Maximum-likelihood decoding of device-specific multi-bit symbols for reliable key generation

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Maximum-Likelihood Decoding of Device-Specific Multi-Bit Symbols for Reliable Key Generation

Meng-Day (Mandel) Yu†‡§, Matthias Hiller†, Srinivas Devadas§
†Verayo, Inc., San Jose, CA, USA
‡COSIC / KU Leuven, Belgium
§Institute for Security in Information Technology / Technische Universität München, Germany
myu@verayo.com
matthias.hiller@tum.de
devadas@mit.edu

Abstract—We present a PUF key generation scheme that uses the provably optimal method of maximum-likelihood (ML) detection on symbols derived from PUF response bits. Each device forms a noisy, device-specific symbol constellation, based on manufacturing variation. Each detected symbol is a letter in a codeword of an error correction code, resulting in non-binary codewords.

We present a three-pronged validation strategy: i. mathematical (deriving an optimal symbol decoder), ii. simulation (comparing against prior approaches), and iii. empirical (using implementation data). We present simulation results demonstrating that for a given PUF noise level and block size (an estimate of helper data size), our new symbol-based ML approach can have orders of magnitude better bit error rates compared to prior schemes such as block coding, repetition coding, and threshold-based pattern matching, especially under high levels of noise due to extreme environmental variation. We demonstrate environmental reliability of a ML symbol-based soft-decision error correction approach in 28nm FPGA silicon, covering -65°C to 105°C ambient (and including 125°C junction), and with 128-bit key regeneration error probability ≤ 1 ppm.

Index Terms—Physical Unclonable Function, Maximum-Likelihood, Non-Binary Codewords, Soft-Decision, Key Generation

I. INTRODUCTION

Silicon Physical Unclonable Function (PUF) [8] [9] circuits exploit CMOS manufacturing variation in security applications, allowing a silicon device to be authenticated based on a challenge/response evaluation of difficult-to-replicate manufacturing variation. Since evaluation of a PUF response is physical, PUF response bits are noisy and noise increases with environmental variation. For a key generation use case, the basic PUF circuit needs error correction to account for evaluation noise; this requires helper data.

Figure 1 offers a codeword-centric view of PUF error correction, using code-offset helper data [6] [12]. During provisioning, keying material KM is used as input to an error correction code (ECC) encoder to produce a codeword CW. The codeword is bit-wise XOR’ed with a reference query of the PUF response RS, and stored as helper data HD. During regeneration, HD enables the return of a noisy query of the PUF response RS into a form where ECC decoding can be applied to remove noise. In effect, PUF is used to “store” KM using chip-unique manufacturing variation where KM can be hashed/down-mixed into a key (hash not shown in Figure).1

It is natural to think of codewords as composed of binary values vs. symbols from a higher-order alphabet. In our work, we perform PUF error correction using symbol-level (vs. bit-level) recovery and non-binary codewords. Our proof-of-concept Maximum-Likelihood Symbol Recovery (MLSR) implementation reduced bit errors to \( \sim 0.01\% \) at a 125°C key regeneration junction temperature (provisioning at room temperature), and produced a soft-decision metric that allows a simple soft-decision decoder to “mop up” remaining errors. Our soft-decision symbol parity decoder also uses ML decoding and corrects the most-likely corrupted symbol based on the least distance between the most-likely and 2nd-most-likely symbols, reducing the \( \sim 0.01\% \) bit failure rate to show an empirical key regeneration failure rate \( \leq 1 \times 10^{-6} \).

A. Contributions

The main contributions of the current work are:

- first to use ML symbol-based decoding on PUF bits
- first PUF error correction scheme capable of orders of magnitude bit error improvement over the standard repetition coding and BCH approaches

Fig. 1. PUF error correction using code-offset for provisioning (left) and regeneration (right): HD relates RS' to CW' to allow error decoding.

1Figure 1 assumes an externally chosen secret [12]. Using the syndrome construction [6], the secret (i.e., KM) can be internally derived, from RS.
new method to help counter recent helper data attacks based on linear-XOR or linear-shift malleability

B. Organization

Section II provides background. Our MLSR method is described in Section III, and validated in Section IV using a three-pronged (mathematical, simulation, empirical) approach. Section V covers recent helper data manipulation attacks which MLSR helps to address. We conclude in Section VI.

II. BACKGROUND

A. Related Work

**PUF Key Generation.** Silicon PUF key generation was first introduced using Hamming codes in [9] and more details were presented in [22]. More practical implementations that accounted for a wider range of environmental variation were described in [1] [16] [23] [19] [24] [17] [14] [11]. The recent results in [18] at 22nm have restricted temperature data (25°C to 50°C) that does not cover many operating conditions.2

**Helper Data Manipulation.** The issue of helper data manipulation was addressed for the biometric realm in [2] [7], with robust fuzzy extractors, but not so much in a manner specific to silicon PUFs. Their use of a helper data hash does not address recent helper data manipulation attacks in [13] [5]. A more recent work [11] also addresses helper data manipulation and in the context of a PUF but requires a hash function.

**Maximum-Likelihood (ML).** ML is a provably optimal decoding method used to recover bits transmitted over a noisy physical medium [20]. Repetition coding, which was used for PUF key generation in [1] [17] [14], uses a decoder that can be interpreted as ML with (degenerate) one-bit symbols. A recent publication [11] uses ML for 2nd stage decode but not 1st. We are the first to apply ML on multi-bit symbols, and to do so early on, during the 1st stage, which results in orders of magnitude bit error improvements over standard methods.

B. PUF Building Blocks

**Arbiter PUFs.** Arbiter PUFs were introduced in [15] (Figure 2, top). PUF output is derived from a race condition formed by successive delay stages each comprising a crossbar switch. A challenge bit \( x_i \), \( 0 \leq i < k - 1 \), for each of the \( k \) stages determines whether the parallel path or cross path is active. Collectively, \( k \) challenge bits determine which path is chosen to create the transition at the top input to the arbiter latch, and similarly for bottom input. The difference comparison between the two delays determines whether the PUF produces a “1” vs. “0” response bit. Design layout has to be symmetrically matched such that random manufacturing variation would affect the response.

**k-sum Ring Oscillator PUFs.** In a \( k \)-sum PUF [24], each of \( k \) stages is replaced with a pair of identically manufactured ring oscillators (Figure 2, middle). Delay propagation is mimicked by digitally adding delay values for each stage depending on the challenge bit associated with that stage. Mathematically, \( k \)-sum PUF and Arbiter PUF can be reduced to the same topology by a linear remapping of the challenge bits [21].

C. n-XOR Construction

The basic Arbiter PUF (and its equivalent \( k \)-sum PUF) can be modeled using machine learning attacks [21] when a sufficient number of challenge/response pairs (CRPs) is exposed. To scale up security against modeling attacks, XORs can be applied to outputs from \( n \) copies of these manufacturing variation building block primitives (Figure 2, bottom), to require more CRPs, computational power, time, etc. to model the composite \( n \)-XOR PUF.

III. MAXIMUM-LIKELIHOOD SYMBOL RECOVERY

A. Adapting n-XOR PUF for Key Generation

The \( n \)-XOR PUF, in its native form, is an “open system” in that user and adversary alike have unrestricted access to its challenge/response interface. When used for authentication, the number of exposed CRPs is potentially very large, easily in the millions or billions, making the PUF susceptible to machine learning predictive attacks. Fortunately, in our specific key generation use case (similar to [19]), we generate a fixed number of keying bits from the PUF, thus we reveal only a limited number of response bits to the adversary, e.g., several thousand bits.

Our implementation uses 4-XOR (programmable to more) and exports \( \leq 3000 \) response bits as helper data for a 128-bit key.3 Multiple research groups with attacks spanning 3+ years [21] [4] require more than 3000 CRPs to break 4-XOR PUF.

2Typical commercial (junction) temperature range: 0°C to 85°C; for industrial: -40°C to 100°C; and for military: -55°C to 125°C.

3The entire challenge schedule, starting from the seed, is hardwired.
B. ML vs. Bounded-Distance/Threshold-Based

We offer two explanations why ML has a lower error probability than a bounded-distance decoding approach associated with most PUF block decoders (e.g., BCH) [1] [23] [17], and equivalently for a threshold-based approach [19]:

1) Visual: Figure 3 shows two sequences, with the goal of deciding whether a received sequence belongs to seq_a or seq_b. In bounded-distance/threshold-based decode, seq'' decodes properly to seq_b, since it is within a certain preset distance or threshold represented by the circle around seq_a. On the other hand, seq' decodes improperly since it falls outside the circle. In ML, since seq' is closer to seq_b than to seq_a (i.e., falls to the right of dotted line), even though it falls outside the circle, it still decodes properly to seq_b.

2) Toy Example: Let seq_a = 1010101, and seq_b = 1010010. If a received sequence is 0101010, although 4 out of 7 bits have flipped with respect to seq_b, ML decodes to seq_b since received sequence is closer to seq_b than seq_a.

C. Device-Specific Symbols with Ease of ML Decode

We treat PUF response bits as l-bit sequences through which we encode and decode α-bit symbols. This is done for β-processing blocks. Now, we establish some notation.

- PUF is a basic PUF building block (with no XOR), PUF_i is the i-th PUF instance.
- KM is keying material, HD is helper data (l · β bits).
- RS is a response sequence of l bits, and RS_{PUFi} is a l-bit response sequence of i-th PUF.
- s is an α-bit symbol. s'' is the ML recovered version.
- seq' is the second-most-likely recovered version.
- xresp is the l-bit response sequence (helper data block) obtained after all RS_{PUFi} selected using s are XOR'ed.
- xresp_f is helper data block read back into the decoder.
- c^f means a value derived during regeneration.
- c' is soft-decision confidence information for a symbol.

Now, we shall encode in a manner that yields a simple ML decoder. We give an α=4-bit example:

1) Symbol Encoder: HD is a function of symbol bits (derived from KM) and RS from multiple PUFs. Symbol bits s[0...3] each drive a mux select line of a two-bit to one-bit mux (Figure 4, top). The construction looks like n-XOR PUF except 2:1 muxes are used to select which PUF response RS_{PUFi} to use prior to the XOR function. The XOR'ed l-bit result, denoted xresp, forms the helper data HD for one symbol. Encoder uses four 2:1 muxes and one XOR gate.

2) Symbol Decoder: Figure 4, bottom, depicts a symbol decoder, which contains a symbol encoder. The helper data from encoding xresp is read in as xresp'' since the symbol has yet to be recovered. To perform ML decoding, a 4-bit counter is used to iterate through all 16 symbols (indicated by s?), and the most likely symbol is stored. The symbol with minimum distance is found by performing bit-wise XOR of xresp'' (from HD) and candidate regenerated l-bit sequence (xresp')^s? representing a symbol, and incrementing a counter whenever a bit mismatch is seen (using one XOR gate and one log_2(l)-bit adder). A comparator (subtractor) is used to find 1st nearest distance and 2nd nearest distance (uses two stage buffering of log_2(l) bits each), and their difference (using a subtractor) computes soft-decision confidence c'.

A 4-bit symbol encoder/decoder with l=128 bits requires 12 SLICES (440 gate equivalents) in a Xilinx Artix-7 FPGA.

D. 4-bit Symbol Encoder / Decoder Pseudo-Code

Here, ⊕ is bit-wise XOR, (.)? is an if-then-else statement, and wt is the Hamming weight operator counting the number of 1s. Note that tmp^f (temporary variable), RS^f, xresp^f are each l bits, and dist (symbol distance) is log_2(l) bits.

**bold italics** indicate a helper data block; (xresp')^s? is not a helper data block but an intermediate result used during symbol recovery.
Symbol Encoding:
\[
\begin{align*}
tmp^a & = (s[0] = 1) \cdot RS_{PUF1} \oplus RS_{PUF0} \\
tmp^b & = (s[1] = 1) \cdot RS_{PUF3} \oplus RS_{PUF2} \\
tmp^c & = (s[2] = 1) \cdot RS_{PUF5} \oplus RS_{PUF4} \\
tmp^d & = (s[3] = 1) \cdot RS_{PUF7} \oplus RS_{PUF6} \\
x_{\text{resp}}^s & = tmp^a \oplus tmp^b \oplus tmp^c \oplus tmp^d
\end{align*}
\]

Symbol Decoding:
\[
\begin{align*}
\text{dist}^0 & = wt[x_{\text{resp}}^s] \oplus RS_{PUF0} \oplus RS_{PUF2} \\
& \quad \oplus RS_{PUF4} \oplus RS_{PUF6} \\
\text{dist}^1 & = wt[x_{\text{resp}}^s] \oplus RS_{PUF1} \oplus RS_{PUF2} \\
& \quad \oplus RS_{PUF4} \oplus RS_{PUF6} \\
\vdots & \\
\text{dist}^{15} & = wt[x_{\text{resp}}^s] \oplus RS_{PUF1} \oplus RS_{PUF3} \\
& \quad \oplus RS_{PUF5} \oplus RS_{PUF7}
\end{align*}
\]
\[
\begin{align*}
s' & = \arg\min_i \text{dist}^i, \ s \in \{0, \ldots, 2^\alpha - 1\} \\
s'' & = \arg\min_i \text{dist}^i, \ s \in \{0, \ldots, 2^\alpha - 1\} \setminus \{s'\} \\
c' & = \text{dist}^{s''} - \text{dist}^{s'}
\end{align*}
\]

E. Concatenating Symbol Encoding/Decoding with ECC

MAP symbol-based provisioning (top) and regeneration (bottom).

MAP symbol decoder would maximize the probability of a correct symbol given a helper data (XOR’ed) block:
\[
s' = \arg\max \Pr[x_{\text{resp}}^s|x_{\text{resp}}] \tag{1}
\]
where \(s \in \{0, 1, \ldots, 2^\alpha - 1\}\). It is not clear what a direct implementation of this looks like and how complex it would be.

Assuming Uniform Symbol Distribution. We now derive an optimal symbol decoder assuming that symbols are uniformly distributed (\(s\) is random if \(KM\) is chosen at random). We can express the MAP decoder equation above in terms of Bayes’ rule and derive a ML [20] decoder with a relatively simple implementation. Applying Bayes’ rule and accounting for uniform distribution and that the helper data appears with a probability of 1, for \(s \in \{0, 1, \ldots, 2^\alpha - 1\}\):
\[
\Pr[x_{\text{resp}}^s] = \Pr[x_{\text{resp}}^s|x_{\text{resp}}] \cdot \Pr[x_{\text{resp}}] / \Pr[x_{\text{resp}}^s] = \Pr[x_{\text{resp}}^s] \cdot 2^{-\alpha} / 1 \tag{2}
\]

So instead of maximizing \(\Pr[x_{\text{resp}}^s|x_{\text{resp}}]\) using MAP, we can alternatively maximize \(\Pr[x_{\text{resp}}^s]\) using ML. However, during regeneration, the only way the decoder can represent \(s\) is by looking at \((x_{\text{resp}})^s\) (i.e., we need the encoder function during the decoding process). So using ML, we maximize:
\[
\Pr[x_{\text{resp}}^s | s] = \Pr[x_{\text{resp}}^s | (x_{\text{resp}})^s] \tag{3}
\]

This is equivalent to minimizing Hamming distance between helper data block and all possible regenerated choices:
\[
s' = \arg\min_i \text{wt}[x_{\text{resp}}^s] \oplus (x_{\text{resp}})^s], \tag{4}
\]
\(s \in \{0, 1, \ldots, 2^\alpha - 1\}\). We have thus shown that our ML decoder (Eqn. (4)) is in fact an optimal decoder if we assume that the symbols \(s\) have uniform distribution, and in fact Eqn. (4) has a simple implementation as was shown in Sections III.C and III.D.

If symbols are not uniform, optimality of ML decoder gets impacted; security is compensated by generating more symbols \(s\) before a down-mixing hash.
B. Simulation: Analytical and Monte Carlo Results

Deriving MLSR error probability. The output bit error probability of a scheme is an important performance measure to evaluate it. We analyze MLSR with a sequence (helper data block) length \( l \) and alphabet size \( z = 2^\nu \). Let (Hamming) distance between two random reference sequences be \( d \). The regenerated sequence is interpreted as one of the reference sequences plus noise.

The correct symbol is detected, if less than \( \left\lfloor \frac{d-1}{2} \right\rfloor \) errors occur in the \( d \) positions where two sequences \( a \) and \( b \) differ. For a distance of \( d_a \) between the regenerated sequence and sequence \( a \), and \( d_b \) between the regenerated sequence and sequence \( b \), an error in the sequence positions with identical values would increase both distances by 1 such that \( d = d_a - d_b \) still remains constant and the decision is not affected.

Therefore, for an input PUF bit error probability \( p \) and a distance \( d \), a detection error \( p_e \) between symbols \( a \) and \( b \) occurs with

\[
p_e(d) = 1 - \sum_{i=0}^{\left\lfloor \frac{d-1}{2} \right\rfloor} \binom{d}{i} p^i(1-p)^{d-i} \tag{5}
\]

Depending on \( l \), distances \( d_i \) occur with a probability

\[
Pr[d = d_i] = \binom{l}{d_i} \cdot 2^{-l} \tag{6}
\]

Summing over all possible distances and their error probabilities gives an average overall error probability for a detection error between any two unknown symbols of

\[
p_{pair} = 2^{-l} \sum_{d=0}^{l} \binom{l}{d} p_e(d) \tag{7}
\]

For \( z \) symbols, the correct symbol could be decoded to any other of the \( z-1 \) symbols, which gives an overall symbol error probability \( p_S \) of

\[
p_S = (z-1) \cdot 2^{-\nu} \sum_{d=0}^{l} \binom{l}{d} p_e(d) \tag{8}
\]

**ML vs. Bounded-Distance/Threshold-Based.** We demonstrate the efficiency of ML vs. other ECC methods with an equivalent block size (which is a first order estimate of helper data size) and a fixed PUF noise level. Figure 6 shows three scenarios, namely embedding 7 information bits into a block size of \( \sim 64 \) (red), 8 bits into a block size of \( \sim 128 \) (blue) and 9 bits into a block size of \( \sim 256 \) (black).

For each scenario, we compare MLSR\( (\nu = l, \kappa = \alpha) \) (solid line) to a Rep\( (\nu, \kappa = 1, \tau) \) code (dotted line) and a BCH\( (\nu, \kappa, \tau) \) code (circles), where \( \nu \) is the block size, \( \kappa \) is the number of information bits per block, \( \tau \) is the number of correctable errors. For MLSR, key bit error probabilities are derived from symbol error probabilities according to Eqn. 8, and assuming that for an \( \alpha \)-bit symbol, in average half the bits in a symbol are flipped if symbol recovery fails. The solid dots represent simulation results that demonstrate that the analytic approach is accurate. To achieve similar rates for the Repetition code for a fair comparison, we divide a block into one sub-block for each information bit. Note that bounded-distance decoded BCH code with maximum error correction (which works \( \sim 25\% \) bit flips in this setting) is equivalent to a threshold-based pattern matching approach [19] with comparison threshold set to 0.25.

Comparing curves of same color shows significant performance increase through using ML. For example, for MLSR\( (128,8) \) (solid blue curve), at a PUF noise factor of \( p = 0.15 \) (x-axis), MLSR has a bit error probability of \( 1 \times 10^{-8} \) (y-axis) prior to any soft-decision decoding to mop up residual errors, while both repetition and BCH (set to maximum correction of 25%) have error probabilities at least 10,000\% higher.

C. Empirical: Reliable Key Generation at 28nm

**Design Under Test.** We implemented our design in 28nm Xilinx Artix-7 FPGA silicon, using \( k=64 \)-sum PUFs [24] in a \( n=4\)-XOR construction (in an ASIC implementation this can be replaced by 64-stage Arbiter PUFs). Each symbol is configured as \( \alpha = 8 \) bits. To derive a 128-bit key, we treat 22 symbols as keying material, producing a 176-bit value that is down-mixable to a 128-bit key (we assume a 1.3 pre-hash expansion as was used in [1]). Our design contained ability to perform single-parity-symbol soft-decision decoding, where we appended a single 8-bit parity column to the 22 8-bit columns (symbols); this is equivalent to \( \beta = 23 \) symbols stored as helper data for a 128-bit key. Each symbol is mapped to a \( l=128\)-bit 4-XOR’ed response sequence, taking up \( l \cdot \beta = 2944 \) bits in the helper data. In terms of raw 4-XOR PUF noise, compared to a reference at 25°C ambient, error rates are 22% at -65°C ambient, 24% at 105°C ambient / 125°C junction, and 15% at 25°C ambient.
TABLE I
ENVIRONMENTAL STRESS TESTING, 28nm FPGA SILICON

<table>
<thead>
<tr>
<th>Prov</th>
<th>Regen</th>
<th>Key Regen Cycles</th>
<th>After MLSR</th>
<th>After SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>25°C</td>
<td>-65°C</td>
<td>3.4 x 10^7</td>
<td>11 Failures (3.24 x 10^-3)</td>
<td>0 Failures (&lt; 2.94ppm)</td>
</tr>
<tr>
<td>25°C</td>
<td>105°C</td>
<td>1.2 x 10^9</td>
<td>3562 Failures (2.95 x 10^-3)</td>
<td>0 Failures (&lt; 0.83ppm)</td>
</tr>
<tr>
<td>1.00V</td>
<td>1.05V</td>
<td>1.7 x 10^6</td>
<td>0 Failures</td>
<td>0 Failures (&lt; 0.59ppm)</td>
</tr>
<tr>
<td>1.00V</td>
<td>0.95V</td>
<td>8.0 x 10^7</td>
<td>0 Failures</td>
<td>0 Failures (&lt; 1.25ppm)</td>
</tr>
</tbody>
</table>

Key Regeneration Under Environmental Stress. Table 1 shows results of key generation stress testing from one device. Another two devices underwent similar tests (covering same extreme conditions) but ran for a smaller number of key regeneration cycles to confirm these results. Results show a 128-bit key regeneration failure rate \( \leq 1 \times 10^{-6} \) in that more than a million key regenerations were run but no failures were seen after soft-decision decode (SD) for the most stressful condition of 105°C ambient / 125°C junction.

V. LINEAR-XOR AND LINEAR-SHIFT ATTACKS

Recently, there have been several papers attacking PUF key generators by manipulating helper data.


2) Pattern Matching: Shift Malleability: [5] crafted an attack noting that a left/right shift in pattern matching helper data produces an “adjacent” helper data in a predictable fashion. The attacks in [13] [5] manipulated helper data based on a codeword structure or shift pattern that is common to all devices. We have resistance to such attacks since our symbol-to-symbol helper data relationships are determined by manufacturing variation unique to each device vs. relying solely on algorithmic attributes common to all devices. It is not obvious how to manipulate a helper data block to cause it to converge predictably to a different symbol on a particular device.

VI. CONCLUSION

We presented a ML symbol coding method that can outperform, with orders of magnitude lower bit error probabilities, prior PUF key generation approaches such as BCH, repetition coding, and pattern matching. We further concatenated symbol recovery with soft-decision ECC to eliminate residual errors.

Future work include reconciling our mathematical, simulation, and empirical silicon results with theoretical symbol-based extensions [3] to the fuzzy extractor that are under development.

ACKNOWLEDGMENT

This work was partly funded by the Bavaria California Technology Center (BaCaTeC), grant number 2014-1/9.

REFERENCES