A Transmission Gate Physical Unclonable Function and On-Chip Voltage-to-Digital Conversion Technique

Raj Chakraborty, Charles Lamech
Intel Corp.
raj.k.chakraborty@intel.com,
charles.d.lamech@intel.com

Dhruva Acharyya
AdvanTest Inc.
Dhruva.Acharyya@advantest.com

Jim Plusquellic
University of New Mexico
jimp@ece.unm.edu

ABSTRACT

A physical unclonable function (PUF) is an embedded integrated circuit (IC) structure that is designed to leverage naturally occurring variations to produce a random bitstring. In this paper, we evaluate a PUF which leverages resistance variations which occur in transmission gates (TGs) of ICs. We also investigate a novel on-chip technique for converting the voltage drops produced by TGs into a digital code, i.e., a voltage-to-digital converter (VDC). The analysis is carried out on data measured from chips subjected to temperature variations over the range of -40^{o}C to +85^{o}C and voltage variations of +/- 10% of the nominal supply voltage. The TG PUF and VDC produce high quality bitstrings that perform exceptionally well under statistical metrics including stability, randomness and uniqueness.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection -- *Authentication.*

General Terms

Security

Keywords

Hardware security, unique identifier, process variations

1. INTRODUCTION

Physical Unclonable Functions (PUFs) continue to gain momentum as an alternative to embedding 'secrets' using fuses and non-volatile memory on ICs. PUFs derive secrets from variations that occur in the physical parameters of the on-chip wires and transistors. These variations are unique to each chip and, depending on the parameter, can be leveraged to produce large numbers of random bits. PUFs can produce repeatedly random bitstrings on the fly, and therefore eliminate the need for a specialized non-volatile on-chip memory to store them.

A PUF produces a bitstring by applying a set of "challenges" to specialized circuit primitives and measuring the corresponding "responses". The challenges are typically 'digital' and therefore can be generated on-chip using a pseudo-random number generator such as a linear feedback shift register (LFSR). The challenges are used to configure one or more PUF circuit primitives prior to the application of a stimulus. The stimulus elicits an analog response from the PUF primitives, which is measured and digitized by other components of the PUF circuit. The digitized responses are then compared in a variety of combinations to produce a digital bitstring.

The PUF response is analog in nature, e.g., it can be a voltage drop or the propagation delay of a signal through the PUF primitive. The analog nature of the underlying random variable make the PUF sensitive to environmental variations such as temperature and power supply noise. Several important applications of a PUF require that they produce the same bitstring for a fixed challenge. Therefore, PUF architectures must be both random and resilient to noise sources.

In this paper, we investigate a PUF primitive that leverages resistance variations that occur in transmission gates (TGs). Hardware experiments are carried out on a set of chips at 9 temperature-voltage (**TV**) corners, using all combinations of the temperatures -40^{o}C, 25^{o}C and 85^{o}C and voltages 1.08 V, 1.2 V and 1.32 V. A novel embedded test structure called a **voltage-to-digital converter (VDC)** is also evaluated under these environmental conditions. The VDC is used to digitize the voltage drops produced by the TG PUF.

Beyond these novel aspects of this work, we also investigate several noise resilient bit-flip avoidance schemes, that are designed to increase the probability that the bitstring can be reproduced under varying environmental conditions. The first technique derives a threshold from a chip's digitized voltage drop distribution profile that is used to decide whether a given comparison generates a **strong** bit or a **weak** bit. A second triple-module-redundancy (TMR-based) scheme is proposed for fixed length bitstrings that further improves bit-flip resilience. Although these techniques discard a significant fraction of bits, they provide several significant advantages. The public (helper) data associated with these methods reveals nothing about the secret bitstrings that they encode. Second, for applications where the PUF responses are made public, the difficulty of model building is significantly increased (assuming the public data is obfuscated) because bitstrings are constructed using only a subset of all possible voltage pairings. These techniques are investigated on data obtained from 63 copies of a test chip fabricated in a 90 nm technology.

2. BACKGROUND

Random bitstrings form the basis for encryption, identification, authentication and feature activation in hardware security. The introduction of the PUF as a mechanism to generate random bitstrings began in [3], although their use for chip identifiers began a couple years earlier [2]. Since their introduction, there have been many proposed architectures that are promising for PUF imple-

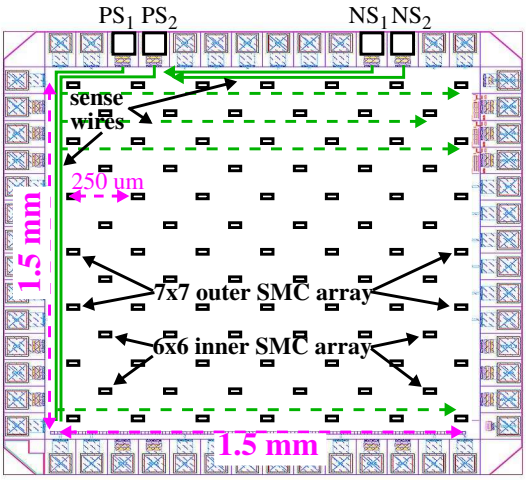

Fig. 1. Block diagram of 90 nm chips with 85 embedded stimulus-measure circuits (SMCs).

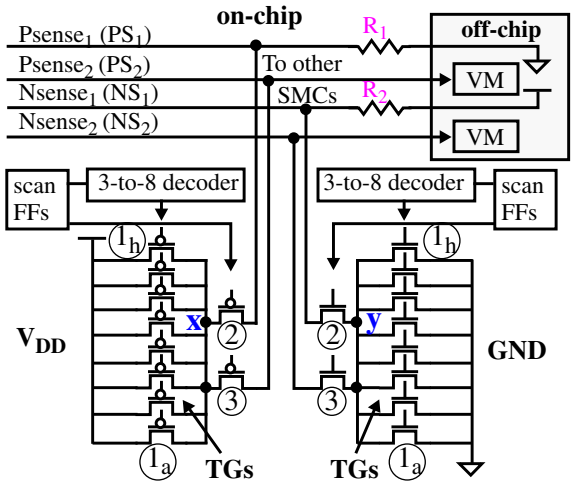

Fig. 2. SMC schematic in 90 nm chips.

mentations, including those that leverage variations in transistor threshold voltages [2], in speckle patterns [3], in delay chains and ROs [4-7], in SRAMs [8], in metal resistance [9][10], in sensors [11], and many others. The TG PUF proposed in this research is also based on resistance variations as in [10]. However, this paper for the first time investigates the reproducibility of the bitstrings across 9 industrial range TV corners after digitization using an on-chip VDC.

3. EXPERIMENT SETUP

3.1 TG Array, TGVs and TGVD

Fig. 1 gives a block diagram of the 90 nm test chip architecture. The chip padframe consists of 56 I/Os, and surrounds a chip area of approx. 1.5 mm x 1.5 mm. Four PADs labeled PS_1, PS_2, NS_1 and NS_2 refer to *voltage sense* connections, the 'P' version for sensing voltages near V_{DD} and the 'N' version for voltages near GND. These four terminals wire onto the chip and connect to 85 copies of a *Stimulus/Measure circuit* (SMC). The SMCs are distributed across the entire chip (see small rectangles) as two arrays, a 7x7 outer array and a 6x6 inner array. Although not shown, a controlling scan chain connects serially to each of the SMCs.

The schematic diagram of the SMC is shown in Fig. 2. A set of 20 'pseudo' pass gates (hereafter referred to as transmissions gates or **TGs**) serve as both the PUF primitives and voltage sensing elements. Eight of the TGs, labeled 1_a through 1_h, connect to the V_{DD} grid, as shown on the left side of Fig. 2, while the other eight connect to the GND grid. Two additional TGs, labeled as *2* and *3*, connect to the drains of the $1_{a\text{-}h}$ TGs. Separate scan FFs control their connection to the chip-wide wires that route to the P/NS_x pins shown in Fig. 1. The PS_1 and NS_1 sense wires are connected off-chip to GND and V_{DD}, resp., to create the stimulus condition described below. PS_2 and NS_2 are routed to off-chip Agilent 34401A voltmeters (VMs).

A voltage drop measurement is carried out by enabling three TGs, both of those labeled *2* and *3* and one from the group 1_a through 1_h. For example, using the PFET TGs, enabling TGs 1_a and *2* create a short between the V_{DD} grid on-chip and a GND node off-chip. The voltage falls across the two TGs as well as the PS_1 wire. The voltage on the node **x** between TG 1_a and *2* can be sensed with TG 3^1. The on-resistances of the TGs (and the resistance of the PS_1 wire) determine how much of the V_{DD} voltage falls across each of TG 1_a and *2*. Random variations in the on-resistances of the TGs 1_a through 1_h (referred to subsequently as the **stack**) produce different voltage drops as each is enabled. We refer to the voltages at node **x** as **TGVs**.

The component of the TGV that falls across the sense wires represents a bias because the length of the sense wires is different for each SMC in the array. The bias is eliminated by creating TGV differences (**TGVDs**) using the 8 TGVs measured within each SMC, separately for NFETs and PFETs. The TGVDs are obtained by subtracting pairs of TGV values. With 8 TGVs, a total of 8*7/2 = 28 TGVDs can be created in each stack. The total number of TGVDs obtained per chip is 2,380 for each of the PFETs and NFETs, obtained as 85 SMCs * 28 TGVDs/SMC.

The NFET and PFET TGVDs, in turn, can be compared under all combinations to produce bitstrings of length 2,380*2,379/2 * 2 = **5,662,020** bits. The NFET and PFET TGVDs cannot be compared with each other primarily because of channel width differences (PFETs are 2.5x wider than the NFETs) and mobility variations with doping (NFET variations are larger than PFET variations). As a consequence, PFET voltage variations are only about half as large as the NFET variations.

In our experiments, the order in which the comparisons are made is randomized using *srand(seed)* and *rand()* from the C programming library. This operation is easily implemented on chip using an LFSR and a seed.

3.2 Voltage-to-Digital Converter (VDC)

In addition to analyzing the TG voltage drops directly, we also analyze a digital representation of them that is produced by an on-chip VDC, similar to designs described in [12]. The architecture of the VDC is shown in Fig. 3. The VDC is designed to 'pulse shrink' a negative input pulse as it propagates down a current-starved inverter chain. As the pulse moves down the inverter chain, it activates a corresponding set of latches to record the passage of the pulse, where activation is defined as storing a '1'. A

1. Only a negligible amount of current flows through TG *3* to the voltmeter so the voltage on node **x** is nearly identical to that at the voltmeter.

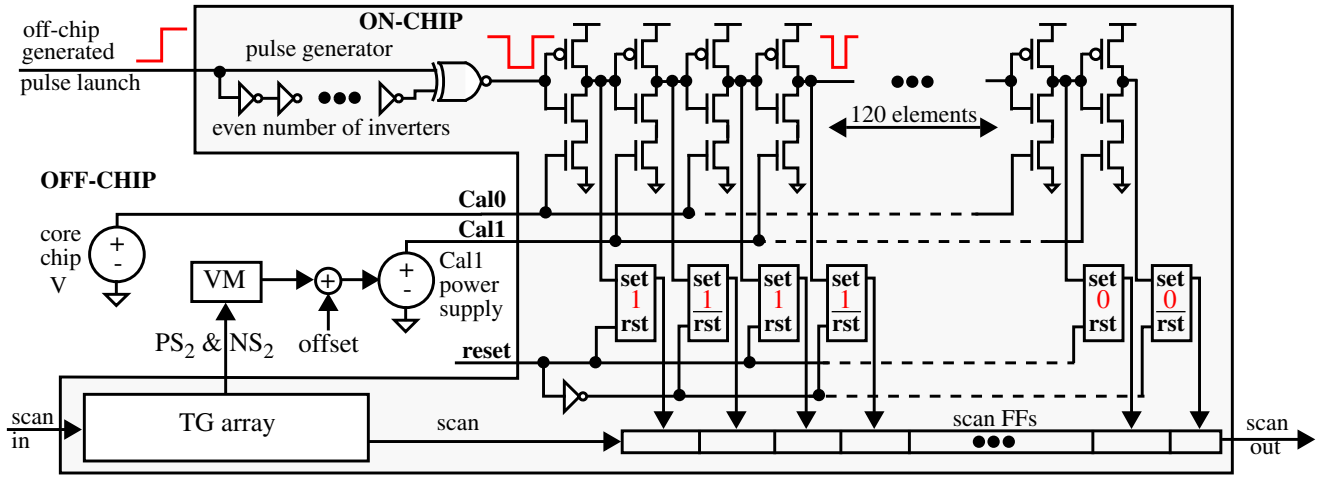

Fig. 3. Voltage-to-Digital Converter (VDC). On the left side is off-chip instrumentation that measures a voltage from the TG array, adds an offset and programs a power supply to drive the Cal1 input of the VDC.

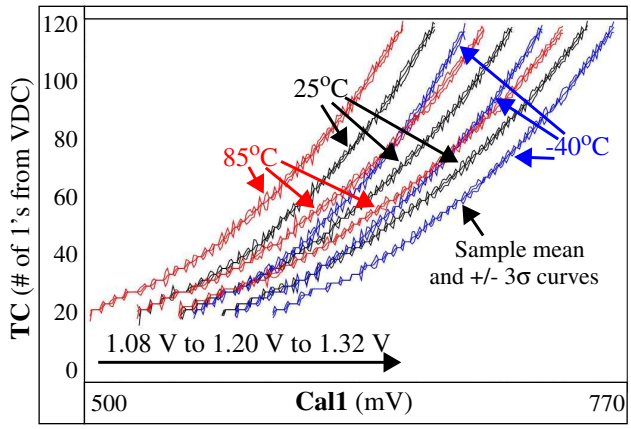

Fig. 4. VDC Cal1 vs. thermometer code (TC) curves across 9 TV corners on one chip.

thermometer code, i.e., a sequence of '1's followed by a sequence of '0's, represents the digitized voltage.

The voltage-to-digital conversion is accomplished by introducing a fixed-width (constant) input pulse, which is generated by the pulse generator shown on the left side of the Fig. 3. Two analog voltages, labeled Cal0 (which is held constant) and Cal1 (the voltage to be digitized) connect to a set of NFET transistors in the inverter chain, with Cal0 connecting to NFETs in even numbered inverters and Cal1 to the NFETs in odd numbered inverters. The propagation speed of the two edges associated with the pulse are controlled separately by these voltages. The pulse will eventually die out at some point along the inverter chain when the trailing edge of the pulse 'catches up' to the leading edge. This is ensured by fixing Cal0 at a voltage higher than Cal1. A digital representation of the Cal1 voltage can then be obtained by counting the number of '1's in the latches.

In order to enable this type of pulse shrinking behavior, Cal1 needs to be set to a value between 500 mV and 800 mV. The voltage-divider (series) arrangement of the identically-sized TGs shown in Fig. 2 should provide voltages at the midpoint of the supply voltage, e.g., approx. 600 mV. This is not the case, however, for two reasons; 1) a portion of the voltage falls across the NS_1 and PS_1 sense wires resistances labeled R_1 and R_2 in Fig. 2, and 2) the series-connected transistors in the shorting path, e.g., 1_a and 2 in Fig. 2, operate in linear mode and saturation modes, resp. (See Section S4 for details.) As a consequence, the range of the TGVs observed in our experiments at node **x** in Fig. 2 for PFETs is between 950 mV to 1050 mV, and at node **y** for NFETs is 150 mV to 250 mV. In order to move Cal1 into the 600 mV range, an **offset** voltage is added (subtracted) to the voltages measured by the VM as shown in Fig. 3 for NFETs (PFETs). This offset voltage is computed using a calibration process described below.

The calibration process is needed because the required offset voltage changes as a function of TV conditions. The curves in Fig. 4 depict the behavior of the VDC over the 9 TV corners for one chip. The graph plots Cal1 on the x-axis against the number of '1' bits in the thermometer code, referred to as **TC**, on the y-axis. The mean and 3σ curves are superimposed. The average 3σ, computed using the individual 3σ in each curve, is less than 1 for all curves. The small non-linearity in the curves does not degrade the statistical properties of the bitstrings, as shown below. The sensitivity of the VDC is approx. 1 TC bit per millivolt change in Cal1. The TGVs for a typical chip vary over the range of 40 to 60 mVs so less than half of the 120 bit range of the VDC is used in our experiments.

Although the VDC remains stable across the TV corners, the shift of the curves along the x-axis causes overflow in the VDC; a situation where the pulse propagates through all 120 delay chain elements. A calibration process is carried out that tunes the 'offset' at each TV corner, and effectively eliminates the adverse effects of the curve shift. The calibration process tests a distributed set of 9 TGs, e.g., of the 680 NFET TGs, and uses binary search to find an offset voltage that produces a 'target' TC, separately for each of the 9 tests. We set the target TCs for NFET and PFET TGVs to 65 and 85, resp. These targets worked well to prevent overflow in all of the 1,360 TG measurements, across all TVs and chips used in our experiments. The **median offset** from the 9 calibration tests is used as the offset during the subsequent data collection process. This calibration procedure only *approximates* the best offset, but does not need to be precise because the goal is only to prevent overflow in the VDC. A more detailed explanation of the process is given in Section S1.

We plan to integrate the instrumentation used to measure the TGVs, to add an offset and to control the Cal1 voltage, as shown

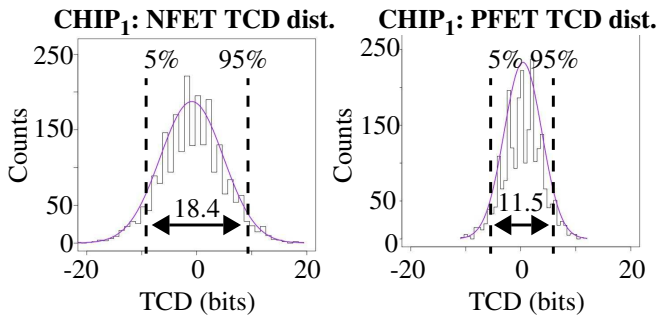

Fig. 5. Enrollment NFET (left) and PFET (right) TCD distributions with 2,380 components from one chip, with inter-percentile ranges delineated.

on the left side of Fig. 3, in the next version of the chip. The Cal1 offset voltages can be derived using a resistor-ladder network [13], and added to the TG voltage using a voltage subtractor/adder circuit [14]. The offset only needs to be accurate to approx. 5 mVs, which significantly reduces the area overhead of the ladder network. With the availability of these on-chip components, a state machine can be easily designed to carry out the calibration process described above.

3.3 Data Collection Process

The calibration process is used to select an offset voltage, separately for the PFET and NFET elements on each of the 63 chips. Each of the 680 components are then enabled, one at a time, and the corresponding TGV is measured using the VM as shown in Fig. 3. The Cal1 power supply is programmed with this TGV plus the offset and 11 TC samples are collected from the VDC. This process is repeated for both the NFET and PFET components. The mean value of the 11 samples is used to compute a 'difference' value, synonymous to the TGVDs described above. We use the term **TCD** to refer to these thermometer code differences in the remainder of this paper.

3.4 Overhead

Each SMC occupies an area of approx. 500 um^2, so the total area occupied by the array of 85 SMCs is approx. 42,500 um^2. If the SMCs are placed adjacent to each other (instead of being distributed as in Fig. 1), the array would occupy a 206 um x 206 um region. The VDC occupies an area of 136 um x 60 um. The area of the digital components, i.e., the LFSR and bit generation engine, is estimated at 300 um x 300 um. On-chip memory requirements for the array of 680 NFET and PFET TGs is approx. 2,380 bytes.

3.5 Thresholding Technique

As discussed above, TCDs are computed by subtracting TCs within the same SMC as a means of eliminating the voltage bias introduced by the sense wires. Computing differences also has the benefit of significantly increasing the number of bits that can be produced from each chip. For example 2,380 TCDs are produced from the 680 NFET TCs.

Using difference values, however, has two main drawbacks. First, subtracting two TCs reduces the signal-to-noise ratio because the noise from two separate measurements is combined in the difference. More importantly, TCDs 're-use' the base entropy of the array, which is defined by the 1,360 NFET and PFET TCs for each chip. Therefore, re-use makes model building attacks possible in cases where the bitstring is made public.

We propose a thresholding technique as a means of dealing with model-building attacks and preventing information leakage in the public helper data. Our thresholding technique discards TCD comparisons that are susceptible to producing bit flips in the bitstring. Bit flips occur when the relative ordering of a pair of TCDs defined during enrollment reverse order during regeneration. This is much more likely to occur for pairs of TCDs that are similar in magnitude. We show in our experimental results that it is possible to define a threshold that filters all TCD pairings that introduce bit flips during regeneration at one or more of the TV corners. The threshold is derived using the distribution characteristics of TCDs obtained during **enrollment**, which is carried out in our experiments at 25^{o}C and 1.20V.

Fig. 5 shows the TCD enrollment distributions for NFETs and PFETs from one of our chips. It is clear from the spread of the distributions that the NFET TCDs have more variation than the PFET TCDs as discussed in Section 3.1. The objective is to derive a threshold from these distributions that serves three primary goals: 1) avoids bit flips under different TV conditions in the subsequent bit generation phase, 2) preserves as many strong bits as possible for each chip and 3) makes the number of strong bits as consistent as possible across chips, i.e., scales with the range of variation that occurs on each chip. We define **strong bits** as those generated by TCD comparisons where the differences in the TCDs exceeds the threshold.

In our experiments, we found the limits defined by the two vertical lines labeled 5% and 95% in Fig. 5 achieve these goals. These limits capture the spread of the distribution while ignoring the outliers on the tails of the distributions, which, when included, introduce large variations in the number of strong bits preserved across the chip population, i.e., they degrade criteria 3 above. We then multiply the 2 **inter-percentile ranges** defined as the distances between these limits by 2 **scaling factors**, one for NFETs and one for PFETs, to define the 2 TCD **thresholds** for the chip.

Figs. 6(a) and (b) provide an illustration of the thresholding process applied using TCD data from one of the chips. The graphs plot bit number along the x-axis against the **differences** of the TCDs being compared. Only the first 390 strong bits are shown. The horizontal lines at 9.7 and -9.7 delineate the threshold boundaries for the NFET TCDs, which are derived from Fig. 5 using a scaling factor of 0.53.

Fig 6(a) shows those TCD differences which produce strong bits during enrollment. In addition to generating the secret bitstring, a **thresholding bitstring** is also constructed during enrollment which indicates which comparisons produce strong bits and which produce **weak bits**. The thresholding bitstring is recorded in public data storage, and using techniques such as run-length encoding, is proportional in size to the secret bitstring (see Section S3). This type of public data reveals nothing about the secret bitstring, and represents the helper data for our PUF.

Fig. 6(b) superimposes the TCD difference data points generated under the remaining 8 TV corner experiments, which represent the regeneration scenarios in our experiments. The thresholding bitstring is consulted to ensure regeneration uses the same comparisons as enrollment[1]. The data points associated with the regenerations appear above and below the enrollment data points. Only those that move toward 0 line are problematic how-

1. The thresholding process is implemented only during enrollment, and is disabled during regeneration.

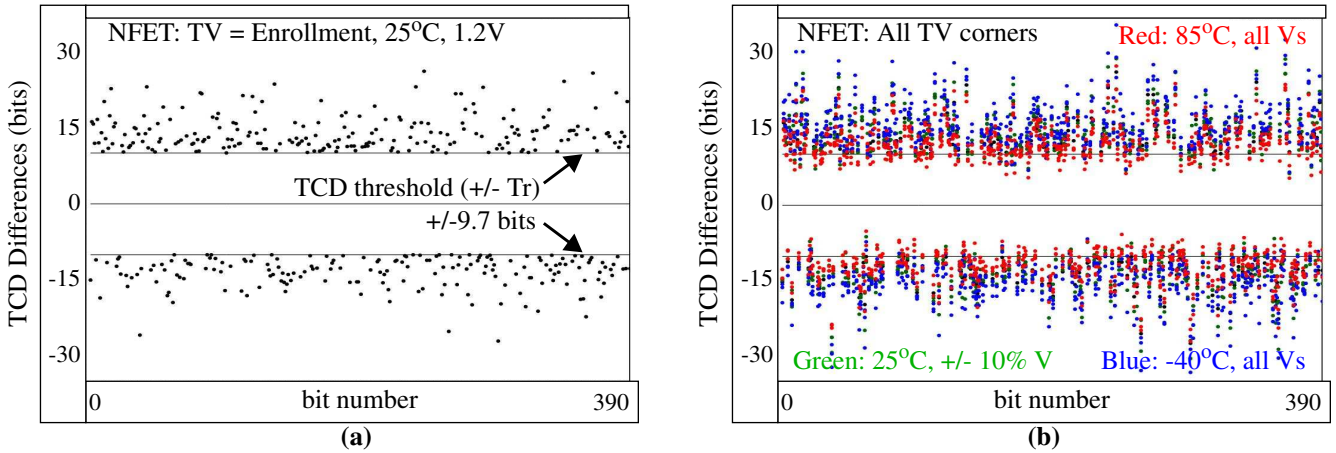

Fig. 6. Threshold method showing the first 390 strong bit comparisons during (a) enrollment and (b) regeneration across 8 TV corners.

ever. Although none occur in these plots, points that move over the 0 line from above or below indicate the relative ordering has changed in the TCD pairing. A bit flip will occur during regeneration if this condition is met.

The TCD differences plotted in the figure span a larger range than the TCDs used to compute the inter-percentile range from Fig. 5 because the TCDs themselves are both positive and negative. Despite their larger range, only about 21% of the 2,831,010 possible comparisons, i.e., approx. 595,000 bits, survive the thresholding for NFETs. A similar analysis using the TGVDs shows approx. 33% surviving the thresholding, which suggests that the digitization process adds to the noise. This is even more dramatic in the PFET analysis, where approx. 7% of the TCDs survive and approx. 36% of the TGVDs survive. The smaller variation in the PFET TCDs reduces the signal-to-noise for the VDC even further. However, the 832,343 TCD-based bits for this chip that survive are reproducible across the TV corners and exhibit excellent statistical characteristics as we show below.

3.6 Fixed Length Bitstrings and TMR

In actual applications, only a fixed number of bits are needed. With encryption, the values vary between 128 to 1024 bits, depending on the encryption algorithm. The large number of bits available from the PUF can be beneficial, however, by allowing a distinct set of fixed-length secret keys to be generated over time during successive enrollments.

A second possible usage scenario leverages this large pool of strong bits to further increase the resiliency to bit flip failures, i.e., beyond that provided by thresholding. We propose a bitstring replication method that mimics a popular scheme used in fault tolerance called triple-module-redundancy or TMR. In this technique, a fixed length, e.g., 1,024-bit, bitstring is generated as described above. TMR is then applied to generate two more copies of the bitstring. The two copies are generated by parsing the strong bit sequence until a match is found to each bit in the first bitstring. During regeneration, a majority voting scheme is applied to each of the columns in the three identically regenerated bitstrings as a means of avoiding single bit flip failures. In other words, the final bitstring is constructed by using the majority of the 3 column bits as the final bit for each bit position, i.e., a '1' is assigned in the final bitstring when 2 or more of the 3 bits in the column are '1', and a '0' otherwise. An illustrative example is given in Section S2.

A PUF that is able to generate strong bit sequences that are locally random (a quality measured by the NIST tests [1] presented in the Section 4) ensures that a match occurs for each bit during the generation of the two copies every 2 bits on average. Under these conditions, it follows that a TMR-based bitstring, and its public data, consumes on average 5 times more strong bits than a non-TMR-based bitstring. The benefit, on the other hand, is a significant decrease in the 'probability of failure', i.e., the likelihood of a bit flip occurring during regeneration, as we show in Section 4. Moreover, this scheme offers flexibility by allowing a trade-off between tolerance to bit flips and public data size.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the several important statistical properties of the TGVD and TCD-derived bitstrings including randomness, uniqueness and probability of bit flips, e.g., failures to regenerate the bitstring under different environmental conditions. As discussed in Section 3.2, the process of digitizing the voltages using the VDC adds noise and reduces the number of corresponding strong bits. The penalty of the digitization process is evaluated by carrying out the same analysis using the TGVDs directly, and serves to illustrate the best that can be achieved in the absence of digitization noise.

Fig. 7a) gives the inter-chip hamming distance (HD) distribution using the TGVDs while Fig. 7b) shows the distribution using TCDs. The graphs plot HD along the x-axis against the number of instances on the y-axis[1]. With 63 chips, the total number of instances is 63*62/2 = 1,953. The distributions are 'fitted' with Gaussian curves to illustrate the level of conformity they exhibit to this distribution.

Since HDs must be computed across bitstrings of equal length, it was necessary to truncate the bitstrings used in Fig. 7 to the length obtained for the chip with the fewest number of strong bits. Truncation reduced the lengths to 1,901,845 for the TGVD analysis and 725,230 for the TCD analysis, which are approx. 33.6% and 12.8%, resp., of the maximum possible length, i.e., 5,662,020 bits. The chip with the longest bitstring, in comparison, uses 35.6% of the maximum for the TGVD analysis and 15.0% for the TCD analysis. The term **truncated bitstrings** is used to refer to the shorter, equal-length bitstrings.

1. HD is computed by counting the number of bits that are different in the bitstrings from two chips.

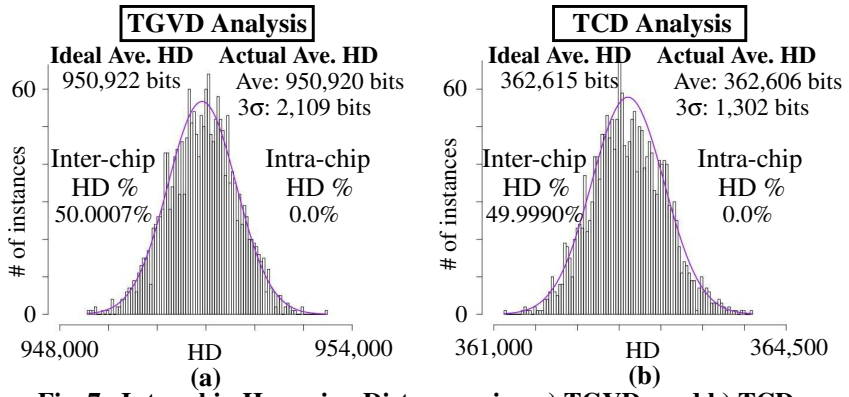

Fig. 7. Inter-chip Hamming Distance using a) TGVDs and b) TCDs.

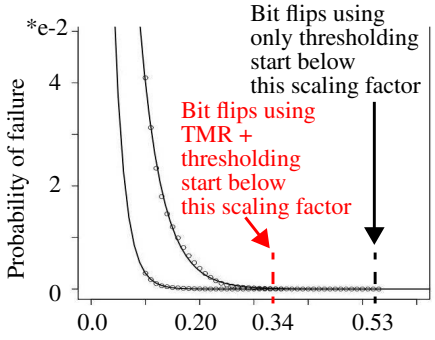

Fig. 8. NFET TCD scaling factor (x-axis) vs. probability of failure (y-axis).

The actual average inter-chip HDs listed in Fig. 7 are nearly equal to the ideal value of 50%. In contrast, the average inter-chip HDs for the bitstrings of length 5,662,020, i.e., those with the weak bits included (not shown), is 48.4% and 48.5% for TGVD and TCD, resp., so removing the weak bits improves the inter-chip HDs. The 3σ values shown in the figure are derived from the Gaussian curves and represent the spread of the distributions (where smaller is better). These values are small relative to the length of the truncated bitstrings, e.g., they are only 0.11% and 0.18% of the lengths for the TGVD and TCD analysis, resp.

The scaling factors are set to 0.42 (NFET) and 0.39 (PFET) for the TGVD analysis and 0.53 (NFET) and 0.78 (PFET) for the TCD analysis. These values were derived by analyzing the bitstrings across all 9 TV corners and tuning the values until no bit flips occurred (Section S2 discusses how this can be done in practice). Therefore, the intra-chip HD is 0.0% as shown in Fig 7 for both analyses. However, the underlying noise levels can be measured by disabling the thresholding technique, yielding intra-chip HDs of 5.11% and 8.68% for the TGVD and TCD analyses, resp. The increase in the TDC intra-chip HD over that given for TGVD reflects the noise added by the VDC digitization process.

We applied the NIST statistical tests [1] to the truncated bitstrings of the 63 chips at a significance level of 0.01 (the default). The TGVD and TCD bitstrings **pass all tests**, with no fewer than 60 passing chips per test (the number required by NIST for the test to be considered 'passed'). Moreover, all tests passed the **Pvalue-of-the-Pvalues** metric.

Fixed-length bitstrings were also created using the TMR-based scheme proposed in Section 3.6. In our experiments, we were able to create, on average, 381 1024-bit TMR-based bitstrings per chip using TGVD data, and 156 on average using TCD data. Although not shown, the statistical test results are similar to those discussed above for the longer bitstrings.

As discussed in Section 3.6, the TMR scheme improves resiliency to bit flips over the thresholding scheme alone. The curves shown in Fig. 8 illustrate the improvement. The scaling factor used for NFETs (the PFET scaling factor is also changed proportionally) is plotted along the x-axis against the probability of failure on the y-axis. The probability of failure is computed at each scaling factor value by dividing the number of bit flips that occur in all 63 chips by the total number of strong bits produced. The curve on the left is the result obtained using the TMR + thresholding technique, while the curve on the right uses only thresholding. Both curves are exponential in shape (see Section S2 for curve fits and further analysis). However, from the positions of the curves, it is clear that the TMR scheme requires a lower scaling factor, 0.34 vs. 0.53, before any bit flips occur. Using 0.53 as the scaling factor, the probability of failure is 1.1e-6 with thresholding but improves significantly to 1.5e-12 after adding TMR.

5. CONCLUSIONS

A transmission gate (TG) PUF and on-chip voltage-to-digital conversion circuit are evaluated on 63 copies of a 90 nm chip, at 9 temperature-voltage corners. Thresholding and triple-module-redundancy techniques are proposed as a means of avoiding bit flips. Results from statistical tests confirm that cryptographic quality bitstrings are obtained using either the TG voltages or their digitized representations. The proposed bit flip avoidance schemes allow the user to trade-off the probability of failure with helper data overhead for applications requiring bitstring regeneration.

6. REFERENCES

[1] NIST: Computer Security Division, Statistical Tests, http://cs-rc.nist.gov/groups/ST/toolkit/rng/stats_tests.html

[2] K. Lofstrom, *et al.*, "IC Identification Circuits using Device Mismatch," *SSCC*, 2000, pp. 372-373.

[3] R. S. Pappu, *et al.*, "Physical One-Way Functions," *Science*, 297(6), 2002, pp. 2026-2030.

[4] B. Gassend, *et al.*, "Controlled Physical Random Functions," *Conference on Computer Security Applications*, 2002.

[5] M. Majzoobi, *et al.*, "Lightweight Secure PUFs", *ICCAD*, 2008.

[6] G. Qu and C. Yin, "Temperature-Aware Cooperative Ring Oscillator PUF", *HOST,* 2009, pp. 36-42.

[7] A. Maiti and P.Schaumont, "Improving the Quality of a Physical Unclonable Function using Configurable Ring Oscillators", *FPLA*, 2009. pp. 703-707.

[8] J. Guajardo, *et al.*, "Physical Unclonable Functions and Public Key Crypto for FPGA IP Protection," *FPLA*, 2007, 189-195.

[9] R. Helinski, *et al.*, "Physical Unclonable Function Defined Using Power Distribution System Equivalent Resistance Variations", *DAC*, 2009, pp. 676-681.

[10] J. Ju, R. Chakraborty, R. Rad, J. Plusquellic, "Bit String Analysis of Physical Unclonable Functions based on Resistance Variations in Metals and Transistors", *HOST*, 2012, pp. 13-20.

[11] K. Rosenfeld, *et al.*, "Sensor Physical Unclonable Functions", *HOST*, 2010, pp. 112-117.

[12] L. Guansheng, Y.M. Tousi, A. Hassibi and E. Afshari, "Delay-Line-Based Analog-to-Digital Converters,", Trans. on *CAS II*, Volume: 56, Issue: 6, 2009, pp. 464-468.

[13] Dan O'Sullivan & Tom Igoe, "Physical Computing: Sensing and Controlling the Physical World with Computers," Thomson Course Technology Publishers, 2004, pp 388-391.

[14] R. Fried and C. C. Enz, "Simple and Accurate Voltage Adder/Subtractor," *Electronics Letters*, vol. 33, 1997, pp. 944-945.

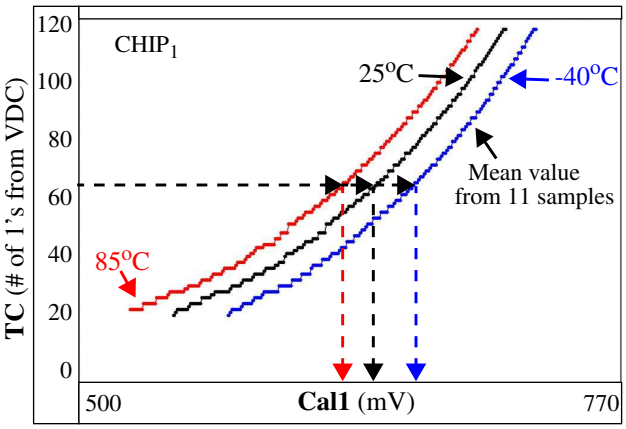

Fig. 9. VDC calibration curves at 85, 25, -40°C and 1.2V illustrating offset calculation process.

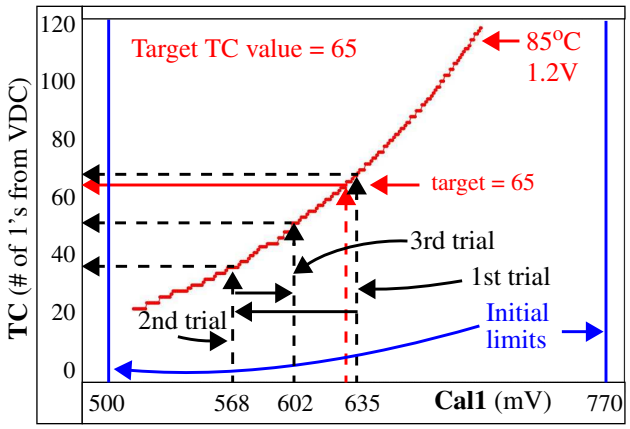

Fig. 10. Illustration of the binary search process used during calibration at 85°C, 1.2V.

SUPPLEMENTARY MATERAL

S1 VDC Calibration Process

The calibration process described in Section 3.2 is further illustrated using the Cal1 vs. TC curves shown in Fig. 9. As indicated earlier, calibration is carried out before enrollment and regeneration, and its objective is to find an appropriate Cal1 voltage offset that prevents overflow in the VDC for any of the TGVs that will be measured during bit generation. We determined that testing a subset of 9 TGs during calibration is sufficient to obtain a good predictor for offset voltage that prevents overflow.

The goal of calibration is to select an offset voltage such that the TG-under-test produces the same TC value independent of the TV corner. This objective is illustrated in Fig. 9 with the horizontal dashed line at TC = 65. The 3 curves shown represent the mean values produced by the VDC on $CHIP_1$ as the Cal1 voltage is swept across a range of values (similar to the process described in Section 3.2 in reference to Fig. 4) at 3 different temperatures. The different positions of the dashed vertical lines from each curve make it clear that the offset voltage needs to change in order to maintain a value of 65 in the VDC. Note that the TGV itself measured from the TG-under-test will also change as a function of temperature. This situation is handled by using the TGVs directly in the calibration process (as opposed to using a special voltage source).

Calibration is carried out by enabling each of a select, distributed group of TGs, one at a time, and performing a binary search. The search process varies the Cal1 voltage offset until the TG-under-test produces a specific TC value. The process is illustrated in Fig. 10 using the 85°C Cal1-TC curve from Fig. 9. The initial limits are set to 500 mV and 770 mV. The 1st trial selects the midpoint between these limits, i.e., 635 mV. Note this midpoint voltage is the sum of the TGV and the offset voltage that is being tuned in the search. The 1st trial produces a TC of approx. 68, which is larger than the target. Therefore, the next trial uses 635 mV as the upper limit and the new midpoint voltage becomes 568. The 2nd trial produces a TC of 35, so 568 is used as the lower limit for the new midpoint. The process continues until an offset is found that produces a TC of 65. The binary search process is repeated using 9 TGs as a means of obtaining a value that best approximates the average behavior. The median value from the 9 calibration tests is used as the final offset, which is added to all subsequent TGVs measured at this TV corner.

S2 Thresholding & TMR-based scheme

The thresholding and TMR-based schemes are described in Sections 3.5 and 3.6. This section of the Supplementary Material is designed to clarify this process with an example. The thresholding scheme shares characteristics with the shielding function proposed in [15] but is simpler because it is based entirely on strong bits, referred to as 'robust' bits in the reference. This fact changes the nature of the public data and eliminates information leakage that, although unlikely, is possible with shielding functions.

Fig. 11 illustrates the proposed thresholding and TMR-based scheme using data from a hypothetical chip. The x-axis plots a sequence of comparisons that would be used to generate a bitstring, while the y-axis plots the differences between the pairings of TCDs. Each difference reflects the relative ordering of the two TCDs, e.g., positive difference values indicate that the first TCD is larger than the second. For strong bits, the TCD difference data points must lie above or below the thresholds, labeled '+Tr' and '-Tr' in the figure. This condition, when met, is recorded using a '1' in the thresholding bitstring shown below the data points. Weak bits, on the other hand fall within the thresholds and are indicated with a '0'. The bold (and blue) '0's indicate strong bits that are skipped under the TMR scheme described below.

As discussed in Section 3.6, the TMR-based method constructs 3 identical bitstrings during enrollment as shown along the bottom of Fig. 11. The left-most bitstring labeled 'Secret BS' is generated from the first 4 strong bits encountered as the sequence of data points is parsed from left to right. The second bitstring labeled 'Redundant BS_1' is produced from the next sequence of data points but has the additional constraint that each of its bits must match those in the first bitstring. During its construction, it may happen in the continued left-to-right parsing of the data points that a strong bit is encountered that does not match the corresponding position in the 'Secret BS'. In the example, this occurs at the position indicated by the left-most bold '0' in the thresholding bitstring. Here, we encountered a strong bit with a value of '0'. But the 'Secret BS' requires the first bit to be a '1', so this strong bit is skipped. This process continues until redundant bitstrings BS_1 and BS_2 bitstrings are constructed.

The number of strong bits required to generate a secret bitstring of length 4 is approx 5x or 20. From the example, this is evaluated by counting the number of '1's and bolded '0's in the thresholding bitstring, which is given as 19. The benefit of creating these redun-

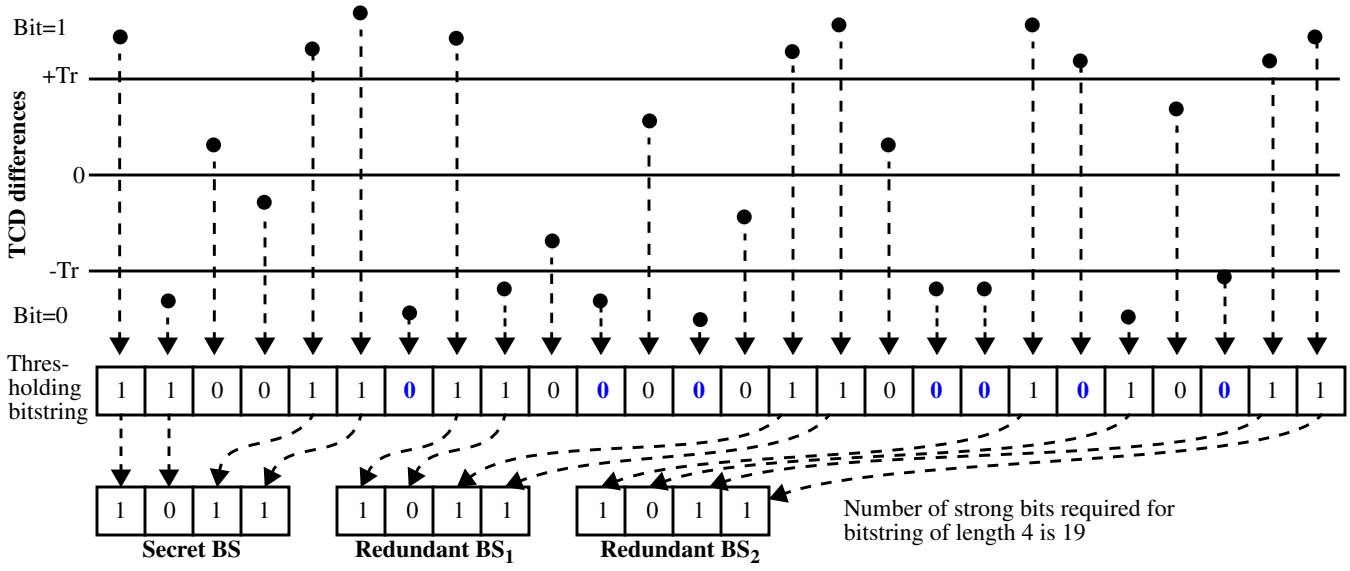

Fig. 11. Secret bitstring generation example using the proposed thresholding and TMR-based method.

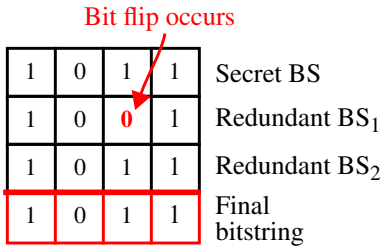

Fig. 12. Bit flip avoidance illustration using example from Fig. 11.

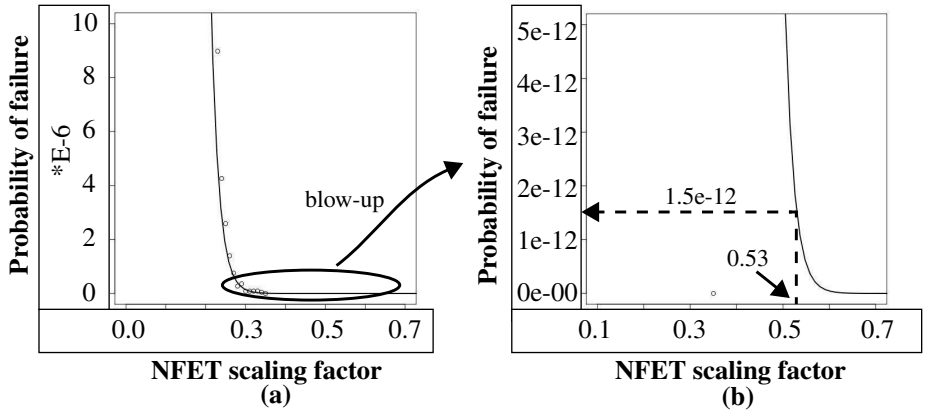

Fig. 13. a) TMR probability of error curve and b) blow-up of the designated region. The discrete curve is fitted with a superimposed exponential function.

dant bitstrings is the improved tolerance that they provide to bit flips. For example, during regeneration, the three bitstrings are again produced, but this time using the thresholding bitstring to determine which TCDs to compare.

In scenarios where the threshold is set too low, it is possible that a strong data point used in enrollment is displaced across both the threshold and the '0' line because of different TV conditions in regeneration, causing a bit flip. However, with TMR, a bit flip can be avoided if no more than 1 bit flip occurs in a single column of the matrix of bits created from the 3 bitstrings. For example, the first 3 rows of the matrix of bits in Fig. 12 is constructed during regeneration in a similar way to those shown in Fig. 11 for enrollment. The bottom row represents the final secret bitstring and is constructed by using a **majority vote** scheme (in the spirit of TMR). The bit flip shown in the third column has no effect on the final bitstring because the other two bits in that column are '1', and under the rule of majority voting, the final secret bit is therefore defined as '1'[1].

In Section 4, the probability of failure using thresholding alone and in combination with TMR was discussed, with the latter improving significantly on the former, from 1.1e-6 to 1.5e-12. These values were obtained by fitting the discrete-valued curves produced from repeatedly running the analysis at different scaling factors with exponential functions. Fig. 13(a) shows the data for the TMR + thresholding curve in Fig. 8 with the fitted exponential curve. The exponential is clearly a good fit to the data points. Fig. 13(b) shows a blow-up of the region around the NFET scaling factor of 0.53 from which the estimate of 1.5e-12 was derived.

S3 Run-Length Encoding of Public Data

The size of the public (helper) data under the thresholding and TMR-based schemes can be reduced using compression techniques such as run-length encoding. The benefit of run-length encoding is its simplicity. Fig. 14 shows an example of a thresholding bitstring with 26 bits. The long strings of '0's can be run-length encoded by simply counting them and replacing the '0' sequence with a field which represents the number of '0's in each sequence. In the example, the run-length encoded bitstring uses 19 bits instead of 26. The longer the sequences of '0's, the more efficient the scheme becomes. The best choice for the field width depends on the nature of the public data, i.e., the average length of the '0' strings.

1. TRM can be extended to include 5, 7, etc. copies of the bitstring to further enhance bit flip resiliency.

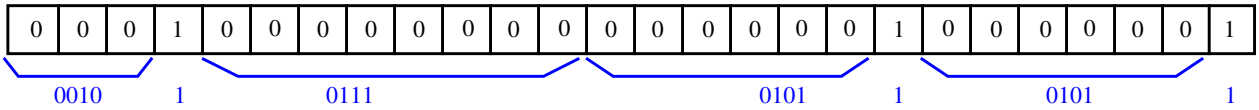

Fig. 14. Examples of run-length encoding as a compression technique to reduce public data size. Original public data string has 26 bits. Run-length encoded using a field width of 4 yields 19 bits.

The public data for the TCD analysis from Section 4 indicates that approx. 14% of the bits survive the thresholding, and even fewer, approx. 8.4%, are marked with '1's in the public data when TMR is added. The public data is therefore expected to contain strings of 0's with average lengths of approx. 11 under thresholding + TMR. Therefore, a field width between 3 and 4 (which allows counting up to 8 and 16, resp.) should be optimal. We found that a field width of 5 is best and yields a 42% reduction on average to the size of the original public data string. We plan to explore other compression techniques in future work.

In addition to compression, obfuscation is required for the thresholding bitstring when the PUF usage scenario involves authentication. This is true because the 'secret' bitstring is not kept on chip as it is for encryption but rather is also made public. With both bitstrings available, an adversary can reverse engineer the relative ordering of the TCDs. In order to prevent this, we propose to obfuscate a portion of the thresholding bitstring as follows. During enrollment, the first *n* strong bits, e.g., 128, are used as a key to encrypt the thresholding bitstring, excluding those public data bits that correspond to the encryption key itself. These bits do not need to be encrypted because the key is never made public.

S4 Underlying Stability Characteristics of the TG PUF

As discussed in Section 3.5, bit flips introduced by TV variations represent the primary threat to the TG-PUF's reliability. Here, we investigate the underlying mechanism that cause TGVs to vary as a function of temperature and voltage. Although the main focus of this paper has been on the TCDs, instability in the TGVs is the primary component of the instability observed in the TCDs and is therefore the focus of our analysis.

Fig. 15 shows a portion of the NFET stack shown on the right side of Fig. 2 for two arbitrary SMCs, SMC_X and SMC_Y. As discussed earlier, TGVDs are created to eliminate the sense wire bias in which the TGVs from two distinct tests in the same SMC are subtracted. The figure includes only 2 NFETs (of the 8) from the stack as an illustration of this operation. In each of the two tests, two transistors are enabled, e.g., TG_{1a} and TG_{1b}, which establishes a current path labeled I_{1ab} from the off-chip power supply, V_{DD}, through the sense wire resistor (R_{w1}) and the two TGs to the on-chip GND grid. The voltage between the two transistors labeled TGV_{1ab} is measured off-chip using a voltmeter (VM) by enabling a third transistor TG_{1S}. A second voltage drop, TGV_{1ac} (not shown), is obtained in similar fashion by enabling TG_{1a} and TG_{1c} within SMC_X. The TGVD is defined as $TGVD_1 = TGV_{1ab} - TGV_{1ac}$. The exact same process is carried out within SMC_Y to obtain $TGVD_2$.

In order to understand how the TGVD change as a function of TV variations, we need to first determine the modes of operation of the two transistors in the shorting path. The shorting path defines a voltage divider network with, e.g., $R_{w1} + R_{1a}$ (the sense wire resistance and **TG_{1a} R_{on} resistance**) as one element and R_{1b} (the **TG_{1b} R_{on} resistance**) as the second element. The R_w's vary from approx. 100 Ohms (upper left-most SMC in Fig. 1) to 1.5 KOhms (lower right-most SMC). Unfortunately, there is no way to measure the R_w's by themselves (the values above are obtained from the layout geometries and the design manual's resistance/square values) so they are lumped together with the transistor resistances R_{1a} and R_{2a} for the purposes of this analysis.

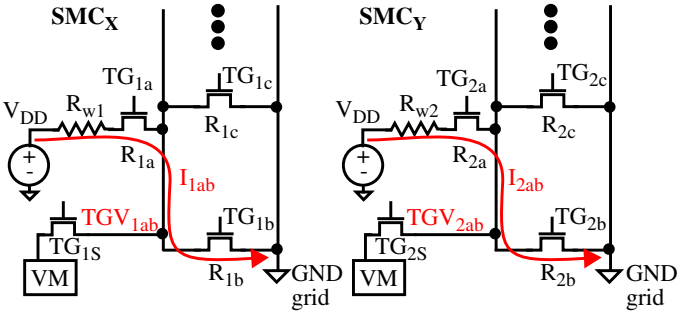

Fig. 15. Example and analysis of TV variations and its impact on TGVDs.

$$TGVD_1 = V_{DD}\left[\frac{1}{1+\frac{R_{1a}(1+Y_{1a})}{R_{1b}(1+X_{1b})}} - \frac{1}{1+\frac{R_{1a}'(1+Y_{1a}')}{R_{1c}(1+X_{1c})}}\right] \quad \textbf{Eq. 1.}$$

$$TGVD_2 = V_{DD}\left[\frac{1}{1+\frac{R_{2a}(1+Y_{2a})}{R_{2b}(1+X_{2b})}} - \frac{1}{1+\frac{R_{2a}'(1+Y_{2a}')}{R_{2c}(1+X_{2c})}}\right] \quad \textbf{Eq. 2.}$$

Eqs. 1 and 2 are the defining equations for the 2 TGVDs. Each equation incorporates two voltage divider network equations, one for each of the 2 TGVs. The R_{on}'s are obtained by dividing, e.g., (V_{DD}-TGV_{1a}) and TGV_{1a} by the current I_{1ab} measured using the off-chip power supply. The two voltage divider subexpressions will be referred to as the **1st term** and the **2nd term** subsequently within which the R_{on} ratios at 25°C, 1.2V (the enrollment corner) are referenced. Note that R_{1a} and R_{2a} from the 1st terms are designated as R_{1a}' and R_{2a}' in the 2nd terms because these resistances are a function of the drain-to-source voltage (V_{DS}), which are different in the two ratios as discussed below. The X and Y terms are defined as the percentage changes in the R_{on} of the associated transistors at a specific TV corner with respect to the R_{on} measured at enrollment.

The magnitude of the R_{on}'s are determined primarily by the mode of operation of the two transistors. NFET transistors whose sources are connected to the on-chip GND grid, e.g., TG_{1b}, TG_{1c}, TG_{2b} and TG_{2c} operate in the linear region. This is true because the V_{DS} for these transistors are in the range of 200 mV while V_{GS} is equal to V_{DD}, e.g., 1.2 V. The design manual specifies that threshold voltages are > 300 mV in this 90 nm technology. There-

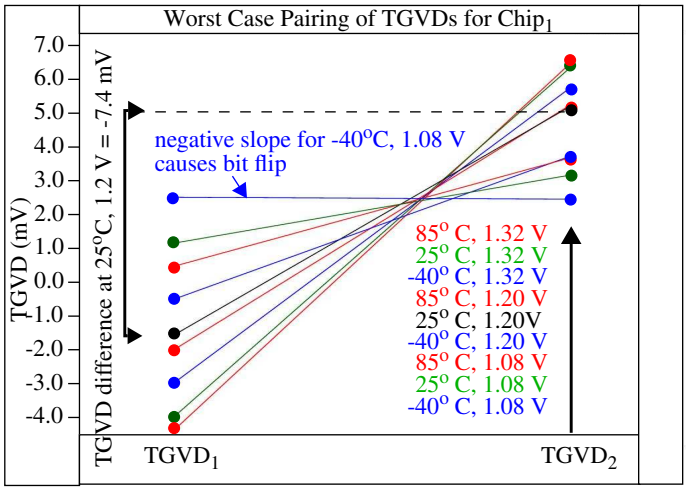

Fig. 16. 'Worst-case' TGVD comparison for Chip$_1$, where worst case is defined to be the comparison that has the largest enrollment TGVD difference and a bit flip.

fore, $V_{DS} < V_{GS} - V_t$ indicating the operating mode is linear.

NFET transistors TG_{1a} and TG_{2a} on the other hand operate in saturation. This is true because the voltage drops across the R_w's are less than 300 mV (typical currents for I_{1ab} and I_{2ab} are approx. 180 uAs). Therefore, with V_{DD} at 1.2 V, V_{GS} is approx. (1.2 - 0.2) = 1.0 V while V_{DS} is, in the worse case, (0.9 - 0.2) = 0.7. Moreover, threshold voltages increase when the V_{SB} (source-to-substrate) is greater than 0, a condition that holds true for these NFET transistors. Therefore, $V_{DS} > V_{GS} - V_t$ indicating the operating mode is saturation.

The resistances given in Eqs. 1 and 2 will change as a function of TV conditions. If the percentage change in all R_{on}'s are identical, i.e., all X and Y are the same, then TV variations would not increase the number of bit flips that occur over the number introduced by measurement noise alone. This is not the case, however. Therefore, the R_{on}'s and the corresponding X and Y percentage change values from the equations must vary at different rates across the TV corners.

This characteristic of the NFET resistances is demonstrated using data from a special, worst-case, pairing of TGVDs. In particular, we analyze the pairing (from the 5,662,020 pairings described in Section 4) from Chip$_1$ that possesses the largest difference in the TGVDs at enrollment AND has a bit flip. This pairing defines the minimum threshold (see Section 3.5) that can be used to avoid bit flips across the 9 TV corners.

Fig. 16 shows the behavior of the two TGVDs used in this pairing. The 9 data points for each TGVD, one for each TV corner, are plotted as a vertical sequence under each TGVD labeled on the x-axis. Each of the points from $TGVD_1$ is line-connected with the point in $TGVD_2$ corresponding to the same TV experiment. If the sign of the difference $TGVD_1 - TGVD_2$ remains the same, then the set of lines would all have positive or all have negative slopes. Instead, they cross over and depict a near complete reversal in order. For example, the ordering from top-to-bottom of the points for $TGVD_1$ is opposite to the legend's ordering, which lists the TV corners in descending order according to voltage and then temperature, while the points for $TGVD_2$ are consistent with it. Note that

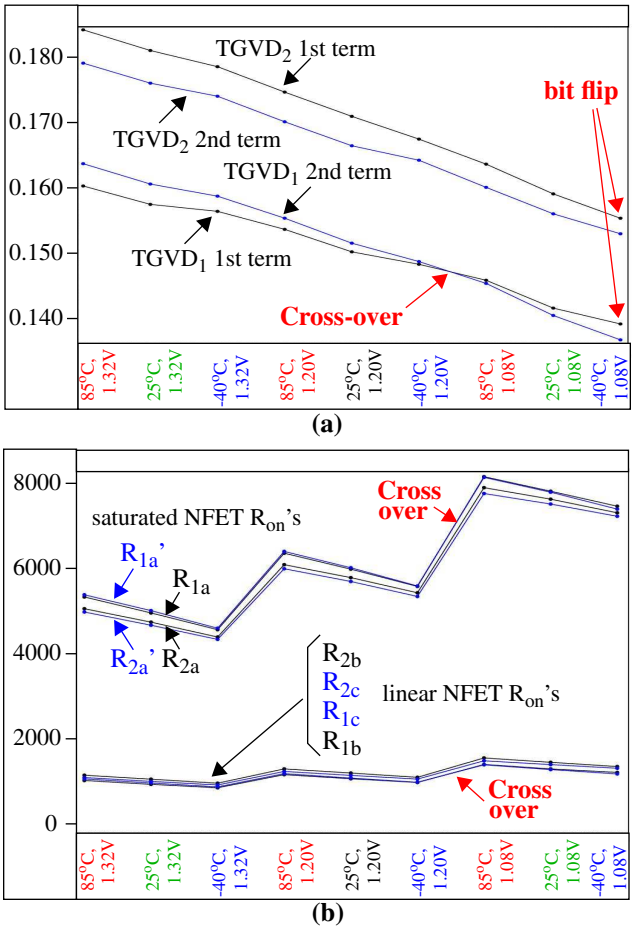

Fig. 17. Behavior of the a) 1st and 2nd terms and b) the individual R_{on}'s from Eqs. 1 and 2 across the TV corners for $TGVD_1$ and $TGVD_2$ given in Fig. 16.

the slope of the line associated with the -40°C, 1.08 V is negative while the others are positive. This condition reflects a bit flip, i.e., $TGVD_1 > TGVD_2$ at this TV corner while $TGVD_1 < TGVD_2$ in the others.

The behavior of the 1st and 2nd terms in Eqs. 1 and 2 as a function of TV corners are shown in Fig. 17(a), which plots the two terms for each TGVD as separate curves. Each curve consists of 9 points (one for each TV corner). Interestingly, all 4 terms decrease monotonically as TV decrease, which illustrate the self-compensation property of the NFET pair. Unfortunately, the rate at which the terms decrease, which is reflected in slope of the curves, is not constant. The larger difference in the slopes between the 1st and 2nd terms for $TGVD_1$ cause the curves to cross over and eventually introduce a bit flip at the -40°C, 1.08V corner. The curves in Fig. 17(b) plot the behavior of the individual R_{on}'s within the ratios of Eqs. 1 and 2. Although the R_{on}'s vary significantly with TV, especially for the saturated NFET R_{on}'s shown along the top of the figure, the corresponding changes in the R_{on}'s of the linear NFETs compensate for most, but not all, of the variations. In particular, the R_{on}'s for all 1.08V TV corners cross over for $TGVD_1$.

S5 Supplementary Material References

[15] B. Skoric, P. Tuyls, W. Ophey, "Robust Key Extraction from Physical Uncloneable Functions", Chapter in Applied Cryptography and Network Security, 2005.