A High Reliability PUF Using Hot Carrier Injection Based Response Reinforcement

Mudit Bhargava, Ken Mai

Department of Electrical and Computer Engineering,
Carnegie Mellon University
{mbhargav,kenmai}@ece.cmu.edu

Abstract. Achieving high reliability across environmental variations and over aging in physical unclonable functions (PUFs) remains a challenge for PUF designers. The conventional method to improve PUF reliability is to use powerful error correction codes (ECC) to correct the errors in the raw response from the PUF core. Unfortunately, these ECC blocks generally have high VLSI overheads, which scale up quickly with the error correction capability. Alternately, researchers have proposed techniques to increase the reliability of the PUF core, and thus significantly reduce the required strength (and complexity) of the ECC. One method of increasing the reliability of the PUF core is to use normally detrimental IC aging effects to reinforce the desired (or "golden") response of the PUF by altering the PUF circuit characteristics permanently and hence making the PUF more reliable. In this work, we present a PUF response reinforcement technique based on hot carrier injection (HCI) which can reinforce the PUF golden response in short stress times (i.e., tens of seconds), without impacting the surrounding circuits, and that has high permanence (i.e., does not degrade significantly over aging). We present a self-contained HCI-reinforcement-enabled PUF circuit based on sense amplifiers (SA) which autonomously self-reinforces with minimal external intervention. We have fabricated a custom ASIC testchip in 65nm bulk CMOS with the proposed PUF design. Measured results show high reliability across environmental variations and accelerated aging, as well as good uniqueness and randomness. For example, 1600 SA elements, after being HCI stressed for 125s, show 100% reliability (zero errors) across ±20% voltage variations a temperature range of -20°C to 85°C.

1 Introduction

Over the past decade, Silicon physical unclonable functions (PUFs) have emerged as highly useful blocks in the design of secure hardware in applications such as identification/authentication and even encryption key generation [1–6]. PUFs derive their randomness from uncontrolled random variation phenomena that occur during the Silicon chip manufacturing process. Rather than store a set of random bits, PUFs generate these random bits every time they are activated. Most PUF implementations do so by amplifying some electrical characteristic (e.g., delay, threshold voltage) from two nominally identical circuit components in the PUF core. For example, delay based PUFs (arbiter [1] or ring oscillator [2]) generate their random bits by amplifying the difference in delay of two nominally identical delay paths. Bi-stable element based PUFs (SRAM [5] or sense amplifier [6]) generate their random bits by amplifying the differences in strength of two (or more) transistors using a positive feedback structure (usually a cross-coupled inverter pair).

These electrical differences, especially when small, often flip polarity across different environmental variations (voltage and temperature), in the presence of ambient noise, or over aging, resulting in some bits of the raw PUF response being unreliable. Previous hardware studies have shown that for some designs $> 25\%$ of the PUF response bits may be unreliable across environmental variations [7,8]. Since differences of larger magnitude require larger variations to flip polarity, larger electrical differences generally result in more reliable PUFs. Although some applications like identification and authentication can be designed to tolerate a few errors in the response without significant loss of security, all applications can benefit from more reliable PUFs, and applications such as key generation require the PUF response to be perfectly reliable. The conventional method to improve PUF reliability use powerful error correction codes (ECC) to correct the raw response from the PUF core. Unfortunately, these ECC blocks generally have significant VLSI overheads, which scale up quickly as the number of bits of correction increases [9–14]. Most ECC implementations require >2000 cycles and an area of >3000 SRAM bits to generate 128 reliable bits (i.e., >23x SRAM bits per reliable bit). Further, they require the generation and storage (typically off-chip in non-volatile memories) of helper data (typically >3000 bits to generate 128 reliable bits), which is later used for correction. The helper data has been shown to be a source of information leakage requiring careful design [9,10]. However, these overheads reduce significantly if the errors in the raw response bits are reduced. For example, the BCH coding in [15] requires 26.7 raw response bits to generate a reliable bit if the raw response bits exhibit 15% errors but requires only 3.68 raw response bits if the errors reduce to 6%.

Hence, researchers have proposed techniques to increase the reliability of the PUF core, and thus significantly reduce the required strength (and complexity) of the ECC [16,17]. One method of increasing the reliability of the PUF core is to use normally detrimental IC aging effects to reinforce the desired (or "golden") response of the PUF by permanently altering the PUF circuit characteristics such that the difference in the electrical characteristic is increased in magnitude, and hence making the PUF more reliable. Previous related work used the IC aging phenomena of negative bias temperature instability (NBTI) to improve reliability of an SRAM PUF by ∼40% [16]. Despite its efficacy in increasing the PUF reliability, NBTI-based response reinforcement requires long baking times (e.g., > 20 hours) that are incompatible with an industrial high-volume-production manufacture and test flow. Further, the high temperatures needed for NBTI-based reinforcement cannot be applied selectively, and thus would detrimentally age all circuits on the chip. Finally, transistor V_{TH} shifts due to NBTI are not permanent, and the transistors return to near their initial characteristics over time. NBTI can typically achieve a permanent V$_{TH}$ shift (after recovery) of only ~10-40mV in PMOS devices over years of stress (equivalent to over tens of hours of accelerated aging under elevated voltage and temperature) [18, 19].

In this paper, we propose to use a different IC aging effect, hot carrier injection (HCI), for PUF response reinforcement which overcomes the shortcomings of the NBTI-based reinforcement enabling extremely high PUF reliability across environmental variations and aging, while requiring only a very short, one-time, reinforcement stress. Previously, researchers have demonstrated the use of controlled HCI stress for gain in performance and reliability in SRAMs [20]. They achieved an increase of > 100mV in transistor V$_{TH}$ in a short stress duration of 10 seconds without increasing temperature or the global supply voltage.

The novel contributions of this paper are as follows:

1. We present a HCI based response reinforcement technique which can reinforce the PUF golden response in short stress times (i.e., tens of seconds) which is over two orders of magnitude shorter than the time needed by NBTI stress (> 20 hours) while simultaneously achieving significantly better reliability improvement. Note that the HCI response reinforcement is a one-time, post-manufacture step and no HCI stress is applied in the field. Once stressed, the effect has high permanence (i.e., does not lessen significantly over time). Further, our technique does not impact the surrounding circuits as it does not require high-temperature or a high-voltage to the global supply. A targeted high voltage supply, decoupled from the global supply voltage, of ~2.5V (nominal ~1.0V) is required for stress and is connected to only a few select devices. This voltage is usually available as the I/O supply voltage for the pads and hence no dedicated extra supply is needed.
2. We present a self-contained HCI-reinforcement-enabled PUF circuit implementation which autonomously self-reinforces with minimal external intervention. No data related to the PUF core response bits are ever sent off chip, minimizing the information leakage. The PUF core circuit is a bi-stable PUF based on StrongARM sense amplifiers [6].
3. The design was implemented on a 65nm bulk CMOS testchip, and we present measured results for the PUF reliability, randomness, and uniqueness from multiple testchip samples. These measured results show that using HCI-based response reinforcement, we can efficiently achieve extremely high PUF response reliability across environmental variations (voltage and temperature) as well as over IC aging. We measure the response across combinations of voltage (0.8V, 1.0V, 1.2V) and temperature variations (-20°C, 27°C, 85°C) for each of the 1600 sense amplifiers per chip which are first evaluated 100 times per voltage/temperature corner and then evaluated 125,000 times at the worst case corner. For the measured testchip samples, after undergoing a one-time HCI stress of 125 seconds, we found zero errors (100% reliability) for all the 1600 sense amplifier outputs across all voltage/temperature corners. Further, one testchip sample was then baked at high temperature and voltage to simulate ~1.7 years of IC aging and did not show any noticeable change in reliability, indicating high permanence for the HCI stress effects.

We present the concept of HCI in Section 2. In Section 3, we present the HCI-SA PUF structure built on the baseline StrongARM sense amplifier PUF [6,16]. In Section 4 and Section 5 we present the details of our 65nm testchip and the measured results from it respectively.

2 Hot Carrier Injection (HCI)

Hot carrier injection (HCI) is a phenomenon by which the threshold voltage (V$_{TH}$) of a transistor may be permanently altered post-manufacturing when high energy carriers become trapped in the gate oxide. The increase in V$_{TH}$ due to HCI stress is usually an undesired phenomenon as it makes the transistors slower and can lower the performance of the circuit, but we leverage HCI to increase the reliability in PUF circuits.

Figure 1 gives an overview of the HCI phenomenon for an NMOS transistor. Figure 1(a) shows an NMOS transistor under normal biasing. The gate-to-source voltage (V_{GS}) and the drain-to-source voltage (V_{DS}) are at nominal supply voltage (assumed to be 1V) and the transistor operates in saturation. As V_{DS} increases, as shown in Figure 1(b), velocity saturation occurs and for today's short channels, it can occur for much of the channel. Electrons moving at saturation velocity continue to acquire kinetic energy, but their velocity is randomized by excessive collisions such that their average velocity along the field direction no longer increases but their random kinetic energy does. These high energy electrons are called hot carriers and their population increases for higher V_{DS}. A small fraction of these hot carriers acquire enough energy to overcome the silicon-oxide barrier energy and get injected into the gate oxide (the brown square in Figure 1(b)).

Transistors with carriers trapped in the oxide require a higher V_{GS} for inversion, effectively increasing their V$_{TH}$. When this stressed NMOS transistor, with trapped electrons, is used under normal V_{DS} biasing as shown in Figure 1(c), the NMOS transistor behaves asymmetrically under the two source-drain biasing directions. When the current flows in the same direction under normal biasing, as in the stressed biasing, such that the trapped electrons are near the drain, the NMOS transistor sees only a slight increase in V$_{TH}$. However, when used with the source-drain directionality reversed, such that the trapped electrons are near the source, the NMOS transistor will see a much higher increase in V$_{TH}$. This is because for inversion, most of the charge accumulates in the channel near the source and with trapped carriers near the source, it requires a larger V_{GS} to attract electrons for inversion. Since the electrons are trapped deep into the oxide, most of the increase in V$_{TH}$ is permanent, making HCI an attractive mechanism to reinforce the PUF response.

3 HCI-Enabled Sense Amplifier (HCI-SA) PUF

In this section we describe a PUF circuit that uses HCI-based response reinforcement to increase reliability over environmental variations and aging. The PUF

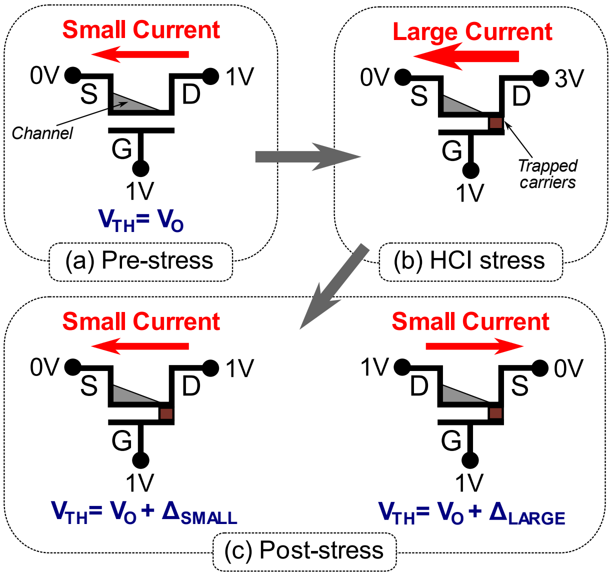


Fig. 1: (a) Pre-stress NMOS transistor with normal biasing. (b) NMOS transistor under HCI stress conditions. A high V_{DS} generates a large current resulting in some hot electrons getting injected deep into the gate oxide (shown as the brown square). (c) After HCI stress, when the NMOS transistor is biased normally, it sees an increased threshold voltage (V$_{TH}$). The increase is significant (>100mV) when current is in the opposite direction as during the stress conditions. The increase in V$_{TH}$, however, is small when current flows is the same direction as during the stress conditions.

is based on a type of bi-stable element PUF that uses sense amplifiers as the core element. We first briefly describe the basic sense amplifier PUF structure before detailing the modifications necessary for HCI reinforcement.

Sense Amplifier (SA) PUF. Sense amplifiers (SAs) are clocked circuits that amplify small differential voltages into full swing digital values, which are used in memory read paths and as voltage comparators. Figure 2 shows a StrongARM sense amplifier circuit topology, which we use as the basis of our PUF. Under ideal conditions, an ideal SA would correctly amplify even the smallest of input differential voltages. In practice, however, variations in the devices of an SA may result in an offset (or bias), a measure of the natural tendency of the SA to resolve to a particular polarity. To ensure correct operation, the SA inputs need to have a difference larger than the offset.

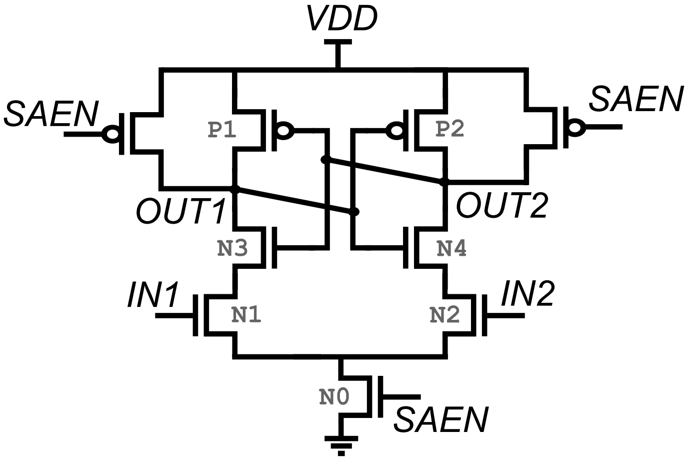


Fig. 2: StrongARM sense amplifier. The sense amplifier is fired by asserting the sense enable (SAEN). Based on the relative values of the inputs (IN1/IN2), it resolves to full swing outputs (OUT1/OUT2). When SAEN is low, the SA resets, with both outputs high.

Offset of a SA results from a combination of systematic and random variations. Systematic variations can be due to manufacturing gradients and layout asymmetries [21], and can be minimized by symmetric layout of matched devices. Random variations are a result of random uncertainties in the fabrication process such as random dopant fluctuation (fluctuations in the number and location of dopants in the transistor channel) [22] and gate line-egde roughness [23]. The effects can be mitigated by using larger devices [24].

To use as a PUF core, SA inputs (IN1 and IN2 in Figure 2) are shorted together (i.e., set to the same voltage, zero differential input) and the SA is fired. The SA will then resolve to a value determined by its individual offset [6]. The offset of the StrongARM SA (Figure 2) is a strong function of the difference

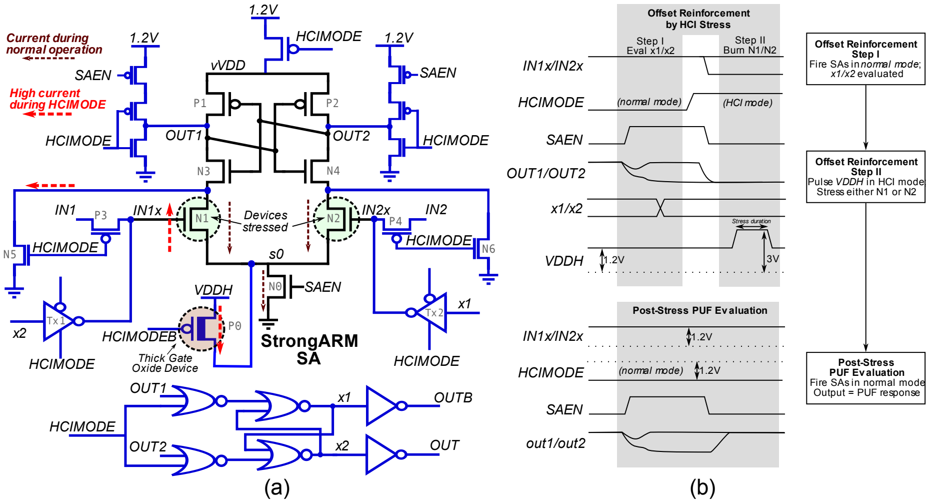


Fig. 3: HCI-enabled sense amplifier PUF (HCI-SAPUF) core schematic. The baseline StrongARM sense amplifier (shown in black) has been modified with peripheral circuits (blue) to enable HCI aging. Note the use of high oxide device to interface with the high 3V *VDDH*.

in V_{TH} of devices N1 and N2. The offset (and hence the reliability) of the SA can be increased by increasing the difference in the V_{TH} of N1 and N2. To maximize the randomness of the response, the SA in a PUF is designed to be as symmetric as possible (minimizing systematic offset), but with small devices in the matched pairs (maximizing offset magnitude). Hardware measurements have previously shown that SA PUFs have good randomness and uniqueness characteristics similar to those of SRAM PUFs [7]. At its core, an SA PUF is very similar to an SRAM PUF, and both belong to the family of bi-stable PUFs.

The offset of a SA is strong indicator of its reliability as a PUF [6, 7]. A SA with high offset (i.e., a strong bias to resolve to a particular polarity) will likely resolve to the same polarity across environmental variations and over aging. Measured hardware results have shown high reliability from SA PUFs with relatively large offsets (> 40mV) [7]. The exact offset cutoff for such high reliability may vary with technology, design specifics, or measurement conditions but the key idea here is that use of SAs with higher offsets will result in PUFs of higher reliability. We note that in practice, even with such high measured reliability, production PUF designs may still use some small amount of ECC to guarantee the needed error probability or bit error rate.

HCI-enabled sense amplifier (HCI-SA) PUF. We modified the StrongARM SA described above to enable the use of HCI to reinforce the individual offset of each SA (Figure 3(a)). The measurement of the offset polarity and reinforcement is done completely autonomously by the circuit. The basic StrongARM SA is shown in black, while the additionally HCI reinforcement circuitry is shown in blue.

Post-manufacturing, and before first use as a PUF, the HCI-SAs can be programmed for higher reliability by stressing either of `N1` or `N2` (Figure 3(a)). This is done individually for each SA, but since the offset reinforcement circuitry is self-contained for each SA, all SAs are reinforced in parallel. If offset_0 is the offset before stress, then the offset after stress (offset_S) should have the same sign as offset_0 and a higher magnitude.

The HCI-SAs operates in two modes, *normal mode* and *HCI mode*, controlled by the signal *HCIMODE*. In the *normal mode* (*HCIMODE*=0), the HCI-SAs act as normal StrongARM SAs. In the *HCI mode* (*HCIMODE*=1), one of `N1` or `N2` is stressed.

The reinforcement of offset is done in two steps: 1) offset polarity measurement and storage, and 2) HCI offset reinforcement.

Offset polarity measurement and storage. Which of `N1` or `N2` is stressed is determined by nodes *x1* and *x2*. The values of *x1* and *x2* are internally generated *normal mode* during the first step of offset reinforcement (Figure 3(b)). For example, if V$_{TH}$ of `N1` is higher than V$_{TH}$ of `N2` for a particular HCI-SA, and all other devices are matched, then in this step, the HCI-SA, when fired, will resolve to a 1 (*OUT1*=1,*OUT2*=0) and this value is latched as *x1*=1, *x2*=0.

HCI offset reinforcement. During the second step of offset reinforcement, *HCIMODE*=1 and `P3` and `P4` are disabled and the tri-state buffers `Tx1` and `Tx2` are enabled. The values at *x1* and *x2* force *IN1x*=1 and *IN2x*=0. This is followed by a pulse of high voltage (∼3V) at *VDDH* resulting in a high current path through `P0`, `N1`, and `N5`. The devices are sized such that most of the voltage drop (∼2.5-2.8V when provided with a pulse of 3V) is seen across the drain-source of `N1`. This creates the stress conditions for `N1` as described earlier in Section 2 and results in an increased V$_{TH}$ of `N1` when the SA is used later as a PUF in the *normal mode*. Note that the high voltage (*VDDH*) is connected only to a single thick gate oxide PMOS device (`P0`) per HCI-SA. A thick gate oxide device can withstand a higher V_{GS} without gate oxide breakdown and is a common process technology option since thick oxide devices are needed in the pads. The amount of stress (i.e., the extent of offset reinforcement) is controlled by the pulse width and the voltage of *VDDH*. A supply of ∼2.5-3V should be readily available on die as the I/O pad supply and hence the offset reinforcement step does not require a separate dedicated voltage supply.

Use case. The proposed use case for the HCI-SA PUF is for it to undergo a one-time HCI reinforcement step immediately post-manufacturing as detailed earlier in this section. This one-time step requires a few tens of seconds of HCI stress resulting in a permanent offset shift in each of the HCI-SA PUF core circuits in the direction determined by random process variations. After this one-time stress, the HCI-SA PUF is used just like a regular SA PUF by activating it in the *normal mode*. HCI is only applied to the devices at this initial reinforcement, and devices in the field are not subjected to any additional HCI stress.

4 Testchip Description

We designed and fabricated a $5.5mm^2$ full custom ASIC testchip in 65nm bulk CMOS technology that contained our HCI-SA PUF design (Figure 4). On each chip, there are 3200 HCI-SAs in a total area of $0.32mm^2$ that includes the scan flops for testing. Each HCI-SA was laid out in an area of $20.8\mu m^2$ that includes one thick gate oxide PMOS per cell (Figure 5(a)). The HCI-SAs were arranged in two arrays of 1600 elements each. All 1600 HCI-SA elements in an array share common *IN1*, *IN2*, and *SAEN* signals. The outputs are sent to flip-flops connected in a scan chain for easy read out. One array had the capability of self-contained offset reinforcement as described in Section 3. The other array was designed as a fail-safe and each HCI-SA element's reinforcement direction is loaded from a scan chain. However, the self-contained reinforcement circuits work as designed and all results in this paper are measurements from those arrays. The chips were packaged in a 132-pin PGA package and tested using a custom designed 4-layer PCB (Figure 5(b)).

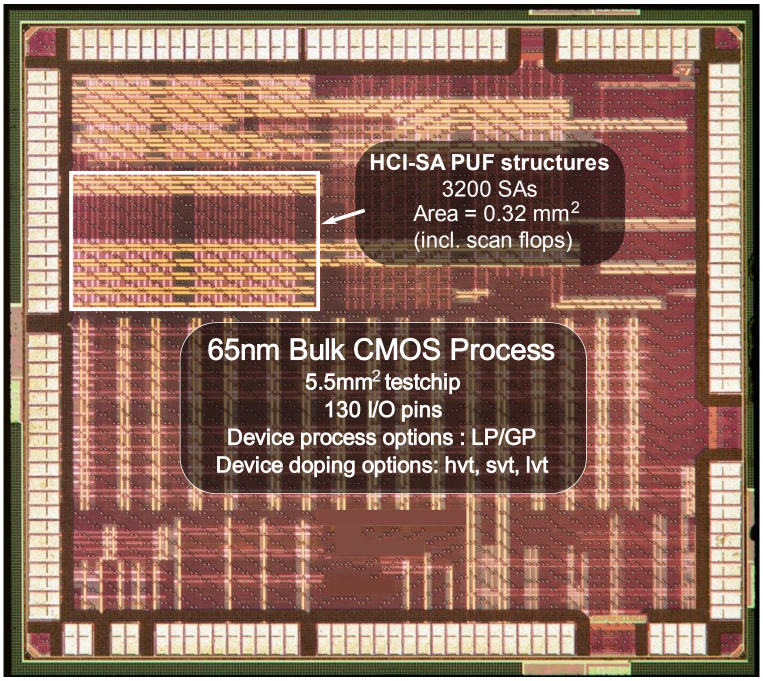
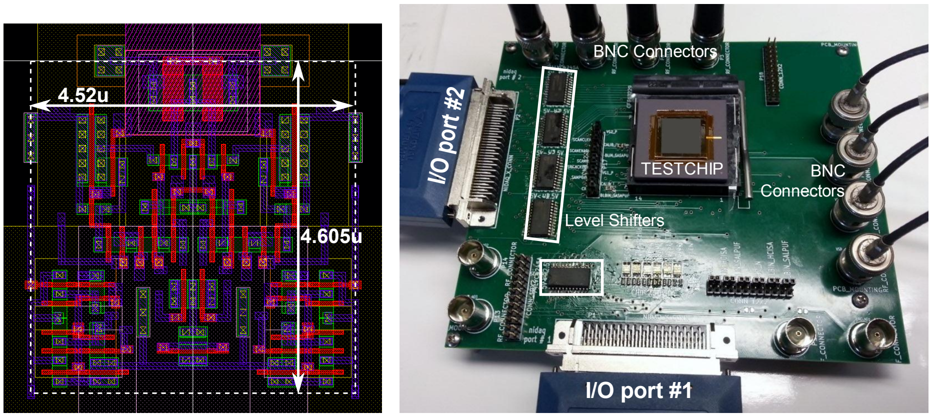


Fig. 4: Die micrograph of 2.5mm x 2.2mm, 130 pad, HCI-SA PUF testchip in 65nm bulk CMOS. There are 3200 HCI-SA PUF elements on each die in a total area of $0.32mm^2$. The die included a number of other unrelated projects.



(a) Layout capture of HCI-SA (b) Printed circuit board (PCB) designed for testing

Fig. 5: (a) Layout of a HCI-SA cell built in an area of 20.8μm^2 and repeat distance of 4.605μm (height) and 4.52μm (width). (b) Custom designed 4-layer test PCB used for testchip testing.

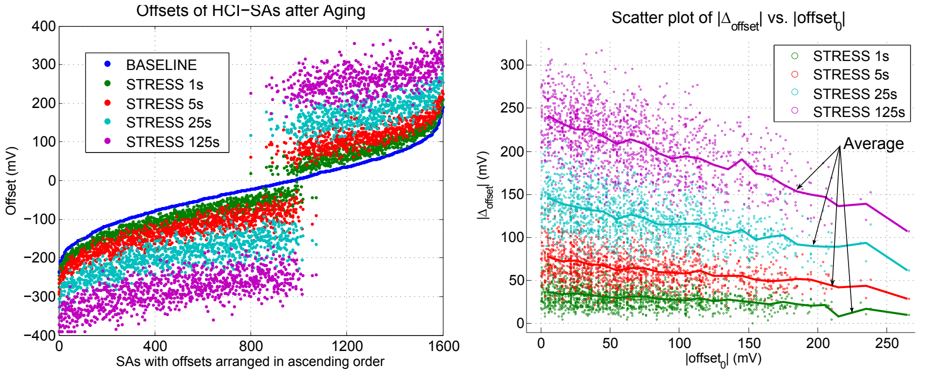
5 Measured Results

Using the testchip implementation described above, we measured the HCI-SA PUF element offset, reliability across environmental variations and aging, uniqueness, and randomness. The measurements are taken across a voltage range of ±20% of the nominal 1V VDD (i.e., 0.8V to 1.2V) and temperatures of $-20°$C, 27°C, and 85°C. We used a TestEquity Model 107 temperature chamber to fully enclose the test PCB during temperature variation testing.

The HCI-SAs were incrementally stressed using 3V *VDDH* pulses of width 1s, 4s, 20s, and 100s which resulted in a cumulative stress of 1s, 5s, 25s, and 125s. As mentioned earlier, the 3V supply is connected only to a thick gate oxide PMOS transistor per HCI-SA to avoid oxide breakdown of other devices in the circuit. Offset reinforcement only requires firing the HCI-SAs once (all fire together, since *SAEN*, *IN1*, and *IN2* are shared across all HCI-SAs in an array) in the *normal mode* followed by a pulse of high voltage at *VDDH* in *HCI mode*.

Offsets before and after stress. To evaluate the efficacy of HCI in altering the SA offset, we measure the offset of each of the 1600 HCI-SAs before and after HCI offset reinforcement. To measure the offset (either before or after HCI stress), the input differential (i.e., the voltage difference between *IN1* and *IN2*) is swept from -400mV to 400mV in steps of 10mV. At each step, the HCI-SAs are fired multiple times. The outputs after every activation are read out of the output scan chain. These are then post-processed to measure the offset of each HCI-SA in the array. Figure 6a shows the measured offset of all HCI-SAs of a chip, before and after HCI stress, when arranged in order of their offset_0 in ascending order. As expected, the shift in offset is higher for longer stress durations. For a 1s stress, the shift in offset is $\sim 10 - 50mV$ and for a stress of 125s, the shift is $\sim 150 - 300mV$.

Figure 6b shows the scatter plot of the magnitude of the shift in offset ($|\Delta_{offset}|$) vs.$|\mathit{offset}_0|$ for different stress durations. As expected, $|\Delta_{offset}|$ is higher for longer stress durations. For a 1s stress, $|\Delta_{offset}|$ is $\sim 10 - 50mV$ and for a stress of 125s, $|\Delta_{offset}|$ is $\sim 150 - 300mV$. Moreover, on average (bold lines in Figure 6b), $|\Delta_{offset}|$ is slightly larger for SAs with low magnitude of offset_0, which is desirable since SAs with low magnitude of offset would need a larger shift for reliability.



(a) Offset voltage of all 1600 HCI-SAs

(b) Scatter plot of magnitude of shift in offset

Fig. 6: (a) Measured offset of all 1600 HCI-SAs on a die before and after HCI stress. The SAs are arranged in order of their offset before stress in ascending order. The minimum magnitude of offset after stress of 1s, 5s, 25s, and 125s was found to be 5mV, 28mV, 92mV, and 158mV respectively. (b) Scatter plot of shift in measured offset after HCI stress vs. the measured offset before stress. The average shift in offset is plotted in bold lines.

Figure 7 shows the measured distribution of the offset for one chip before and after HCI stress at 1.0V and 27°C. The offset_0 has a typical normal spread as expected. The mean is slightly skewed ($\mu_0 = -23mV$) and may be due to a layout systematic bias. The standard deviation of offset_0 is 79mV. After stress, we see that the distribution splits into two groups. SAs with negative offset_0 have their offset shifted to the left (more negative) and SAs with positive offset_0 have their offset shifted to the right (more positive). The minimum magnitude of offset after stress of 1s, 5s, 25s, and 125s was found to be 5mV, 28mV, 92mV, and 158mV respectively.

Reliability across environmental variation. Figures 6a and 7 show that HCI stress can increase offset in SA. However, offset is an indirect measure of reliability and PUF reliability can be directly measured by multiple evaluations

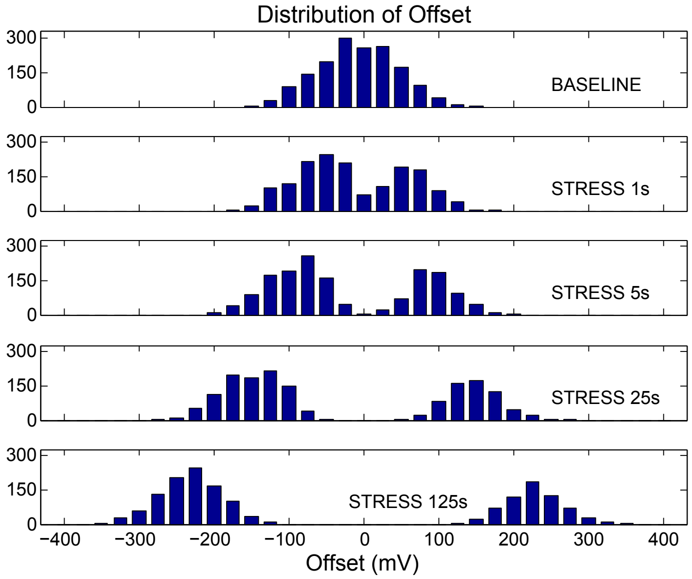
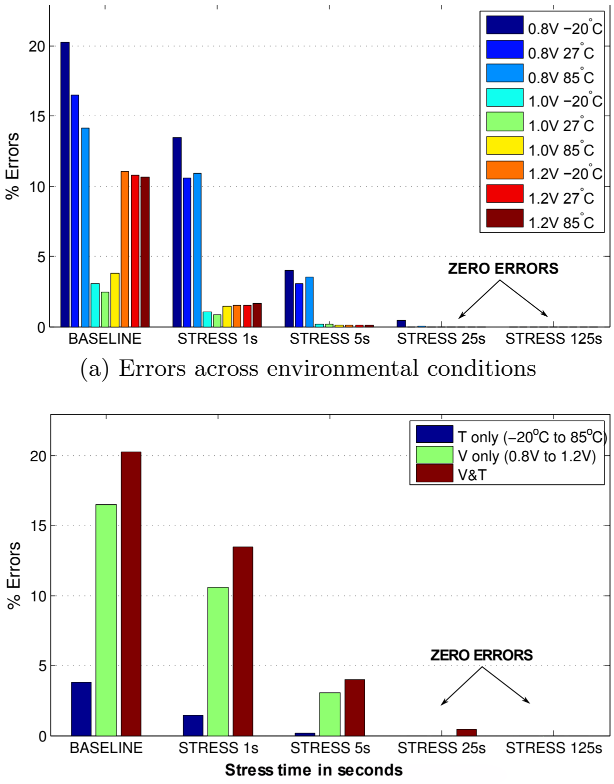


Fig. 7: Distribution of measured offset of all of 1600 self-programmable HCI-SAs on a die before and after HCI stress (measured at 1.2V and 27°C). After stress, we see that the distribution splits into two groups. SAs with negative $offset_0$ have their offset shifted to the left (more negative) and SAs with positive $offset_0$ have their offset shifted to the right (more positive).

across environmental variations and over aging. We first do the following for *small-scale* reliability measurements:



(a) Errors across environmental conditions

(b) Overall errors

Fig. 8: Reliability of HCI-SAPUFs shown as a percentage of errors (100 - % reliability). % Errors shown are the maximum errors across 100 evaluations. (a) Errors across all the environmental conditions. Errors were measured for voltage variations of ±20% from nominal 1V and temperatures of −20°C, 27°C, and 85°C. (b) Errors across only voltage, only temperature, and *all* voltage and temperature variations.

1. Perform 100 PUF evaluations at each possible combination of voltage (0.8V, 1.0V, 1.2V) and temperature (−20°C, 27°C, 85°C). Each PUF evaluation generates 1600 response bits corresponding to the 1600 HCI-SA elements in the array. The majority vote of the 100 responses at the nominal conditions conditions (27°C and 1.0V) is considered the golden response against which the response at other conditions will be compared.

2. At every combination of voltage and temperature, each of the 100 evaluations is compared to the golden response. We define $Error_{i,V_1,T_1}$ as the number of bits out of the 1600 HCI-SA outputs that do not match the golden response in the i^{th} evaluation at voltage=V_1 V and temperature=T_1°C.
3. We define % errors at a voltage-temperature combination ($Error_{V_1,T_1}$) as the maximum $Error_{i,V_1,T_1}$ across the 100 evaluations. The % errors across voltage-only variations ($Error_{V only}$) is defined as the maximum errors across the 100 evaluations at all voltage variations and at nominal temperature i.e., maximum of $Error_{0.8,27}$, $Error_{1.0,27}$, and $Error_{1.2,27}$. Similarly, % errors across temperature-only variations ($Error_{T only}$) is defined as the maximum errors across the 100 evaluations at all temperature variations and at nominal voltage i.e, maximum of $Error_{1.0,-20}$, $Error_{1.0,27}$, and $Error_{1.0,85}$. The overall % errors ($Error_{V\&T}$) is defined as the maximum errors across the 100 evaluations at all voltage and temperature combinations. Using this methodology, the reported % errors is the largest % of bits that were erroneous for any of the voltage/temperature conditions in any of the 100 PUF evaluations performed at that voltage/temperature. In other words, if we had an ECC that had the capability to correct that % of the bits, we would have a perfect response (i.e., matching the golden response) every evaluation.

These measures of reliability were taken for a die before and after different stress durations. Figure 8 shows the improvement in reliability for SAs with different stress durations and when expressed as % errors. Figure 8(b) shows the overall errors (across *all* voltage and temperature variations considered) reduce from 20.3% to 13.5%, 4.0%, 0.43%, and 0% when stressed for 1s, 5s, 25s, and 125s respectively. Note that this means that with 125s stress, there were no errors for any of the 1600 SAs, across all of 100 evaluations at all voltage and temperature combinations. Variations in voltage have a stronger impact on reliability as compared to temperature variations. For temperature only variations, the % errors reduce from 3.8% to 1.4%, 0.19%, 0%, and 0% when stressed for 1s, 5s, 25s, and 125s respectively; and for voltage only variations, the % errors reduce from 16.5% to 10.6%, 3.1%, 0%, and 0% when stressed for 1s, 5s, 25s, and 125s respectively. Figure 8(a) shows that highest number of errors are seen at low-temperature and low-voltage (0.8V and −20°C).

Large-scale measurements. Once we have identified the worst case corner for reliability as low-temperature and low-voltage (0.8V and −20°C), we perform large scale measurements at the corner. We ran 125,000 measurements at both worst case corner and the nominal conditions and found no errors for all of the 1600 SAs. Hence, emperically we demonstrated response bit errors $< 5 * 10^{-9}$. This is equivalent to a 128-bit key error rate of $< 10^{-6}$ which is the typical targeted failure rate for ECC implementations.

Permanence of reliability reinforcement with aging. To measure permanence of HCI-stress over aging, we measure the reliability of a HCI-stressed chip before and after aging, simulated in a shorter duration using elevated temperature and voltage. We bake a chip, originally stressed for 125s, at 150% of nominal 1V (= 1.5V) and 100°C for 93 hours, resulting in a Acceleration Factor of 161.4 and hence an aging of ~1.7 years for a chip operating at nominal conditions (1.0V and 27 °C) [8]. The large-scale reliability measurements were done after accelerated aging, suggesting that the impact of HCI-stress is not significantly reversed with aging.

Uniqueness. Uniqueness is a measure of how uncorrelated the response bits are across chips, and ideally the response bits should differ with a probability of 0.5. The Hamming distance of a k-bit response from ideally unique chips should follow a binomial distribution with parameters $\mathcal{N} = k$ and $p = 0.5$ and the mean of the HD distribution should be equal to k/2. For our case, we create 100 16-bit response words (i.e., k=16) from the measured outputs of the 1600 HCI-SAs on three chips. These words are generated at 27°C and 1.0V after the HCI-SAs have been stressed for 25s. Figure 9 shows that the the pair-wise HD of response bits from the three chips is close to ideal with means of 7.32, 7.36, and 7.50.

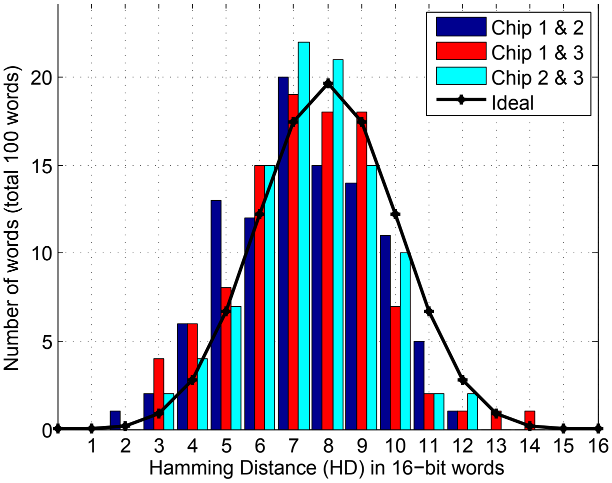


Fig. 9: Histogram of Hamming distance (HD) of response words from the HCI-SAs across three chips. Also shown is the probability mass function of the HD in responses from ideally unique chips. For the HD comparison, the response bits from 1600 HCI-SAs on a die are grouped to create 100 words of size 16 each. The pair-wise HD of response bits from the three chips is close to ideal with means of 7.32, 7.36, and 7.50.

Randomness. PUF randomness is a measure of the unpredictability of the response bits. In an ideal random response, the %1's and %0's in the response should be equal. In our measured response of 1600 HCI-SAs from three chips, after a 25s stress, the %1's were found to be 60.6%, 63.6%, and 61.4% which corresponds to entropy of 0.967, 0.946, and 0.962 respectively. The HCI-SAs were designed and laid out symmetrically (Figure 3 and Figure 4) and we suspect this small but consistent bias across chips to be due to some undesired systematic bias in the layout.

We note that previous studies of SA PUFs have concluded that their uniqueness and randomness characteristics are equal to or better than other PUF types [6, 7, 16].

Discussion of Results. Unlike conventional ECC techniques, the HCI-SAs do not require any helper data. The response reinforcement step is a one-time, self-contained step and does not require connection to an IC tester, and only requires that the chip be powered. It does not require any additional power supply and uses the available I/O pad supply for reinforcement. Reinforcement can be done at any time during manufacture test or even in-the-field after deployment. The required HCI stress time of 1-2 minutes, and could be done during IC burn-in or in-system-test, both of which are typically longer than the time needed for HCI stressing. Further, bit generation takes 1 cycle (compared to >2000 cycles for conventional ECC techniques) and can be achieved in ∼ access time of a SRAM (<1 ns for HCI-SAs in 65nm from our simulations). HCI-SA cell proposed is ∼20x the area of a SRAM or ∼10x the area of a SA. However, the core HCI-SA cell has much fewer devices. Many of the devices exist to enable a parallel, autonomous reinforcement of each cell. If we allow a serial and externally controlled reinforcement by removing the offset polarity storage circuitry (20 transistors), and sharing the thick gate oxide device (single largest transistor) across cells, we can reduce the HCI-SA cell to $< 10\mu m^2$, equivalent to 10x SRAM cell area or 5x SA area (compared to >23x SRAM cell area per bit for conventional ECC techniques).

6 Conclusions

We have presented a novel PUF response reinforcement technique based on hot carrier injection (HCI) to enhance the reliability of the PUF core. We described a novel sense amplifier circuit implementation that enables HCI-based stress response reinforcement and hence reliability improvement. Measured results from our custom ASIC testchip fabricated in 65nm bulk CMOS showed significant improvement in reliability across environmental variations (±20% of nominal VDD and temperature range of $-20°$C to 85°C) and good uniqueness and randomness.

The proposed technique can reinforce the PUF response in short stress times (i.e., tens or low hundreds of seconds), without impacting the surrounding circuits, and with high permanence (i.e., does not degrade significantly over time). The reinforcement mechanism is self-contained and requires minimal external intervention. No data related to the PUF core response bits are ever sent off chip, minimizing the information leakage. If we do not wish to use a dedicated supply, the existing I/O pad supply can be used as the high HCI supply voltage used for reinforcement.

By using response reinforcement techniques such as the proposed HCI method, the core PUF reliability can be significantly enhanced. Thus, the desired reliability can be achieved by much smaller or no ECC blocks. As the ECC overheads (area, power, delay, complexity) scale quickly with the ECC strength, we contend that the overall PUF efficiency can be boosted by using such response reinforcement techniques. While these methods are particularly useful for applications requiring high PUF reliability such as key generation, all applications can benefit from improved PUF reliability and efficiency.

References

1. Gassend, B., Clarke, D., van Dijk, M., Devadas, S.: Silicon physical random functions. In: CCS '02: Proceedings of the 9th ACM conference on Computer and communications security, New York, NY, USA, ACM (2002) 148–160
2. Suh, G.E., Devadas, S.: Physical Unclonable Functions for Device Authentication and Secret Key Generation. In: Proceedings of 44th ACM/IEEE Design Automation Conference DAC '07. (2007) 9–14
3. Lee, J.W., Lim, D., Gassend, B., Suh, G.E., van Dijk, M., Devadas, S.: A technique to build a secret key in integrated circuits for identification and authentication applications. In: Proceedings of Digest of Technical Papers VLSI Circuits 2004 Symp. (2004) 176–179
4. Lim, D., Lee, J.W., Gassend, B., Suh, G.E., van Dijk, M., Devadas, S.: Extracting secret keys from integrated circuits. **13**(10) (2005) 1200–1205
5. Holcomb, D.E., Burleson, W.P., Fu, K.: Power-Up SRAM State as an Identifying Fingerprint and Source of True Random Numbers. **58**(9) (2009) 1198–1210
6. Bhargava, M., Cakir, C., Mai, K.: Attack resistant sense amplifier based PUFs (SA-PUF) with deterministic and controllable reliability of PUF responses. In: Proceedings of IEEE Int Hardware-Oriented Security and Trust (HOST) Symp. (2010)
7. Bhargava, M., Cakir, C., Mai, K.: Comparison of Bi-stable and Delay-based Physical Unclonable Functions from Measurements in 65nm bulk CMOS. In: Custom Integrated Circuits Conference, 2012. CICC '12. IEEE. (Sept 2012)
8. Maes, R., Rozic, V., Verbauwhede, I., Koeberl, P., van der Sluis, E., van der Leest, V.: Experimental evaluation of Physically Unclonable Functions in 65 nm CMOS. In: ESSCIRC (ESSCIRC), 2012 Proceedings of the. (Sept. 2012) 486 –489
9. Yu, M.D., Devadas, S.: Secure and Robust Error Correction for Physical Unclonable Functions. IEEE Design & Test of Computers **27**(1) (2010) 48–65
10. Maes, R., Herrewege, A.V., Verbauwhede, I.: PUFKY: A Fully Functional PUF-Based Cryptographic Key Generator. In: CHES, Springer (2012) 302–319
11. Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In Cachin, C., Camenisch, J., eds.: Advances in Cryptology - EUROCRYPT 2004. Volume 3027 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2004) 523–540
12. Guajardo, J., Kumar, S.S., Schrijen, G.J., Tuyls, P.: FPGA Intrinsic PUFs and Their Use for IP Protection. In: Proceedings of the 9th international workshop on Cryptographic Hardware and Embedded Systems. CHES '07, Berlin, Heidelberg, Springer-Verlag (2007) 63–80
13. Bosch, C., Guajardo, J., Sadeghi, A.R., Shokrollahi, J., Tuyls, P.: Efficient Helper Data Key Extractor on FPGAs. In Oswald, E., Rohatgi, P., eds.: Cryptographic Hardware and Embedded Systems CHES 2008. Volume 5154 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2008) 181–197

14. Yu, M.D.M., M'Raihi, D., Sowell, R., Devadas, S.: Lightweight and secure PUF key storage using limits of machine learning. In: Proceedings of the 13th international conference on Cryptographic hardware and embedded systems. CHES'11, Berlin, Heidelberg, Springer-Verlag (2011) 358–373
15. Guajardo, J., Kumar, S.S., Schrijen, G.J., Tuyls, P.: Physical Unclonable Functions and Public-Key Crypto for FPGA IP Protection. In: Proceedings of Int. Conference Field Programmable Logic and Applications FPL 2007. (2007) 189–195
16. Bhargava, M., Cakir, C., Mai, K.: Reliability enhancement of bi-stable PUFs in 65nm bulk CMOS. In: Hardware-Oriented Security and Trust (HOST), 2012 IEEE International Symposium on. (june 2012) 25 –30
17. Vivekraja, V., Nazhandali, L.: Circuit-level techniques for reliable physically uncloneable functions. In: Hardware-Oriented Security and Trust, 2009. HOST '09. IEEE International Workshop on. (July 2009) 30 –35
18. Pobegen, G., Aichinger, T., Nelhiebel, M., Grasser, T.: Understanding temperature acceleration for NBTI. In: Electron Devices Meeting (IEDM), 2011 IEEE International. (Dec. 2011) 27.3.1 –27.3.4
19. Bhardwaj, S., Wang, W., Vattikonda, R., Cao, Y., Vrudhula, S.: Predictive Modeling of the NBTI Effect for Reliable Design. In: Custom Integrated Circuits Conference, 2006. CICC '06. IEEE. (Sept. 2006) 189 –192
20. Miyaji, K., Suzuki, T., Miyano, S., Takeuchi, K.: A 6t sram with a carrier-injection scheme to pinpoint and repair fails that achieves 57% faster read and 31% lower read energy. In: Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International. (Feb. 2012) 232 –234
21. Agarwal, K., Nassif, S.: Characterizing Process Variation in Nanometer CMOS. In: Proceedings of 44th ACM/IEEE Design Automation Conference DAC '07. (2007) 396–399
22. Keyes, R.W.: Effect of randomness in the distribution of impurity ions on FET thresholds in integrated electronics. **10**(4) (1975) 245–247
23. Oldiges, P., Lin, Q., Petrillo, K., Sanchez, M., Ieong, M., Hargrove, M.: Modeling line edge roughness effects in sub 100 nanometer gate length devices. In: Proceedings of Int. Conference Simulation of Semiconductor Processes and Devices SISPAD 2000. (2000) 131–134
24. Pelgrom, M., Duinmaijer, A., Welbers, A.: Matching properties of MOS transistors. **24**(5) (October 1989) 1433–1439