Exploiting Weak PUFs from Data Converter Non-Linearity - E.g. A multibit CT $\Delta \Sigma$ Modulator
Andreas Herkle, Joachim Becker, Member, IEEE, and Maurits Ortmanns, Senior Member, IEEE

Abstract—This paper presents a novel approach of deriving Physical Unclonable Functions (PUF) from correction circuits measuring and digitizing non-linearities of data converters. The often digitally available correction data can then be used to generate a fingerprint of the chip. The general concept is presented and then specifically evaluated on an existing Delta-Sigma ($\Delta \Sigma$) modulator whose outermost feedback DAC mismatches are greatly influencing the overall performance and thus need correction. The applied mixed-signal correction scheme reveals the intrinsic mismatches which are firstly used to linearize the $\Delta \Sigma$ modulator, but which can also be further analyzed. The intra-Hamming distance is initially determined to values less than 6% at nominal conditions and could be further reduced to less than 2% by applying different encodings. Regarding the distinctness of devices, the inter-Hamming distance is highly stable under all circumstances with a value very close to 50%. Though the influence of varying environmental conditions on the stability of repeated PUF readouts is negligible, inevitable deviations in the correction coefficients increase the intra-Hamming distance with respect to nominal conditions. As a result, 80 highly stable identification bits are obtained from the exemplarily used $\Delta \Sigma$ modulator.

Keywords—Physical unclonable function (PUF), DAC linearization, delta sigma modulation

I. INTRODUCTION

Over the last decade, Physical Unclonable Functions (PUF) became a topic of increasing interest in cryptographic and authentication applications. Utilizing physical variations in the silicon resulting from limited fabrication accuracy, PUFs offer great characteristics for this field namely uniqueness, unpredictability and unclonability [1], [2]. When stimulated with a digital input, they generate an individual digital response based on their unique mismatches. So called "weak PUFs" [3] are capable of generating only few or even a single response. Their outputs usually cannot be used directly in an authentication scheme based on challenge-response-pairs (CRP) as the secrecy of the information is revealed in that process. Still weak PUFs can be used for either identification purposes (fingerprinting) or as sources of randomness for key generation in cryptographic hardware.

Integrated analog-to-digital or digital-to-analog converters (ADC/DAC) suffer from random mismatches in their binary or equally weighted unit elements as they introduce Differential Non-Linearities (DNL) which reduce the effective resolution of the conversion result [4]. While the influence of a single mismatching unit element might be small enough to be neglected, the summed up output of many elements in a data converter results in an Integral Non-Linearity (INL) with the potential to significantly decrease the effective number of bits per conversion. Many techniques have been proposed in the literature to linearize such mismatch induced non-linearities by either using digital or analog techniques of compensation.

In this work, we propose to use the digitized representations of mismatch as unique, non-predictable and physical unclonable identifiers. This approach can be applied to almost any data converter architecture with deterministic unit element mismatch digitization and calibration, e.g. in SAR, Pipeline, $\Delta \Sigma$ and other ADCs as well as DACs. As a testing device and for proof of concept on a real hardware, a previously designed and published $\Delta \Sigma$ modulator is used [5], without restriction of the generality of the approach.

The paper is organized as follows. Section II outlines the general idea of extracting PUFs from the non-linearity correction of data converters. Section III provides a short introduction to the exemplary $\Delta \Sigma$ modulator, the measurement technique of the mismatching DAC and the applied correction scheme, which are used as test vehicle. Section IV presents the concept of deriving and improving identifiers from the correction circuitry at nominal conditions. Environmental influences are characterized by measurements analysis in Section V. Section VI concludes our work.

II. EXPLOITING WEAK PUFs IN DATA CONVERTERS

The general idea of the paper is to use the static non-linearities of data converters, which result from the device mismatch of binary or equally weighted unit elements, as
a unique identifier or weak PUF response. Those device mismatches manifest themselves as major source of DNL and INL in ADCs and DACs, and the main claim is that e.g. the DNL of two identical, but randomly mismatching converters, will never be the same. In the literature, a vast amount of data converter calibration techniques has been proposed over the last decades, from which a large number relies on the measurement of the mismatching unit elements, the digitization of this measurement, and the use of this digital information for linearizing the data converter either in the analog or digital domain.

The following short list shows the potential usage of the proposed technique, i.e. using unique digital calibration data of ADCs and DACs as PUF response. E.g. in [11], DAC unit elements have been measured by a slope converter and thereafter been calibrated. In [12], a SAR ADC has been implemented mismatch self-calibration with 6x6 bit digital correction data stored in the digital domain. In [13], a binary weighted current-steering DAC has been implemented based on current comparison, where 6x11 bit correction data is stored and used on-chip. In [14], the switched capacitor DAC in a multibit $\Delta\Sigma$ modulator has been calibrated, where 4 bit correction words for the 16 feedback codes of a 4 bit internal DAC have been used, yielding 4x16 bit. Also in [15], a split ADC architecture for a 6 bit flash ADC has been proposed, where the authors calibrate for non-monotonicity and need a minimum of 127x7 bit of calibration data.

$\Delta\Sigma$ modulators have been known for a long time to be rather tolerant to device mismatch, but only as long as no internal multibit quantizer and DAC are used. In almost all recent high-speed $\Delta\Sigma$ modulators, this is indeed the case. Therefore, the mismatches and thus non-linearities in the outermost feedback DAC as depicted in Fig. 1 become critical, since even though it is of low resolution, its non-linearity does directly affect the linearity of the overall $\Delta\Sigma$ modulator. Also for multibit $\Delta\Sigma$ modulators, many solutions have been presented to linearize the internal DAC. One set of solution for this problem is the principle of estimating the mismatches of the unit sources and to derive digital representations of those, which are then used to control specific correction mechanisms [16]–[18].

This paper now proposes to use such digital representations of data converter element mismatch or non-linearity as unique, non-predictable and physical unclonable identifiers. As a testing device and for proof of concept including measurement results, the previously designed $\Delta\Sigma$ modulator from [5] is used. Based on the digital estimation of the outermost feedback DAC unit element mismatch [18], we will show that reliable unique and random mismatch information can be extracted from the multibit DAC of a $\Delta\Sigma$ modulator with minimal overhead to the existing measurement circuit.

In order to highlight the competitiveness of this approach, Table I compares the results of the exemplarily analysed $\Delta\Sigma$ modulator and its DAC intrinsic PUF, called the DAC-PUF, to other published implementations of weak PUFs. The weak PUFs have mainly been selected to show different sources of randomness. Many other implementations of course exist exploiting these sources in similar ways. Besides the SRAM-PUF, many recent publications investigated various other types of memory. As can be seen in Table I, the metrics of our exemplary DAC-PUF are in the same range as previous implementations. In Table I, the functionality of the weak PUFs is also classified as either specifically built for its functionality, therefore dedicated, or as intrinsic when the functionality is derived from hardware serving other original purposes. Dedicated PUFs usually hold their IDs while for intrinsic PUFs, their value either has to be stored in an external memory or the initial functionality of the weak PUF has to be discarded as long as its ID is in use. In comparison, our approach is not only intrinsic but also keeps its original functionality while the PUF is either readout or stored. Just like weak PUFs based on memory circuitry, ADCs and DACs can be found in almost each mixed-signal circuit and System-on-Chip (SoC), thus their usage as chip identifiers or random number generator could be largely used without significant overhead.

In the following, we will restrict our analysis on the test vehicle, a continuous-time sigma delta modulator, which serves as a proof of concept and as an example how to characterize PUFs in data converters. The generality of the approach, extracting PUFs from data converters, remains thereby unaltered.

### III. Theoretical Preliminaries

#### A. Influence of mismatch on multibit $\Delta\Sigma$ modulators

$\Delta\Sigma$ modulation based analog-to-digital converters achieve high resolution by eliminating quantization errors in the converted analog signal by oversampling and a noise-shaping control loop. As a digital-to-analog converter is necessary in the feedback loop to oversample the system’s digital output from the analog input signal (DAC$_1$ in Fig. 1), mismatches in this DAC greatly influence the linearity of the $\Delta\Sigma$ ADC if this DAC is of multibit nature. The mismatch in each unit element

### TABLE I. WEAK PUF IMPLEMENTATIONS

<table>
<thead>
<tr>
<th>PUF Name</th>
<th>Source of randomness</th>
<th>Functionality</th>
<th>Storage</th>
<th>Metrics</th>
<th>Source of randomness</th>
<th>Functionality</th>
<th>Storage</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coating PUF [6]</td>
<td>capacitance of local dielectric particles within a coating layer</td>
<td>dedicated</td>
<td>implicit</td>
<td>intra-FHD</td>
<td>4.4%</td>
<td>50%</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Cross-coupled logic gates [7]</td>
<td>transistor threshold voltage variation</td>
<td>dedicated</td>
<td>implicit</td>
<td>inter-FHD</td>
<td>3.03%</td>
<td>46.15%</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>RO-PUF [8]</td>
<td>pairwise frequency comparison of ring oscillators</td>
<td>dedicated</td>
<td>implicit</td>
<td>intra-FHD</td>
<td>3.57%</td>
<td>49.97%</td>
<td>6536</td>
<td></td>
</tr>
<tr>
<td>SRAM-PUF [9]</td>
<td>initial value of SRAM-cells after power-up</td>
<td>intrinsic</td>
<td>explicit</td>
<td>inter-FHD</td>
<td>4%</td>
<td>49.01%</td>
<td>8192</td>
<td></td>
</tr>
<tr>
<td>Buskeeper-PUF [10]</td>
<td>initial value of buskeepers after power-up</td>
<td>intrinsic</td>
<td>explicit</td>
<td>inter-FHD</td>
<td>3.49%</td>
<td>49.96%</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>DAC-PUF [This work]</td>
<td>differential non-linearities of DAC unit elements</td>
<td>intrinsic</td>
<td>implicit</td>
<td>inter-FHD</td>
<td>4.4%</td>
<td>50.54%</td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>
within the multibit DAC results in a Gaussian distributed DNL, contributing to an overall integral non-linearity as is illustrated in Fig. 2. For high oversampling rates, dynamic element matching (DEM) techniques are often sufficient to compensate for these mismatches while in low oversampling \( \Delta \Sigma \) ADCs, which are commonly employed for higher bandwidth modulators, the effectiveness of DEM drastically drops and explicit correction of the DAC elements is necessary. Several state-of-the-art methods applicable to low oversampling \( \Delta \Sigma \) modulators propose to measure or estimate the mismatching DAC elements and to correct them in either the digital or the analog domain.

### B. Mismatch estimation and correction from [5]

The outermost feedback DAC\(_1\) in Fig. 1 of the exemplary modulator implementation is controlled by a thermometer coded flash quantizer such that DAC unit current sources are successively switched on for increasing output values, which makes the variation of the full-scale feedback being dependent on all unit element mismatches.

The method of measuring the mismatch of the current sources is based on the sequential inband insertion of a single-bit binary pseudo-random test sequence \( e_t(n) \) to each unit element of the outermost DAC\(_1\) in the analog feedback loop. By cross-correlating the modulator’s digital output \( D_{\text{out}} \) to the test sequence \( e_t(n) \), itself, the mismatch \( g_{et} \) of each unit element is revealed [18]:

\[
CCF_i = \frac{1}{L} \sum_{n=1}^{L} e_t(n)y_d(n) = g_{nom}(1 + g_{et})gyfss_t^2
\]  

(1)

with \( L \) being the number of samples taken for the computation and \( y_d(n) \) being the digital output \( D_{\text{out}} \) of the 4 bit flash quantizer. \( g_{et} \) is the individual gain error of the tested unit element with index \( i \). The computed cross-correlation value including the individual gain mismatch of the unit element is then used to compute a correction factor. This is achieved by relating the individual cross-correlation values to a chosen reference \( CCF_{ref} \):

\[
c_i = g_{nom} \left( \frac{CCF_i}{CCF_{ref}} - 1 \right)
\]  

(2)

This method was applied in [18] for digital post-correction and then in [5] by correction coefficients used in a binary weighted auxiliary DAC\(_{aux}\). The output of DAC\(_{aux}\), placed within the analog loop filter in parallel to the outermost feedback DAC\(_1\) in Fig. 1, corrects the output of DAC\(_1\) for each digital input code by adding or subtracting the feedback code dependent, summative current mismatches of the activated elements of the main DAC; thus, it inserts the INL of DAC\(_1\) with opposite sign in order to linearize it.

Single correction coefficients with a bitwidth of \( N_{\text{coeff}}=6\) bit are calculated for each unit element of DAC\(_1\) similar as in Eq. (2). The mean value of all CCF\(_i\) has been chosen as reference and additionally, the gain error of DAC\(_{aux}\) itself is taken into account [5]:

\[
c_{auxDNL,i} = \left( \frac{CCF_i}{CCF_{mean}} - 1 \right) \frac{CCF_{mean}}{CCF_{aux}} \cdot 2^{N_{coeff} - 1}
\]  

(3)

This basically reveals the DNL of the \( i \)-th DAC element versus the chosen reference. The INL based correction coefficients are calculated as the cumulative sum of the DNL correction coefficients:

\[
c_{auxINL,i} = \sum_{k=1}^{i} c_{auxDNL,k}
\]  

(4)

These correction coefficients are stored in an on-chip Lookup-Table (LUT) whose output controls the output current of DAC\(_{aux}\). Resulting from that method, each value in this LUT is the sum of the INL of the currently investigated DAC\(_1\) unit element and the previous element’s INL.

Following the main idea from Section II, we can now exploit these correction bits as weak PUF based on the random non-linearity of the DAC. But, even though this calibration method yields an INL correction of DAC\(_1\), it makes the LUT content not directly suitable as a weak PUF due to the strong correlation of its entries. This is obvious, since the true random value is the DNL of the individual sources, while the INL additionally contains the random information from previous sources. Nonetheless, the calculated DNL values are of course available during the time of estimation. The DNL values can either be saved with only little circuit overhead in an additional memory, or they could be recalculated from the INL LUT upon request with little combinatorial logic. DAC\(_{aux}\) has 8 bit of resolution and corrects 16 thermometer codes for the implemented 4 bit internal quantizer and 4 bit DAC\(_1\); thus, the overall size of the LUT is 128 bits which was found to be sufficient to linearize the \( \Delta \Sigma \) modulator to above 80 dB in this example [5]. But at this point it is important to note, that even when the INL values are stored as 8 bit representation in order to cover the worst case INL of the feedback DAC\(_1\) in Fig. 1 [5], single DNLs are obviously smaller, statistically by \( \sqrt{N_{\text{sources}}=16}=4 \), meaning 2 bits less.

In the rest of the paper, DAC\(_1\) is defined as the weak PUF-structure. It consists of 15 mismatching unit current sources (plus 1 exchangeable spare current source for background
implementations with an intra-HD below 5% are often rated
erroneous PUF readouts and thus non-zero intra-HD are caused
intra-HDs are calculated as the arithmetic mean of the FHD of
the response:

\[ FHD(R_k; R'_k) = \frac{\# \{ x, y : R(x, y) \neq R'(x, y) \}}{L} \]

(5)

To give statistical insights on the following encodings, their
intra-HDs are calculated as the arithmetic mean of the FHD of
N_d devices and M measured responses, given in percentage:

\[ HD_{intra} = \frac{100\%}{N \cdot M} \sum_{k=1}^{N_d} \sum_{j=1}^{M} FHD(R_k; R'_k) \]

(6)

Because the weak PUF functionality offers no CRP behav-
ior, the extra dimension of varying challenges is omitted.

While the ideal value of this metric is 0%, in reality
erroneous PUF readouts and thus non-zero intra-HD are caused
by electronic noise and environmental changes such as tem-
perature, supply voltage and aging. In the state of the art, PUF
implementations with an intra-HD below 5% are often rated
as sufficient for authentication applications when the inter-
HD is sufficiently good, as the first is usually handled with
error correction anyway [3]. For cryptographic applications, in
which error-free readouts are required, further techniques like
Temporal Majority Voting (TMV), Burn-in Hardening, Pre-
aging [20] and error correction [21] can be applied, which
are used to improve the implicit intra-HD of the source of
randomness on the device level. Those techniques are of
course easier to implement for already small intrinsic intra-
HD. For this reason, in the remainder of the paper we will
only investigate the intrinsic PUF response, and we do not try
to implement any further method to correct it, since the basis
of the paper is to exploit a new source of randomness, namely
data converter non-linearity, and not investigations to improve
PUF errors.

In terms of uniqueness, the inter-HD, as the number of differ-
futing outputs bits for the same challenge applied to different
devices, ideally follows a Gaussian distribution with a mean of
50%. Its standard deviation converges to \( \sqrt{N} \cdot 50\% \), according
to the central limit theorem for N-dimensional vectors of
independent evenly distributed random variables. Similar to
equation (5), the inter-HD between two distinct PUF devices of
the same length L is calculated as the bitwise distance between
their reference responses R_i and R_j:

\[ FHD(R_i; R_j) = \frac{1}{L} \| R_i \oplus R_j \| \]

(7)

In the rest of the paper, the inter-HD is calculated as the
arithmetic mean of all pairwise FHD of N devices, given in
percentage:

\[ HD_{inter} = \frac{100\% \cdot 2}{(N - 1) \cdot N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} FHD(R_i; R_j) \]

(8)

2) Shannon entropy based advanced metrics [22]: While
the intra- and inter-HD are sufficient measures of distinctness
and stability, the issue of predictability is not covered. There-
fore advanced performance metrics utilizing Shannon entropy
for independent identically distributed (iid) random variables
X_1, ..., X_z with probability mass function \( P(x) = P(X = x) \)
and joint Shannon entropy are calculated, which are defined as

\[ H(X) = - \sum_{x=0}^{1} P(x) \cdot \log_2(P(x)) \]

(9)

and

\[ H(X_1, ..., X_z) = - \sum_{x_1=0}^{1} \cdots \sum_{x_z=0}^{1} P(x_1, ..., x_z) \cdot \log_2(P(x_1, ..., x_z)) \]

(10)

These values will be presented and used to reveal potential
bias of the weak PUF introduced by design. In the following
we will use the "*"-selector similar to [22], which will indicate
that a calculation includes all bits of that dimension. I.e.
regarding \( R_k(x, y) \) in equation (11), it refers to all bits in the
fixed column y of the reference response \( R_k \), resulting in a
vector of length 16. Because of the missing CRP functionality, only the **Bias of PUF-Structure** measure \( H_{d,c} \) and the **Bias of PUF-Cell** measure \( H_{p,c} \) can be applied. **Dependency on local correlations** \( H_0 \) on DAC1 are also considered in order to examine if adjacent unit elements are influencing each other. Due to the absence of challenges, the global measure \( H_0 \) is identical to \( H_c \).

\( H_{d,c} \) is the measure for a general bias in the design, the entropy calculation is applied to all PUF-cells on a device. As the output of a single PUF-cell is of multibit nature, \( H_{d,c} \) is calculated column-wise and averaged over all devices:

\[
\hat{H}_{d,c}(y) = \frac{1}{N_d} \sum_{k=1}^{N_d} H_{d,c}(R_k(*, y)) \tag{11}
\]

A bias in specific PUF-cells on a device can be detected by calculating \( H_{p,c} \) as the entropy of a single PUF-cell on all device averaged over all devices:

\[
\hat{H}_{p,c}(x) = \frac{1}{N_d} \sum_{k=1}^{N_d} H_{p,c}(R_k(x,*)) \tag{12}
\]

To evaluate dependencies on local correlations, the measure of \( H_c \) is used. Since the response of single PUF-cells in our exemplary implementation of Fig.1 is a multibit value, the joint entropy can be applied in two different ways: the first calculation is done on the response of adjacent PUF-cells in order to reveal a potential correlation between the current sources, which are placed side-by-side in the transistor layout. The result of this calculation as an indicator of correlated sources is given as \( H_{c,adj} \) averaged over all devices:

\[
\hat{H}_{c,adj}(x) = \frac{1}{N_d} \sum_{k=1}^{N_d} H(R_k(x,*), R_k(x+1,*)) \tag{13}
\]

Gradient based correlations could also be checked for, when not only adjacent PUF-cells are considered but all of them together. However, as stated in [22], such a calculation requires a sufficiently high number of PUF-structures and therefore is insignificant in our case of [5], since only 20 prototype devices were available. The second calculation of joint entropies is done on the correction coefficients themselves in order to reveal a potential bias in the distribution of the values:

\[
\hat{H}_{c,val}(y) = \frac{1}{N_d} \sum_{k=1}^{N_d} H(R_k(*, y), R_k(*, y+1)) \tag{14}
\]

Due to the Gaussian nature of the mismatch in the sources, we have to expect \( \hat{H}_{c,val} \) to be much lower than 100, but its value still provides an insight about the average amount of predictable bits. \( \hat{H}_{c,val} \) as the mean of all bits also provides an additional qualifier when the extracted data is post-processed in different ways.

All of those measures are in the following used to characterize our exemplary DAC-PUF based on the non-linearity of the current source feedback DAC in a CT \( \Delta\Sigma \) modulator.

---

**Fig. 3.** Measured DNL of DAC1 of 20 devices with one exemplarily highlighted

**IV. MEASUREMENT RESULTS AND ANALYSIS AT NOMINAL CONDITIONS**

In order to prove the general idea of extracting randomness from data converter non-linearity and to analyze stability and uniqueness of the chosen exemplary implementation from Fig.1, 20 \( \Delta\Sigma \) modulator devices with the LUT based correction from [5] were measured and analyzed at nominal conditions of 25°C temperature and a supply voltage of 1.2 V. For each \( \Delta\Sigma \) modulator device, the measurement of the unit source mismatches of DAC1 according to Section III-B was performed 300 times.

**A. PUF characteristics of raw data at nominal conditions**

1) **Metrics of raw INL data:** First, the original 16x8 bit LUT entries for the INL correction scheme were analyzed as a reference for further improvements. Their metrics, listed in Table II as INL, showed to be not suitable for authentication purposes: the intra-HD was determined to a mean of 5.89 % and a standard deviation of 5.37 % while the inter-HD was determined to a mean of 43.79 % and a standard deviation of 12.47 %. Due to the cumulative summation of the measured random DNL values into INL values stored in the LUT, noise induced measurement variations not only influence the correction coefficients of their respective unit element but also spread to further coefficients, which artificially increases the intra-HD. This cumulative effect can also be seen in all of the entropies in Table III, indicating highly correlated values. Especially \( \hat{H}_{c,val} \) and \( \hat{H}_{c,adj} \) are very low. Therefore, this direct approach of using the LUT content as PUF is not further explained nor investigated.

2) **Metrics of raw DNL data:** An alternative approach is to use the 16x6 bit DNL values instead of the 16x8 bit INL values in order to remove the correlation between adjacent rows of the INL based LUT. As explained throughout Section III-A, the original unit element mismatch measurement reveals the DNL, from which the INL is build and the LUT filled. Thus, the DNL could be either stored right after measurement or extracted at any time from the INL by combinatorial logic.

---

**TABLE I**

<table>
<thead>
<tr>
<th>Device</th>
<th>Mean DNL</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAC1</td>
<td>43.79 %</td>
<td>12.47 %</td>
</tr>
<tr>
<td>DAC2</td>
<td>5.89 %</td>
<td>43.79 %</td>
</tr>
<tr>
<td>DAC3</td>
<td>12.47 %</td>
<td>5.89 %</td>
</tr>
</tbody>
</table>

**A3.2.2**

**TABLE II**

<table>
<thead>
<tr>
<th>Device</th>
<th>Mean DNL</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAC1</td>
<td>43.79 %</td>
<td>12.47 %</td>
</tr>
<tr>
<td>DAC2</td>
<td>5.89 %</td>
<td>43.79 %</td>
</tr>
<tr>
<td>DAC3</td>
<td>12.47 %</td>
<td>5.89 %</td>
</tr>
</tbody>
</table>
In Fig. 3, an illustration of the measured DNL values for 20 devices is given with an exemplary one highlighted. The index of the single unit elements corresponds to the index on the x-axis and the y-axis is the decimal representation of measured DNL values of $\text{DAC}_1$ given in LSBs (least significant bit) of the auxiliary DAC in Fig. 1. On first sight it appears that the DNL values of single unit DAC elements are random; thus, even when some elements from different devices might have similar values, the combination as a whole DNL curve is highly unique and distinguishable. The Gaussian shaped distribution of the unit element mismatches from those DNL estimations are shown in Fig. 4. The characteristics of the DNL values have then been further analyzed to find and improve their suitability as PUF response. Despite being designed for DNL values of 6 bits in [5], which was done in order to cover worst case unit element mismatches, the estimated DNL values from the measurements do not exceed a bitwidth of 5 bits, since the unit elements came from production with less variation as expected in worst case. Thus, the most significant bits (MSB) is not changing and therefore is statistically irrelevant, leaving significant 5 bits for the DNL. This is also seen in Figs. 3 and 4, where the decimal values of the DNL is limited within a range of less than ±15. When we calculate the metrics from Section IV of these 5 bit DNL readouts, already competitive intra- and inter-HD are obtained (see Table II under DNL of 5 b): the first is 3.49% without any post processing while the second approaches a nearly perfect value of 49.96%, confirming the method as a reasonable source of unique identifiers.

While using the DNL and restricting it to the statistically relevant 5 bit improves most of the entropy metrics, shown in Table II and III as DNL, a side-effect of the underlying Gaussian distribution in contrast to the ideally wanted uniform distribution can now be observed, namely in $\hat{H}_{\text{c,adj}}$ which is 77.6 indicating a predictability rate of about 22% for all bits. The effect can be explained as follows: as the encoding still covers all possible values, therefore also values with an absolute value exceeding ±7, it can be obviously seen in Fig. 4, that these values occur less often. Therefore the entropy reveals that the uncertainty about the DNL values is reduced by an estimate of 20% which equals to one of the 5 bits.

More specifically it indicates whether a value is expected to have an absolute value exceeding 7. The impact of this issue could be reduced by either limiting the values or by mapping back higher values to the encoding of lower values and thereby reducing the bitwidth to 4 or less. A short test of this assumption was investigated and resulted in an increased $\hat{H}_{\text{c,adj}} = 94.10$. However, further investigation on obtaining a uniform distribution from a Gaussian one lies out of the scope of this paper and instead, methods to further improve the intra-HD are presented next.

1) Investigation on stability and bias: A bitmap illustration of the lower 5 bit of the DNL statistics of all measured devices is given in Fig. 5. The instability shown in Fig. 5 (a) for each bit position is proportional to its rate of changes from its reference value over multiple measurements of the same chip; thus, it relates directly on the quality of the achievable intra-HD. Values near 0 in Fig. 5 (a) relate to highly stable outputs. Larger values are more likely to flip over multiple measurements and especially bits with a value close to 1 result in a bit probability of 50%, at which the output is random and only determined by noise. Fig. 5 (b) shows the averaged response of 20 measured devices against each other; thus, indicating the randomness over various PUF responses and indicating the achievable inter-HD. Most bits are near the ideal value of 0.5 while few show a slight bias; the first yield very unique responses of the LUTs, the latter a tendency to the same bit values from different devices.

The measure to analyze the influence of biased bits on the PUF quality is $\hat{H}_{\text{p,c}}$, which is 96.4 for the raw DNL values. For our limited sample size of only 20 devices, this value is still a good indicator that larger sets of devices might deliver a higher and thus even more ideal value. Removing slightly biased bits from the generated ID by masking would improve the distinctness of the devices even more but was not investigated in this work as the mean inter-HD was already almost an ideal 50%. Apart from that, masking of individual, randomly located bits would imply additional memory effort on a chip.

4) Influence of noise: Since in Fig. 1 the analog $\Delta\Sigma$ modulator itself is used for the estimation of the unit cell mismatches of its own feedback DAC, this measurement is prone to noise, and the lower bits are increasingly influenced by that circuit...
 occur many bit flips when the estimation varies by just a single LSB due to noise. This effect originates from the encoding of the unit element mismatches as two’s complement bit values, which is intended and advantageous for the correction circuit as DAC_aux is a binary weighted current source array and therefore of small size. But in a worst case scenario, a change of 1 LSB in the measurement of the DNL changes the value of all bits within one row of the LUT due to an over- or underflow (e.g. from ’01111111’ to ’10000000’ and vice versa). This yields heavily worse values of the intra-Hamming distance even though the underlying effect was very minor, i.e. one LSB change results in an 8 bits change within the LUT. The same is true for the 5 or 6 bits DNL and thus the impact on the intra-HD is strong. In the following, improved encodings are investigated in order to avoid this effect.

### B. PUF characteristics of improved encodings at nominal conditions

Two encodings mitigating this error were analyzed, namely thermometer coding and Gray coding. While thermometer coding as a continuous code is highly robust and its intra-HD is really low, it comes with a drastically decreased inter-HD. This results from the larger number of bits where the MSB are most likely set to 0 while the LSBs are most likely set to 1. An applied thermometer coding was proven to be insufficient regarding the inter-HD, which shrinks to 25.50%, and it also introduces a strong correlation of the single bits as also stated in Tables II and III. Due to this unacceptable outcome, thermometer coding will not be investigated any further.

#### 1) Gray coding

Gray coding is another continuous code where neighboring values differ by only a single bit in order to decrease errors in digital transmissions. It is also highly robust and distributes the bit values better which affects the inter-HD significantly less. Applying arbitrary Gray coding to the unmasked 5 bit DNL values, the intra-HD shrinks to 1.84% and therefore provides a larger number of 80 PUF bits with very low intra-HD. As before, a Gaussian distribution in the individual bits is observable as the application of Gray coding would cause a random shape of the curve to decrease errors in digital transmissions. It is also highly robust and distributes the bit values better which affects the inter-HD significantly less.
codewords in its standard form starts by filling the lower bits while the higher bits are less frequently set to 1. The application of a Gray code with such an order of codewords thus suffers from the Gaussian distribution of the calculated correction values. It can be seen in Fig. 7(d) as the red bars, that by applying Gray coding, the randomness of different bit positions drastically worsens from 50% towards 0% or 100%, which results in a worse inter-HD of only 39.53% and thus a penalty in uniqueness of the DUT; this is also seen in Table II. While the MSB and the 2 LSB hold reasonable probabilities close to 50%, the middle bits are very predictable due to their corresponding mean values of 0.97 and 0.30, which can also observed in a drastically drop of the entropies in Table III.

2) Customized Gray coding: The previous drawback is solved by a modified Gray coding which permutates the Gray codewords and matches them to the Gaussian probability distribution in such a way that the probability for a bit value equalling 1 for each individual bit is close to 50% and therefore unpredictable. In order to get the best results for the distribution of each individual bit, a search algorithm was used; thereby, we even allowed codewords which violate the minimum Hamming distance condition, i.e. neighboring codewords may have differences in up to 2 bits. Table IV shows a partial view on the permuted mapping of the DNL values to a standard Gray code and the modified version together with the probabilities of the DNL value according to the histogram distribution given in Fig. 4; please note that the original histogram was approximated with a best fit Gaussian distribution to obtain the probability values in Table IV.

The impact of the permutation within the Gray code on the probability distributions of individual bits is illustrated in Fig. 7. In Fig. 7(a), only the DNL = 0 is considered and therefore the probabilities of the 5 bits are equal to the probability of 10.8% listed in Table IV, i.e. there is only one code word for both the Gray code and its modified version, this code word is also shown in Fig. 7(a). In Fig. 7(b)-(d), the codeword mapping range is then expanded to consider more and more DNL values in both the positive and the negative direction, respectively 2, 4 and all DNL values. It can already be seen in Fig. 7(c), showing the accumulated bit probabilities for the DNL values -4 to 4, that the probability of the standard Gray coded 3rd bit approaches 1. This will not change significantly for the full range of DNLs as the probabilities for larger DNL values are already below 1% at that point as seen in Fig. 4. In contrast to this strong biasing, the customized Gray coding approaches more ideal distributions closer to 0.5 for each bit position in Fig. 7(a)-(c). When applying this customized Gray coding to all DNL values, their bit probabilities range again from 0.48 to 0.56, consequently improving the inter-HD from 39.53% (original) to 50.42% (modified). The resulting full range distributions are shown in Fig. 7(d) in comparison to the distribution of the original Gray encoding. Additionally, the entropy metrics of this encoding also show that the uncertainty about single bits on the device is again comparable to those of the unaltered 5 bit and masked 3 bit DNL values, see Table III.

As mentioned the modified Gray code was allowed to have neighboring code words with non-minimum HD; the presented code thus contains 3 neighboring codewords with a Hamming distance of 2, e.g. the DNL values -2 and -1 in Table IV. This minor violation of continuous coding could be proven to have only a negligible influence to the intra-HD: In comparison to the DNLunt., the intra-HD is increased in a negligible way to a value of 1.91%.

Overall, the modified Gray coding of the 5 LSBs of the DNL gives excellent intra- and inter-HD together with 80 unique bits, thus offering the best compromise for the exemplary ∆Σ modulator. The final distributions of the Hamming distances with the application of the customized Gray coding are illustrated in Fig. 8.

3) Achieving ideal equipartition by sign decision: A completely different approach is to use only the sign of the DNL values as source of randomness. A Gaussian distributed random variable with a mean of 0 is equally distributed in respect to the sign of its possible outcome. Therefore, a sign decision omits the effort of encoding while providing a bit
probability close to 50%. As some of the DNL values are either close to 0 or more unstable than others, they switch their sign more often. Masking such dark bits [20] and fixing them to a pre-determined value reduces our measured intra-HD to 0.01% while the inter-HD still holds a good value of 50.51% over the 20 available devices. Another positive side effect of this approach is that the sign of a Gaussian distribution with a mean value close to 0 is highly random. Regarding Fig. 4, an equipartition of the sign values can obviously be assumed. $H_{c,adj}$, in this case the uncertainty about the sign of the DNL values, confirms the assumption with a near perfect value of 99.9, coming from a ratio of 159 bits of positive sign and 161 bits of negative sign for the 16 DNL values per device and 20 measured devices. Additionally, the expected standard deviation for the Hamming distance between vectors of 16 independent uniformly distributed variables is 2 bits = 12.5%, see Section III-C1. Comparing this expected value to the obtained value from Table II also confirms the previous assumption.

For this encoding, an additional statistical test can be performed, namely the Fisher’s exact test, to investigate potential correlations between adjacent PUF-cells. This test is suitable for small sample sizes, when chi-squared tests become insignificant, and in our case is used to investigate potential shifts in the frequencies of the concatenated sign bits of adjacent PUF-cells. The number of occurrences of all possible combinations are filled into a 2x2 contingency table, which perfectly fits the original requirements for the test. The null hypothesis of evenly distributed frequencies was not rejected at the chosen significance level of 0.01, significantly indicating that the assumption of non-correlation holds true even for larger sample sizes. The lowest p-value found for the test pairs is 0.0546, which is an exception amongst all other p-values which are at minimum 0.1968. The outlying value of 0.0546 comes from the 15th unit element, in which a slight bias in its upper 2 bits can be observed, see Fig. 5(b).

The trade-off for these near perfect metrics is the lower number of usable bits in our example, which is 1 per unit element and therefore 16 per ADC. Nonetheless, in a transceiver where at least 2 ADCs are commonly present, an almost perfect 32 bit ID could be generated with the exemplary $\Delta \Sigma$ modulator.

V. MEASUREMENT RESULTS AND ANALYSIS AT CHANGING ENVIRONMENTAL CONDITIONS

The output of a circuit used in key generation is required to be stable over variations of multiple environmental parameters, as e.g. supply voltage, temperature or aging. For this purpose, often error correction encoding techniques are employed, which are adapted from channel coding theory. But obviously, the more stable the intrinsic response of the random source is, the less coding effort is needed to generate a stable ID. This is the same in channel coding, where more effort is needed for the code and redundancy when the errors due to channel noise increases.

The correction scheme from [18] can be used to improve the linearity of the DAC not only regarding mismatch but also for environmental changes such as supply voltage variation and temperature change. Considering that the coefficients are estimated using a test signal which passes through the whole system, it is affected not only by the mismatch of the unit elements but by each circuitry the $\Delta \Sigma$ consists of. Estimating under changed environmental conditions therefore implies a change of the coefficients for the unit elements. This makes it highly stable regarding its original purpose which is the linearization of the feedback DAC. It will be shown that an expected increase in the intra-HD in this case is a deceptive information as the stability between single measurements at random, yet fixed environmental conditions is not affected but only in reference to measurements nominal conditions.

A. Impact of supply voltage variations

For the examined $\Delta \Sigma$ modulator correction technique a measurement over supply voltage variation was already given [5]. The results of these measurements showed that correction based on a once generated LUT content at nominal supply voltage of 1.2 V was still sufficient to obtain the required DAC linearization over supply voltage variations in the range from 1.15 V to 1.35 V. While this approach is adequate for the original linearization purpose, recalculating the correction coefficients at different supply voltages results in slight variations in the LUT content, thereby increasing the intra-HD with respect to nominal conditions.

In order to analyze the stability of the PUF response over supply changes, additional measurements were done with the same supply voltage variations. The LUT contents, which reflect the measured DAC non-linearity, calculated at different voltages were then compared to the reference values obtained at the nominal supply of 1.2 V. As already mentioned in Section V, an increase in the intra-HD has obviously to be expected as any environmental influence might slightly change the DAC non-linearity as source of randomness. Moreover it will change the $\Delta \Sigma$ modulator itself as measurement tool of
said DAC non-linearity, which is then reflected in changing LUT content and PUF response.

Exemplary measurements of the highlighted DNL curve from Fig. 3 are shown for the mentioned range from 1.15 V to 1.35 V in Fig. 9(a). While single measurements at any supply voltage have been found comparably stable, the DNL curve slightly differs over various supplies when compared to the reference obtained at the nominal supply voltage of 1.2 V.

The analyzed intra-HDs of the LUT contents at various supply voltages compared to the reference from 1.2 V are illustrated as blue curve in Fig. 10(a). Increased intra-HDs were determined to a maximum of 11% at the lowest functionally supply voltage of 1.1 V and 9% at the maximum allowed supply stress of 1.35 V. Especially at low supply the DAC cascode and then current source transistors approach linear mode, and thus significantly more non-linearity and DNL changes are expected, as also measured in [5]. Also, the modulator itself as measurement device becomes worse. Nonetheless, within a more realistic supply variation of ±5%, the intra-HD is still below 5%, which is a remarkable stability for a prototype design.

Additionally, the more important measure is how stable a repeatedly extracted PUF response at a single changed supply voltage can be obtained, and thus what intra-HD of a device is measured at one non-nominal supply, in contrast to the intra-HD of a device compared over different supplies. This is because with a coarse measurement of the supply, and the unique PUF response at this supply the device would still present a very unique and stable device ID without encoding. The green curve in Fig. 10 (a) shows that these dynamic intra-HDs are very stable with values less than 2%. This reveals that even though the PUF response noticeably changes between various supply voltages, the response is very stable at any single supply voltage. Also shown in Fig. 10 (a) as red curve are the inter-HDs at the changed supply voltages which all hold mean values of about 50%. This shows that the distinctness between the devices is quite unaffected by supply changes.
B. Impact of temperature variations

In addition to supply voltage variations, changes in temperature are known as strong environmental influence on PUF responses. Obviously, also the feedback DAC in the ∆Σ modulator in Fig. 1 as well as the ∆Σ modulator itself is changed due to temperature variations. In [5] temperature measurements similar to those of supply voltage variations were done with a single estimation of the DAC non-linearity at room-temperature and a subsequently non-changing LUT content; again, the successful linearization could be proven over a temperature range from -30°C to 60°C, even though a drop at smaller temperatures could be seen. As for the supply voltage variations, we consequently also measured the non-linearity extraction and thus LUT content and PUF response over this temperature range in steps of 5°C, because again changes in the PUF response have to be expected. Exemplary measurements of the highlighted DNL curve from Fig. 3 are shown for the mentioned range from -30°C to 60°C in Fig. 9 (b). Firstly, the trend of the curve is still highly unique and distinguishable as the one in Fig. 3 but obviously some of the unit elements are strongly influenced by the temperature variations.

Similar to the results from supply voltage measurements, the average intra-HD and its standard deviation have been determined for all 20 devices. As expected, when comparing the 25°C reference with the readouts at non-nominal temperature, the intra-HD rises by about 2% per 5°C, shown in Fig. 10 (b) as blue curve. It is notable that cooler temperatures are affecting the intra-HD in a slightly stronger way resulting in a value of 8.87% at -10°C in opposite to a value of 6.1% at 60°C. This difference results from the modulators own resolution being more affected at lower temperatures [5]. Since the modulator is used as measurement tool for the DAC and thus PUF response, it must be expected that worse resolution yields worse precision in the PUF extraction.

If the dynamic calculation is repeated, where the stability of the PUF response at the non-nominal temperature is analyzed by multiple readouts at that temperature and a comparison to the reference at that same temperature, the intra-HD is again very low and constant, which is illustrated in Fig. 10 (b) as green curve. Equivalent to the variations of the supply voltage, the inter-HD is nearly unaffected holding a value of about 50% illustrated in Fig. 10 (b) as red curve.

In conclusion, the PUF response is highly unique at various temperatures, similar to the case of supply voltage measurements; thus, together with a coarse measurement from a temperature sensor, the PUF response still presents a stable device ID without any further encoding.

VI. Conclusion

In this paper, we present a novel approach of deriving weak physical unclonable functions (PUF) e.g. for unique device identifiers from the non-linearity of data converters by taking advantage of already implemented mismatch correcting techniques. Therefore, the general idea is outlined and examples from the state of the art are given. In order to emphasize the approach with hardware measurements, one of our previously built multibit CT ∆Σ is further investigated. In the shown example, the digital coefficients of the applied feedback DAC linearization scheme are analyzed. We show that their metrics are sufficient to fulfill a weak PUF functionality. The inter-Hamming distance of the device’s differential non-linearities are determined to an almost perfect value of 49.96%. By applying a suitable encoding to prevent over- and underflow bit flips of two’s complement encoded decimal values, the intra-Hamming distance is reduced to values less than 2%.

Though the correction method is very stable over supply voltage and temperature variations concerning the linearity of the ∆Σ modulator, deviations caused by environmental influences yield slightly different non-linearity coefficients and thus PUF responses. Beside that this effect results in greater intra-Hamming distances in relation to a nominal reference of $V_{dd} = 1.2$ V and $T_{nom} = 25°C$, the distinctness of the devices is unaffected. Therefore coding effort is limited to environmental influences and could be omitted when operating close to nominal conditions. The ∆Σ modulator under test provides 80 unique bits for identification purposes or cryptographic key generation. Keeping in mind that e.g. almost every transceiver has at least two of those ∆Σ modulators for I/Q channel digitization, 160 unique identification bits could be readily used on those transceiver ASICs with very high reliability and very low coding effort for stability over environmental conditions. We believe that this non-linearity extraction is a generally applicable way to generate weak PUFs on almost every System-on-Chip which usually include a significant number of DACs and ADCs. Further research could investigate other DAC principles presumably being less prone to environmental influences, e.g. discrete time ∆Σ modulators or SAR ADCs based on more stable capacitor ratios.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The authors would also like to thank M. Hiller (Technical University of Munich) and S. Puchinger (University of Ulm) for technical discussions.

References


