] Chi-En Yin, Gang Qu, "Temperature-aware cooperative ring oscillator PUF," host, pp.36-42, 2009 IEEE International Workshop on Hardware-Oriented Security and Trust, 2009.

[7] Yu, H.; Leong, P.H.W.; Hinkelmann, H.; Moller, L.; Glesner, M.; Zipf, P.; "Towards a unique FPGA-based identification circuit using process variations," FPL 2009., pp.397-402