Optimized quantization in Zero Leakage Helper data systems

Optimized Quantization in Zero Leakage Helper Data Systems

Taras Stanko, Fitria Nur Andini, and Boris Škorić

***Abstract*— Helper data systems are a cryptographic primitive that allows for the reproducible extraction of secrets from noisy measurements. Redundancy data called helper data makes it possible to do error correction while leaking little or nothing (Zero Leakage) about the extracted secret string. We study the case of non-discrete measurement outcomes. In this case, a quantization step is required. Recently, de Groot *et al.* described a generic method to perform the quantization in a Zero Leakage manner. We extend their work and show how the quantization intervals should be set to maximize the amount of extracted secret key material when noise is taken into account.**

***Index Terms*— Helper data, fuzzy extractor, secure sketch, PUF, biometrics.**

I. Introduction

A. Helper Data Systems

SECURITY with noisy data is the art of reproducibly extracting secret data from noisy measurements on a physical system. The two main applications are storage of cryptographic keys using Physical Unclonable Functions (PUFs) [2], [10], [15], [16], [18], [19] and privacy-preserving storage of biometric data. Power-off storage of keys in *digital* memory can often be considered insecure. PUFs provide an alternative way to store keys, which allows the designer to exploit the inscrutability of *analog* physical behavior. Keys stored in this way are referred to as Physically Obfuscated Keys (POKs) [9].

In both the biometrics and the PUF/POK case, noise has to be eliminated, but under the constraint that the redundancy data (which is visible to attackers) does not endanger the secret. This problem was addressed by the introduction of a special security primitive, the *Helper Data System* (HDS) [14]. A HDS in its most general form is shown in Fig. 1. The `Gen` procedure takes as input a measurement X. `Gen` outputs a secret S and public Helper Data W. The helper data is stored. In the reproduction phase, a fresh measurement Y is obtained. Typically Y is close to X but not identical. The `Rec` procedure takes Y and W as input. It outputs $\hat{S}$, an estimate of S. If Y is sufficiently close to X then $\hat{S} = S$. (See Def. 2 for a more precise formulation.)

Manuscript received July 13, 2016; revised December 15, 2016; accepted January 23, 2017. Date of publication April 25, 2017; date of current version May 10, 2017. This work was supported by The Netherlands Organisation for Scientific Research NWO (Cyber Security Project ESPRESSO) under Grant 628.001.019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Negar Kiyavash. *(Corresponding author: Boris Škorić.)*

The authors are with the Eindhoven University of Technology, 5600MB Eindhoven, The Netherlands (e-mail: b.skoric@tue.nl).

Digital Object Identifier 10.1109/TIFS.2017.2697840

Two special HDS cases are the Secure Sketch (SS) and the Fuzzy Extractor (FE) [8]. The SS has $S = X$ (and $\hat{S} = \hat{X}$, an estimator for X). If X is not uniformly distributed, then S is not uniform. The SS is suitable for privacy-preserving biometrics, where the stored biometric enrollment data is a cryptographic hash of X, just like hashed storage of passwords; high entropy of S (given W) is required, but not uniformity. The FE is required to have a (nearly) uniform S given W. The FE is typically used for extracting keys from PUFs and POKs. Note that there is a generic construction to obtain a FE from a SS: privacy amplification on X by applying a suitable information-theoretic hash function. This can be either a Universal Hash Function (UHF) [3], [13], [17] or, more sophisticatedly, a q-wise independent hash [7].

We consider the general HDS case (occasionally denoted as 'gHDS' in this paper). The general HDS is of particular interest when X is a continuum variable: (i) The least significant digits of X are not interesting for key extraction and (ii) In view of the excellent performance of q-wise independent hashes [7] it is best to first extract from X a non-uniform high-entropy discrete secret and then compress it.

B. Zero Leakage Quantisation

In biometrics and in several PUF/POK scenarios the X is analog. A typical HDS then consists of two stages: (1) a HDS that discretizes X; (2) a HDS acting on a discrete source, e.g. the Code Offset Method [1], [6], [8], [12], [21]. Both stages use helper data. In the first stage it is possible to make a construction such that W leaks nothing about S. Intuitively, W contains the 'least significant bits' of X, which are noisy, while S contains the 'most significant bits'. A HDS that achieves independence of S and W is called a Zero Leakage HDS (ZLHDS). Verbitskiy *et al.* [20] introduced a Zero Leakage Fuzzy Extractor (ZLFE) for $X \in \mathbb{R}$.[1,2] They divided the space $\mathbb{R}$ into N equiprobable intervals $\mathcal{A}_0, \ldots, \mathcal{A}_{N-1}$. At enrollment, if X lies in interval $\mathcal{A}_j$ then S is set to j. For the helper data they introduced a further division of each interval $\mathcal{A}_j$ into m equiprobable subintervals $(\mathcal{A}_{jk})_{k=0}^{m-1}$. If the enrollment measurement X lies in interval $\mathcal{A}_{jk}$ then the index

[1] A high-dimensional measurement is usually split into one-dimensional components, e.g. using Principal Component Analysis or similar methods. A HDS is then applied to each component individually. The results are combined and then serve as input for the 2nd stage.

[2] In the literature the FE and SS terminology is not always used in exactly the same way. The definition in [6] considers a *class* of source distributions, whereas [20] and many other works target a specific variable X with a fixed distribution. We adopt the latter approach.

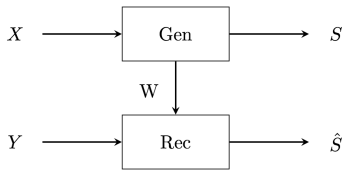

Fig. 1. Data flow in a generic helper data system.

k is stored as helper data. The fact that all these subintervals are equiprobable leads to independence between the helper data and the secret. De Groot *et al.* [5] took the limit $m \to \infty$ and showed that the resulting scheme is not just *a* ZLFE but *the* generic best performing ZLFE for $X \in \mathbb{R}$; other ZLFEs for $X \in \mathbb{R}$ can be derived from the generic scheme. Furthermore, de Groot *et al.* generalized the scheme of [20] from ZLFEs to general ZLHDSs by allowing intervals $\mathcal{A}_0, \ldots, \mathcal{A}_{N-1}$ that are not equiprobable. Several questions were left open regarding the `Rec` procedure in general ZLHDSs and the performance of ZLHDSs compared to ZLFEs.

C. Contributions and Outline

Our contributions relate only to the first stage of ZLHDSs. We investigate the case $X \in \mathbb{R}$.

- We derive an optimal `Rec` procedure that minimises the probability of errors. We obtain analytic formulas for Gaussian noise and for Lorentz-distributed noise, which is also sometimes encountered [11].
- Using this `Rec` procedure we study the performance of ZLHDSs compared to ZLFEs. We define performance as the mutual information between S and $\hat{S}$ conditioned on W. This mutual information $I(S; \hat{S}|W)$ represents the maximum amount of secret key material that can be extracted from X using a ZLHDS. The intricacies of the `Rec` procedure cause the mutual information to become a complicated function of the choice of quantisation $\mathcal{A}_0, \ldots, \mathcal{A}_{N-1}$. We resort to numerics. Our numerical results for Gaussian source and Gaussian noise show that optimisation of the quantisation intervals yields an improvement with respect to the ZLFE in terms of mutual information as well as reconstruction error probability. In most cases the gain in $I(S; \hat{S}|W)$ is modest, but the reduction of the error rate can be substantial. *We conclude that it is better to use a ZLHDS than a ZLFE.*

In Section II we introduce notation and summarize the results of [5]. In Section III we derive the optimal `Rec` procedure and provide analytic expressions (as far as possible) for the mutual information and the error rate. Section IV presents the numerical results for Gaussian source and Gaussian noise.

II. Prior Work

A. Notation and Terminology

We use capitals to represent random variables, and lowercase for their realizations. The input and output variables of the HDS are as depicted in Fig. 1. Sets are denoted by calligraphic font. The set $\mathcal{S}$ is defined as $\mathcal{S} = \{0, \ldots, N-1\}$. For $\alpha \in \mathcal{S}$ we define $p_\alpha = \Pr[X \in \mathcal{A}_\alpha]$. The expected value with respect to a random variable Z is denoted as $\mathbb{E}_z$. The mutual information (see e.g. [4]) between X and Y is $I(X; Y)$, and the mutual information conditioned on Z is $I(X; Y|Z)$. The probability density function (pdf) of the random variable $X \in \mathbb{R}$ in written as $f(x)$ and its cumulative distribution function (cdf) as $F(x)$.

B. Zero Leakage Definition

De Groot *et al.* used the following definition of the ZL property.

Definition 1 (Zero Leakage): Let $W \in \mathcal{W}$. We call a HDS a Zero Leakage HDS if and only if

$$\forall_{\mathcal{V} \subseteq \mathcal{W}} \ \Pr[S = s | W \in \mathcal{V}] = \Pr[S = s]. \tag{1}$$

The property (1) implies $I(W; S) = 0$. The converse is not the case, since for continuous $\mathcal{W}$ one may allow isolated points $w \in \mathcal{W}$ where (1) does not hold.

C. Noise Model

We adopt the noise model from [5]. The X and Y are noisy versions of an underlying 'true' value. Without loss of generality X is taken to have zero mean. The standard deviations of $X, Y \in \mathbb{R}$ are denoted as σ_X and σ_Y respectively. The verification sample Y is related to the enrollment measurement as $Y = \lambda X + R$, where $\lambda \in [0, 1]$ is the *attenuation parameter* and R is zero-mean additive noise, independent of X. We have $\sigma_Y^2 = \lambda^2 \sigma_X^2 + \sigma_R^2$. The correlation between X and Y is

$$\rho \stackrel{\text{def}}{=} \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y} = \lambda \frac{\sigma_X}{\sigma_Y}, \tag{2}$$

with $\rho \in [-1, 1]$. Furthermore $\lambda^2 = \frac{\rho^2}{1-\rho^2} \frac{\sigma_R^2}{\sigma_X^2}$. Two special cases are often considered:

1) Perfect Enrollment: During enrollment there is no noise. The X equals the 'true' value. In this situation it holds that $\sigma_Y^2 = \sigma_X^2 + \sigma_R^2$ and $\lambda = 1$.

2) Identical Conditions: The amount of noise is the same during enrollment and reconstruction. In this situation $\sigma_Y^2 = \sigma_X^2$ and $\lambda^2 = \rho^2 = 1 - \sigma_R^2/\sigma_X^2$. The pdf of Y given $X = x$ is denoted as $\psi(y|x) = \upsilon(y - \lambda x)$. The noise is symmetric and fading, i.e. $\upsilon(-z) = \upsilon(z)$ and $\upsilon(z)$ is a decreasing function of $|z|$. The cdf corresponding to υ is denoted as V.

D. The ZL Scheme of [5]

The helper data is considered to be continuous, $W \in \mathcal{W} \subset \mathbb{R}$, and without loss of generality de Groot *et al.* set $\mathcal{W} = [0, 1)$. The left boundary of the quantisation region $\mathcal{A}_\alpha$ is denoted as Ω_α, $\alpha \in \mathcal{S}$. (See Fig. 2.) It holds that

$$\Omega_\alpha = F^{\text{inv}}\left(\sum_{i=0}^{\alpha-1} p_i\right), \tag{3}$$

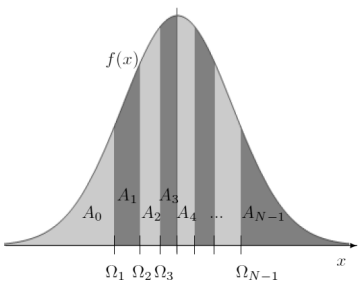

Fig. 2. Illustration of the quantization boundaries Ω_α and regions $\mathcal{A}_\alpha$.

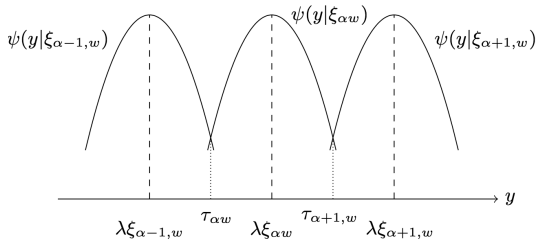

Fig. 3. Decision boundaries. (Reconstruction phase.)

where F^{inv} stands for the inverse function of F. Note that $\Omega_0 = -\infty$. The Gen procedure is written as $s = Q(x)$, $w = g(x)$, where the Q and g functions are given by

$$Q(x) = \max\{\alpha \in \mathcal{S} :\ x \geq \Omega_\alpha\}$$
$$g(x) = \frac{F(x) - F(\Omega_{Q(x)})}{p_{Q(x)}} = \frac{F(x) - \sum_{i=0}^{Q(x)-1} p_i}{p_{Q(x)}}. \quad (4)$$

The relation between x, s and w can be written in a more friendly form as

$$F(x) = F(\Omega_s) + wp_s = \textstyle\sum_{i=0}^{s-1} p_i + wp_s. \quad (5)$$

The thus defined $w \in [0, 1)$ is called *quantile* helper data since it measures which quantile of the probability mass p_s is located between $F(\Omega_s)$ and x. It was shown that the random variable W, given S, has a uniform pdf. Consequently the scheme is a ZLHDS. The mapping of x to (s, w) is a bijection. For the mapping of (s, w) to x the following notation is used,[3]

$$\xi_{s,w} \stackrel{\text{def}}{=} F^{\text{inv}}(\textstyle\sum_{i=0}^{s-1} p_i + wp_s). \quad (6)$$

In the FE case (uniform $p_\alpha = 1/N$ for all $\alpha \in \mathcal{S}$) the optimal reconstruction procedure was found to be the following maximum-likelihood 'decoder',

$$\hat{s} = \texttt{Rec}^{\text{FE}}(y, w) = \arg\max_{\alpha\in\mathcal{S}} \psi(y|\xi_{\alpha w}). \quad (7)$$

Eq. (7) can be conveniently implemented by defining decision boundaries $(\tau_{\alpha w})_{\alpha=0}^{N}$. If $y \in [\tau_{\alpha w}, \tau_{\alpha+1,w})$, then $\hat{s} = \alpha$. In the case of symmetric fading noise the location of the decision boundaries dictated by (7) was found to be

$$\tau_{\alpha w}^{\text{FE}} = \lambda(\xi_{\alpha-1,w} + \xi_{\alpha w})/2. \quad (8)$$

Here $\xi_{-1,w} = -\infty$ and $\xi_{Nw} = \infty$, resulting in $\tau_{0w} = -\infty$, $\tau_{Nw} = \infty$. Fig. 3 shows how to understand (8). Each pdf $\psi(y|\xi_{\alpha w})$ in (7) is centered around $y = \lambda\xi_{\alpha w}$ and drops off symmetrically. The point where one α-value becomes more likely than another lies halfway between the centers of two neighbouring pdfs; the crossing point is a decision boundary.

[3] We often omit the comma and write ξ_{sw} instead of $\xi_{s,w}$.

III. General ZLHDS Optimization

In this section we extend the results of de Groot *et al.* [5]. We generalize equations (7) and (8). Then we derive analytic expressions for $I(S; \hat{S}|W)$ and the reconstruction error probability P_{err} in terms of the scheme's parameters. We also discuss the relation between P_{err} and the bit error rate.

A. Formal Definition of a HDS

In Section I-A we discussed HDSs without providing a definition. For completeness we give a definition in the spirit of [14] using the terminology of [8]. The definition makes use of a distance measure 'dis' on the source space $\mathcal{X}$ and of the statistical distance 'SD' (total variation distance). Note that $\mathcal{X}$ can be a continuum, and that a HDS is defined for a specific distribution on $\mathcal{X}$.

Definition 2: A $(X, \mathcal{S}, t, \varepsilon)$-HDS for a variable $X \in \mathcal{X}$ is a pair of functions $\texttt{Gen} : \mathcal{X} \to \mathcal{S} \times \{0, 1\}^*$ and $\texttt{Rec} : \mathcal{X} \times \{0, 1\}^* \to \mathcal{S}$, with the following properties.

1) **Correctness**. Let $(S, W) = \texttt{Gen}(X)$. If $\text{dis}(X, Y) \leq t$ then $\texttt{Rec}(Y, W) = S$.
2) **Security**. Let $(S, W) = \texttt{Gen}(X)$. Let f_{SW} denote the probability distribution of (S, W). Let f_S and f_W be the marginal distributions. It holds that $\text{SD}(f_{SW}, f_S f_W) \leq \varepsilon$.

Optionally, the Gen is randomized using public randomness.

B. ZLHDS Reconstruction

For the sake of completeness we explicitly show that W given $S = s$ is uniform. (This fact was implicit in [5] but it was not separately stated.)

Lemma 1: The probability density function of the helper data W given the secret S is uniform.

Proof: For the pdf of W given $S = \alpha$ we write $\rho(w|\alpha)$. We start from $p_\alpha \rho(w|\alpha)\text{d}w = f(\xi_{\alpha w})\text{d}\xi_{\alpha w}$. (The validity of this equation is readily verified. Applying $\int_0^1$ to the left hand side yields p_α by definition; on the right hand side the equivalent operation is integration over $\xi_{\alpha w}$ on the interval $\mathcal{A}_\alpha$, which also yields p_α.) Now we can write $\rho(w|\alpha) = \frac{f(\xi_{\alpha w})}{p_\alpha \text{d}w/\text{d}\xi_{\alpha w}} = \frac{f(\xi_{\alpha w})}{\text{d}F(\xi_{\alpha w})/\text{d}\xi_{\alpha w}} = \frac{f(\xi_{\alpha w})}{f(\xi_{\alpha w})} = 1$. In the second equality we used (5) with $s = \alpha$ kept constant while w varies. ■

Lemma 2: For the general HDS the optimal reconstruction procedure is given by

$$\hat{s} = \texttt{Rec}(y, w) = \arg\max_{\alpha\in\mathcal{S}} p_\alpha \psi(y|\xi_{\alpha w}). \quad (9)$$

Proof: This is a slight modification of [5, Lemma 5.1], with the same starting point.

$$\texttt{Rec}(y, w) = \arg\max_{\alpha\in\mathcal{S}} \Pr[S = \alpha | Y = y, W = w]$$
$$= \arg\max_{\alpha\in\mathcal{S}} \frac{\Pr[S = \alpha, Y = y, W = w]}{\Pr[Y = y, W = w]}. \quad (10)$$

The denominator does not depend on α. It can be eliminated.

$$\begin{aligned}\texttt{Rec}(y,w) &= \arg\max_{\alpha\in\mathcal{S}} \Pr[S=\alpha, Y=y, W=w] \\ &= \arg\max_{\alpha\in\mathcal{S}} \Pr[Y=y|S=\alpha, W=w]\rho(w|\alpha)p_\alpha.\end{aligned}$$

Using Lemma 1 we get

$$\hat{s} = \texttt{Rec}(y,w) = \arg\max_{\alpha\in\mathcal{S}} p_\alpha \Pr[Y=y|S=\alpha, W=w]. \tag{11}$$

Since (α, w) uniquely defines $\xi_{\alpha w}$, the probability $\Pr[Y = y|S = \alpha, W = w]$ equals $\Pr[Y = y|X = \xi_{\alpha w}]$, for which the notation $\psi(y|\xi_{\alpha w})$ is used. ∎

From (9) we can derive an optimal placement of the boundaries $\tau_{\alpha w}$ for general noise and general HDS.

Lemma 3: For a ZLHDS the reconstruction boundary $\tau_{\alpha w}$ obtained using pdf intersections satisfies the following equation:

$$p_{\alpha-1}\psi(\tau_{\alpha w}|\xi_{\alpha-1,w}) = p_\alpha \psi(\tau_{\alpha w}|\xi_{\alpha w}). \tag{12}$$

Proof: From Lemma 2 we see that the decision boundary is the point y where the function $p_\alpha\psi(y|\xi_{\alpha w})$ intersects the function $p_{\alpha-1}\psi(y|\xi_{\alpha-1,w})$. ∎

In the FE case, $p_{\alpha-1} = p_\alpha$ and (12) reduces to $\psi(\tau_{\alpha w}|\xi_{\alpha-1,w}) = \psi(\tau_{\alpha w}|\xi_{\alpha w})$, which directly yields (8). In the general HDS case, however, the difference between the p_α parameters changes the heights of the pdfs $\psi(y|\cdots)$ in Fig. 3, which leads to a more complicated solution.

Theorem 1: Let the noise be zero-mean Gaussian with variance σ_R^2. Then the intersection points as specified in (12) are

$$\tau_{\alpha w} = \lambda\frac{\xi_{\alpha-1,w}+\xi_{\alpha w}}{2} + \frac{\sigma_R^2 \ln\frac{p_{\alpha-1}}{p_\alpha}}{\lambda(\xi_{\alpha w}-\xi_{\alpha-1,w})}. \tag{13}$$

Proof: The Gaussian noise is given by $\psi(y|x) = \frac{1}{\sqrt{2\pi}\sigma_R}\exp[-\frac{(y-\lambda x)^2}{2\sigma_R^2}]$. Eq. (12) then becomes

$$\frac{p_{\alpha-1}}{\sqrt{2\pi}\sigma_R} e^{-\frac{(\tau_{\alpha w}-\lambda\xi_{\alpha-1,w})^2}{2\sigma_R^2}} = \frac{p_\alpha}{\sqrt{2\pi}\sigma_R} e^{-\frac{(\tau_{\alpha w}-\lambda\xi_{\alpha w})^2}{2\sigma_R^2}}. \tag{14}$$

Taking the logarithm on both sides of the equation yields a *linear* equation in $\tau_{\alpha w}$, with solution (13). ∎

Theorem 2: Let the noise be Lorentz-distributed, $\psi(y|x) = \frac{1/\sigma_R}{1+\pi^2(y-\lambda x)^2/\sigma_R^2}$. Let $p_\alpha \neq p_{\alpha-1}$. If

$$p_\alpha p_{\alpha-1}(\lambda\xi_{\alpha,w}-\lambda\xi_{\alpha-1,w})^2 \geq \sigma_R^2\frac{(p_\alpha-p_{\alpha-1})^2}{\pi^2}, \tag{15}$$

then the reconstruction boundary $\tau_{\alpha w}$ is given by

$$\tau_{\alpha w} = \frac{p_{\alpha-1}\lambda\xi_{\alpha w} - p_\alpha\lambda\xi_{\alpha-1,w}}{p_{\alpha-1}-p_\alpha} - \frac{1}{p_{\alpha-1}-p_\alpha}\cdot \times\sqrt{p_\alpha p_{\alpha-1}(\lambda\xi_{\alpha w}-\lambda\xi_{\alpha-1,w})^2 - \frac{\sigma_R^2}{\pi^2}(p_{\alpha-1}-p_\alpha)^2}. \tag{16}$$

Proof: Substitution of the distribution into (12) yields

$$\frac{p_\alpha}{1+\pi^2\sigma_R^{-2}(\tau_{\alpha w}-\lambda\xi_{\alpha w})^2} = \frac{p_{\alpha-1}}{1+\pi^2\sigma_R^{-2}(\tau_{\alpha w}-\lambda\xi_{\alpha-1,w})^2}. \tag{17}$$

Inverting both sides gives a quadratic equation in $\tau_{\alpha w}$. (If $p_\alpha = p_{\alpha-1}$ then it reduces to a linear equation with (8) as the solution.) The quadratic equation has solutions if the discriminant is nonnegative, which is equivalent to condition (15). We have to choose the correct sign preceding the square root of the determinant. We choose the sign such that $\lambda\xi_{\alpha-1,w} < \tau_{\alpha w} < \lambda\xi_{\alpha w}$. We verify that (16) indeed satisfies these inequalities. On the one hand, (16) can be written as

$$\tau_{\alpha w} = \lambda\xi_{\alpha w} + \frac{p_\alpha\lambda(\xi_{\alpha w}-\xi_{\alpha-1,w}) - \sqrt{\cdots}}{p_{\alpha-1}-p_\alpha}. \tag{18}$$

Note that $\xi_{\alpha w}-\xi_{\alpha-1,w} > 0$. If $p_{\alpha-1} > p_\alpha$ then the $\sqrt{\cdots}$ 'wins' and the numerator of the fraction is negative, as it should be. If $p_{\alpha-1} < p_\alpha$ then the denominator is negative and the $\sqrt{\cdots}$ 'loses', making the numerator positive. On the other hand, (16) can also be written as

$$\tau_{\alpha w} = \lambda\xi_{\alpha-1,w} + \frac{p_{\alpha-1}\lambda(\xi_{\alpha w}-\xi_{\alpha-1,w}) - \sqrt{\cdots}}{p_{\alpha-1}-p_\alpha}. \tag{19}$$

If $p_{\alpha-1} > p_\alpha$ then the $\sqrt{\cdots}$ 'loses' and the fraction is positive. If $p_{\alpha-1} < p_\alpha$ then the $\sqrt{\cdots}$ 'wins' and the fraction is positive. ∎

Remark: If one adopts (13) as decision boundaries, an incorrect reconstruction procedure may result under some pathological circumstances. This can happen, for example, if for some α it happens that $p_\alpha \ll p_{\alpha-1}$ and $p_\alpha \ll p_{\alpha+1}$; then in Fig. 3 the middle curve is located beneath the intersection of its neighbours, and $\hat{s}$ cannot equal α even if $s = \alpha$. In practice we will never see this pathological case.

C. Optimization of the Quantization Intervals

As announced in Section I-C, we want to maximize the amount of key material extracted from X by the ZLHDS. We have to take into account two effects: the noise, which limits how much of the entropy of X can be recovered in the reconstruction phase, and the fact that the adversary knows W. The quantity of interest is the mutual information between S and $\hat{S}$ given W: $I(S;\hat{S}|W)$. This represents the 'secrecy capacity' or quality of the channel from S to $\hat{S}$ created by the ZLHDS. If a perfect error correction mechanism is used as the second-stage HDS, i.e. one that achieves the Shannon bound, then $I(S;\hat{S}|W)$ is the achievable key length.

We note that even though $\mathsf{H}(S|W) = \mathsf{H}(S)$, we have $I(S;\hat{S}|W) \neq I(S;\hat{S})$ because $\hat{S}$ is not independent of W.

Lemma 4: For a zero leakage helper data system the mutual information can be expressed as

$$I(S;\hat{S}|W) = \mathsf{H}(S) - \mathsf{H}(S|\hat{S},W) = I(S;\hat{S},W). \tag{20}$$

Proof: We write $I(S;\hat{S}|W) = \mathsf{H}(S|W) - \mathsf{H}(S|\hat{S},W)$. Due to the ZL property it holds that $\mathsf{H}(S|W) = \mathsf{H}(S)$. ∎

The mutual information $I(S;\hat{S}|W)$ can be seen as a function of the system parameters $p_0,\ldots,p_{N-1}$. These parameters completely fix the Gen and Rec procedures. (The λ, σ_X and σ_R are given by nature and cannot be chosen). Hence we want to determine how to set vector $(p_\alpha)_{\alpha\in\mathcal{S}}$ as a function of λ, σ_X, σ_R so as to maximize our target function. Unfortunately, $I(S;\hat{S}|W)$ depends on the p_α parameters in a very complicated way. The Gen is simple enough, but the Rec procedure

has decision boundaries $\tau_{\alpha w}$ (12) that depend on $p_0, \ldots, p_{N-1}$ not only directly but also via the $\xi_{\alpha w}$ points as specified in (6); this dependence is quite convoluted as the $\xi_{\alpha w}$ invoke the non-smooth stepwise function Q as well as the nonlinear F^{inv}. Analytic maximisation of $I(S;\hat{S}|W)$ is intractable. It is clear, however, that a maximum must exist. Consider the ZLFE at fixed $N \geq 3$. Not all intervals $\mathcal{A}_\alpha$ have equal width, which leads to unequal probabilities for jumping from one interval to another due to noise. Making the narrowest intervals slightly broader reduces the reconstruction error probability (with a positive effect on our target function) and the entropy of S (with a negative effect). It is intuitively clear that at large σ_R the effect of reconstruction errors weighs more heavily than the $\mathsf{H}(S)$ effect; then we expect a nontrivial maximum at a p_α setting different from the FE's $p_\alpha = 1/N$. The numerics in Section IV show that this is indeed the case.

For the efficiency of the numerical optimisation we now look for a simple form in which to represent $I(S;\hat{S}|W)$. We introduce the following notation,

$$\Upsilon_{\hat{s}|sw} \stackrel{\text{def}}{=} \Pr[\hat{S}=\hat{s}|S=s,W=w] = \int_{\tau_{\hat{s}w}}^{\tau_{\hat{s}+1,w}} \psi(y|\xi_{sw})\mathrm{d}y$$
$$= V(\tau_{\hat{s}+1,w} - \lambda\xi_{sw}) - V(\tau_{\hat{s}w} - \lambda\xi_{sw}). \quad (21)$$

The first equality follows from the reconstruction rule $y \in (\tau_{\hat{s}w}, \tau_{\hat{s}+1,w}) \implies \hat{S} = \hat{s}$. The last step follows from $\psi(y|x) = v(y-\lambda x)$ and V being defined as the integral of v (Section II-C). We can express the mutual information entirely in terms of the p_α and $\Upsilon_{\hat{s}|sw}$ parameters.

Lemma 5: For the ZLHDS it holds that

$$I(S;\hat{S}|W) = \sum_{s=0}^{N-1}\sum_{\hat{s}=0}^{N-1}\int_0^1 \mathrm{d}w\ p_s\Upsilon_{\hat{s}|sw}\log\frac{\Upsilon_{\hat{s}|sw}}{\sum_{\beta=0}^{N-1} p_\beta\Upsilon_{\hat{s}|\beta w}}. \quad (22)$$

Proof:

$$I(S;\hat{S}|W) = \mathbb{E}_{s\hat{s}w}\log\frac{\Pr[S=s,\hat{S}=\hat{s}|W=w]}{\Pr[S=s|W=w]\Pr[\hat{S}=\hat{s}|W=w]}$$
$$= \mathbb{E}_w\textstyle\sum_{s,\hat{s}=0}^{N-1}\Pr[S=s|W=w]\Upsilon_{\hat{s}|sw}\cdot$$
$$\times\log\frac{\Pr[S=s,\hat{S}=\hat{s}|W=w]}{\Pr[S=s|W=w]\Pr[\hat{S}=\hat{s}|W=w]}. \quad (23)$$

In the last line we used the chain rule $\Pr[S=s,\hat{S}=\hat{s},W=w] = \mathbb{E}_w\Pr[S=s|W=w]\Upsilon_{\hat{s}|sw}$. Next we use $\mathbb{E}_w(\cdots) = \int_0^1 \mathrm{d}w(\cdots)$ as implied by Lemma 1, and $\Pr[S=s|W=w] = p_s$ by the ZL property. We apply these rules, and $\Pr[\hat{S}=\hat{s}|W=w] = \sum_s p_s\Upsilon_{\hat{s}|sw}$, inside the logarithm. ∎

D. Reconstruction Errors

While we are mainly interested in the mutual information, we also care about the practical implementation aspects of the second-stage HDS. The second-stage HDS typically employs an Error-Correcting Code (ECC). If the output of the first-stage HDS has a high bit error rate, this causes problems for the ECC. In our numerics we keep track of the error rate.

TABLE I

THREE-BIT GRAY CODE USED FOR $N=5$ AND $N=6$. THE HIGHLIGHTED CELL SHOWS THE TWO-BIT GRAY CODE FOR $N=3$ AND $N=4$

s	1st bit	2nd bit	3rd bit
0	0	0	0
1	0	0	1
2	0	1	1
3	0	1	0
4	1	1	0
5	1	1	1

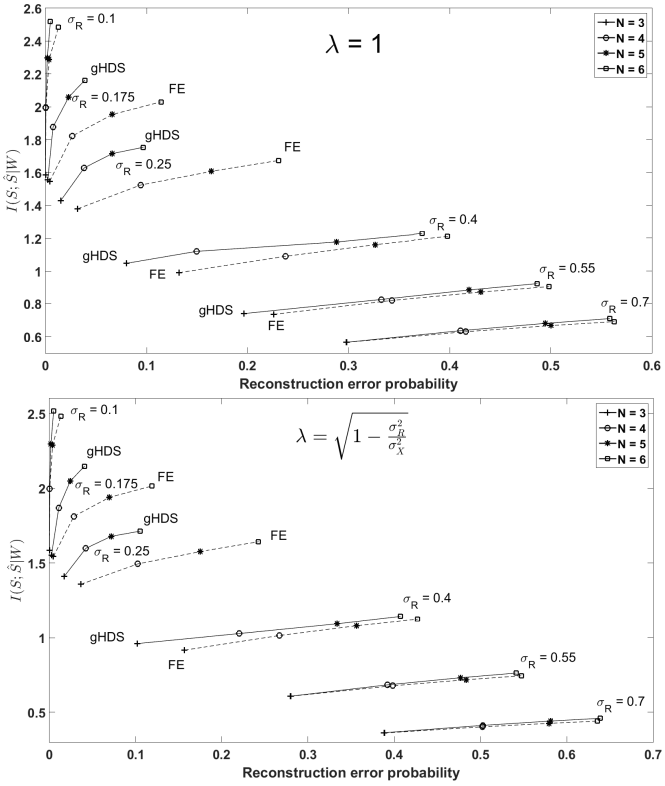

Fig. 4. Mutual information versus P_{err} for perfect enrollment (upper figure) and identical conditions (lower figure). At fixed σ_R, data points for the general HDS are connected with a solid line, while a dashed line corresponds to the FE.

We write $P_{\text{err}} = \Pr[\hat{S} \neq Q(X)]$ for the overall probability that $\hat{S}$ is not equal to S. This is an averaged quantity, i.e. averaged over X. For fixed x we have

$$\Pr[\hat{S}=Q(X)|X=x] = \Upsilon_{Q(x)|Q(x),g(x)}. \quad (24)$$

Averaging over x gives

$$1-P_{\text{err}} = \mathbb{E}_x\Pr[\hat{S}=Q(X)|X=x] = \mathbb{E}_x\Upsilon_{Q(x)|Q(x),g(x)}$$
$$= \textstyle\sum_{s\in\mathcal{S}} p_s\int_0^1\mathrm{d}w\ \Upsilon_{s|sw}. \quad (25)$$

In the last step we used that x uniquely maps to $(s,w) = (Q(x),g(x))$. Eq. (25) together with (21) is the most convenient way to analytically express the error probability.

We consider the case where s is encoded as a Gray code. This is a well known technique to reduce the number of bit flips when a reconstruction error occurs. Table I lists the Gray

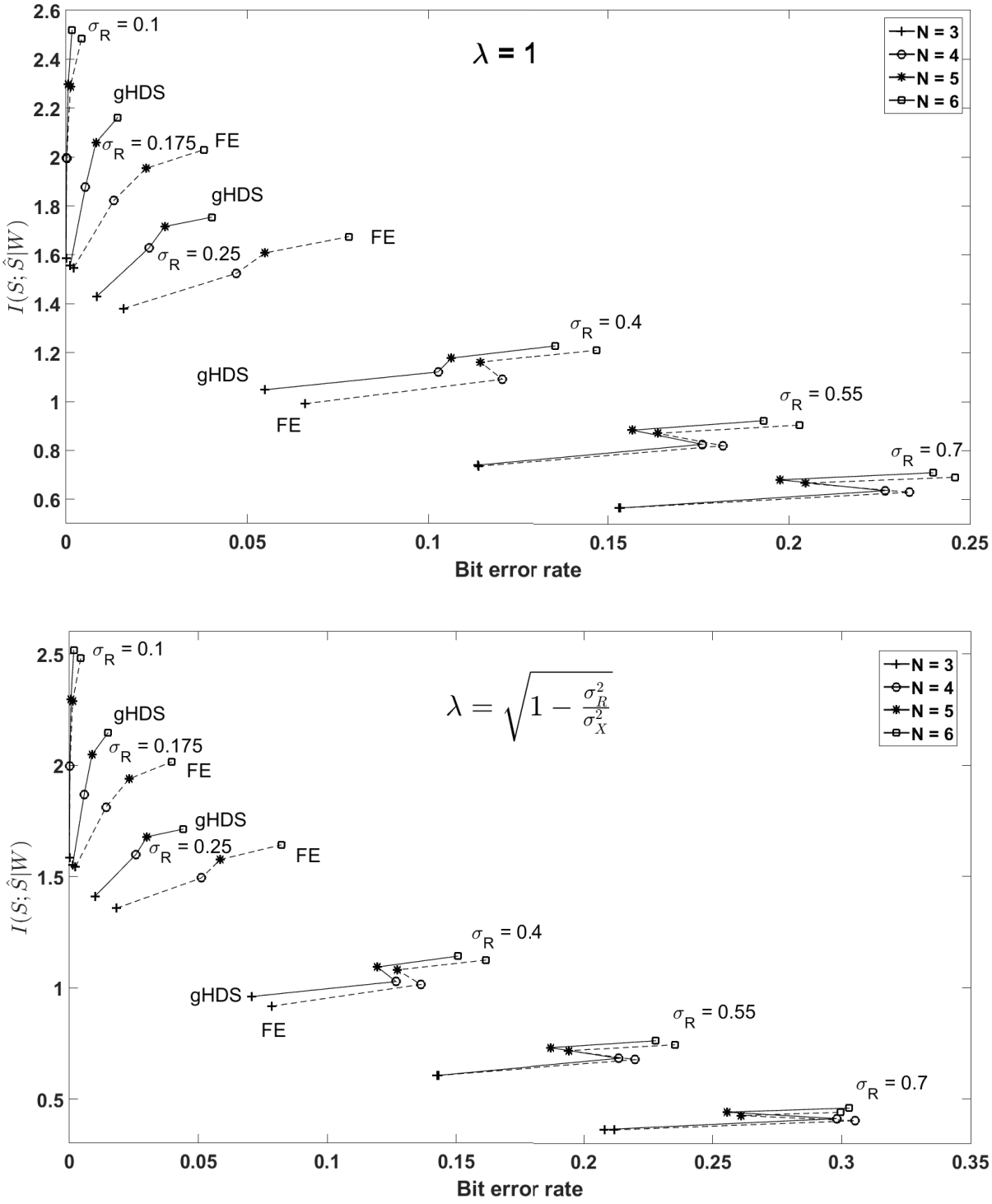

Fig. 5. Mutual information versus BER for perfect enrollment (upper figure) and identical conditions (lower figure). At fixed σ_R, data points for the general HDS are connected with a solid line, while a dashed line corresponds to the FE.

code. (Other, equivalent, encodings are possible.) We will look at $N \in \{3, 4, 5, 6\}$. The length of the Gray code is $\lceil \log N \rceil$.

The Bit Error Rate (BER) is given by

$$\text{BER} = \frac{\mathbb{E}[\#\text{ bit errors}]}{\lceil \log N \rceil} = \frac{1}{\lceil \log N \rceil} \sum_{t=0}^{\lceil \log N \rceil} t \Pr[\#\text{bit errors} = t]. \tag{26}$$

We introduce the following notation,

$$\Delta_{\hat{s}|s} \stackrel{\text{def}}{=} \Pr[\hat{S} = \hat{s} | S = s] = \mathbb{E}_w \Upsilon_{\hat{s}|sw}. \tag{27}$$

All the probabilities in (26) can be calculated in terms of the $\Delta_{\hat{s}|s}$ probabilities. The details are given in the Appendix.

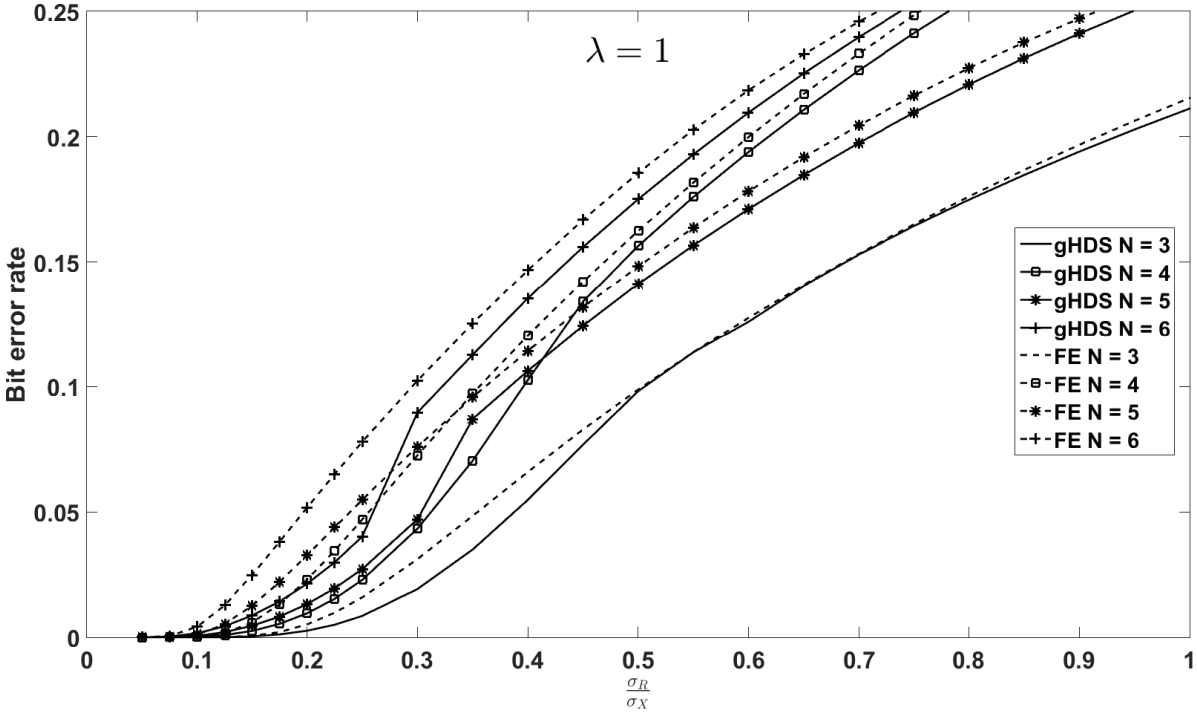

Fig. 6. Bit Error Rate as a function of the noise parameter σ_R/σ_X. Perfect enrollment.

IV. Numerical Results

We present numerical results for the optimization described in Section III, for $N \in \{3, 4, 5, 6\}$. We consider a Gaussian source X and Gaussian noise. (This is already a rather accurate model for Coating PUFs [18]). Without loss of generality we set $\sigma_X = 1$. Only the ratio σ_R/σ_X matters. We consider the two cases defined in Section II-C: *perfect enrollment* and *identical conditions*. We implemented (22) in Wolfram Mathematica 10.2 as a symbolic function. We used `FindMaximum` to obtain optimum values for $p_0, \ldots, p_{N-1}$. To reduce the dimension of the search space we imposed the symmetry $p_{N-1-\alpha} = p_\alpha$ by hand. In the figures the label 'gHDS' stands for *general* HDS (as opposed to FE). Fig. 4 shows $I(S; \hat{S}|W)$ versus P_{err} for various σ_R.

- For small σ_R, the optimum setting of the gHDS is close to the FE setting $p_\alpha = 1/N$, and it is visible that increasing N has a huge benefit for the mutual information.
- For somewhat larger σ_R, there is a clear difference between the optimised gHDS and the FE. For example, in the $\lambda = 1$ graph at $\sigma_R = 0.25$ we see that at $N = 6$ the transition from FE to gHDS brings a modest improvement of the mutual information and a reduction of P_{err} from $\approx 23\%$ to $\approx 10\%$. The reduced P_{err} means that the ECC in the second stage is much easier to implement for the gHDS than for the FE.
- At $\sigma_R > 0.5$ the noise is so bad that the gHDS and the FE perform almost equally badly (though the gHDS is always slightly better). Increasing N improves the mutual information only slightly, and at the cost of a large increase in P_{err}.

The choice of algorithm is not important. Eq. (22) is smooth and does not have spurious local maxima in the parameter region of interest.

Fig. 5 shows the same data, but with the BER on the horizontal axis. The 'zigzag' at the transition from $N = 4$ to $N = 5$ occurs because the Gray code jumps from a 2-bit representation of s to a 3-bit representation, with little noise in the first of the three bits.

Fig. 6 shows the BER as a function of σ_R/σ_X. The curves for $N = 4$ and $N = 5$ cross each other; this causes the 'zigzag' in Fig. 5. Apart from this crossing, in general increasing N increases the BER. The graphs of P_{err} as a function of σ_R/σ_X (Fig. 7) are much smoother. Unsurprisingly, for extremely strong noise the P_{err} for the gHDS and the FE is practically the same. For completeness Fig. 8 plots the BER versus P_{err}. The relation is nonlinear. Again it is visible that the gHDS outperforms the FE at every combination of σ_R/σ_X and N.

Fig. 9 shows the optimal values of $p_0, \ldots, p_{N-1}$ for the perfect enrollment case ($\lambda = 1$). At $\sigma_R = 0$ it holds that $p_\alpha = 1/N$ for all α, which is the FE configuration. When σ_R increases, the outer regions A_0, A_{N-1} shrink while the central region(s) become broader. (This makes intuitive sense: the FE's central intervals are more narrow than the outer intervals and therefore more susceptible to noise.) Then this trend reverses. At very large σ_R the p_α values stabilize, but not into the FE configuration.

Figs. 6–9 show $\lambda = 1$. The results for the identical conditions case look similar.

V. Summary

We have extended the results of de Groot *et al.* [5] in the case of non-equiprobable quantisation intervals. Lemma 3 gives the recipe for finding the optimal decision boundaries used in `Rec`. The result for Gaussian and Lorentzian noise is given in Theorems 1 and 2.

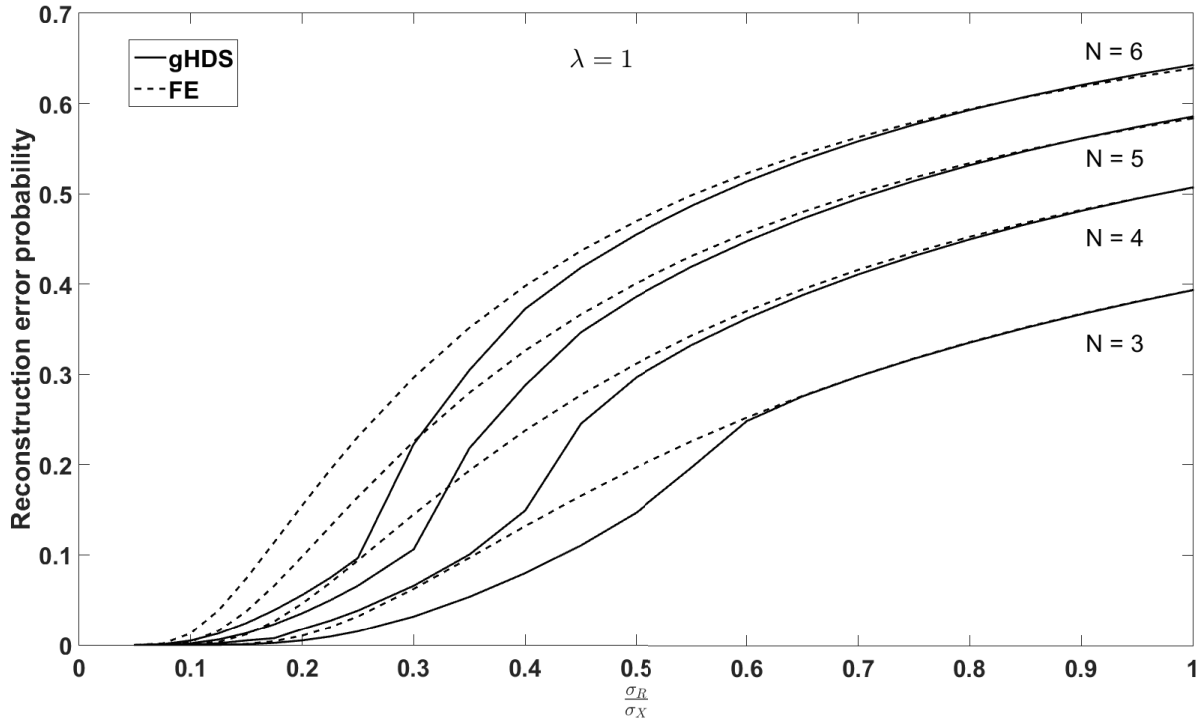

Fig. 7. P_{err} as a function of the noise parameter σ_R/σ_x. Perfect enrollment.

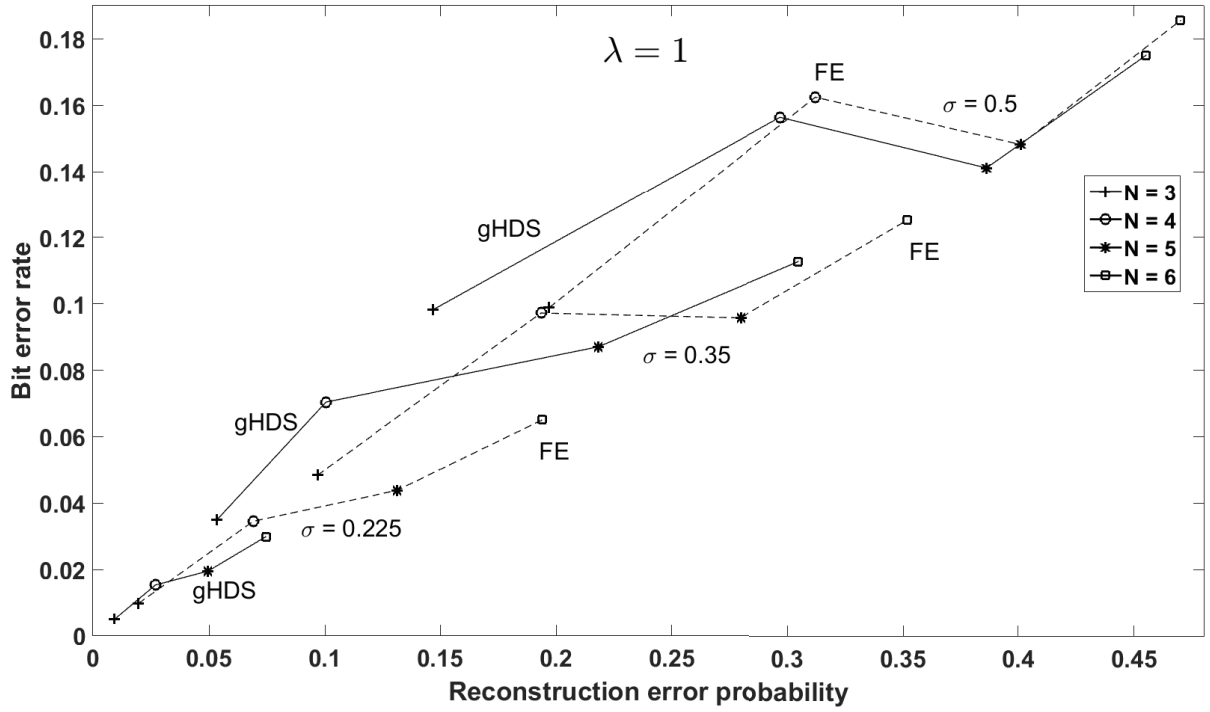

Fig. 8. BER versus reconstruction error probability P_{err}. Perfect enrollment. At given σ_R, data points for the HDS are connected with a solid line, while a dashed line corresponds to the FE.

We have studied the mutual information $I(S;\hat{S}|W)$, which is an upper bound on the amount of secret key material that can be robustly extracted from X. The mutual information is most conveniently expressed in terms of the p_s and $\Upsilon_{\hat{s}|sw}$ parameters (22). The dependence of the $\Upsilon_{\hat{s}|sw}$ on $p_0, \ldots, p_{N-1}$ is so complicated that optimisation of $I(S;\hat{S}|W)$ cannot be done analytically. The figures in Section IV show the results of numerical optimisation in a simple model where the source and the noise are Gaussian. Such a model is reasonably accurate for Coating PUFs. For every combination $(N, \sigma_R/\sigma_X)$ the optimized ZLHDS clearly performs better than the ZLFE in terms of both mutual information and bit error rate. The

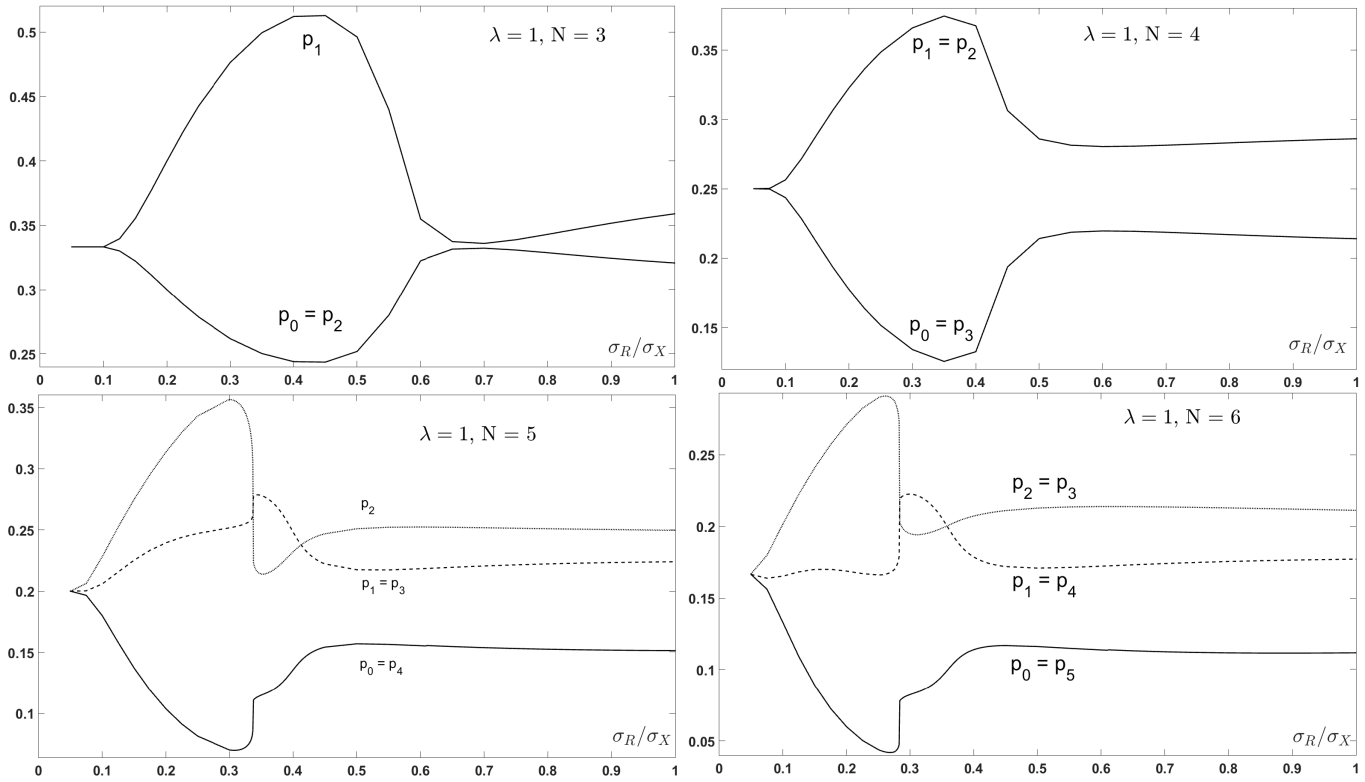

Fig. 9. The optimized p_α values as a function of the noise parameter σ_R/σ_X, for $\lambda = 1$, $N = 3, 4, 5, 6$.

reduction in the BER is substantial. This makes the design of a second-stage HDS much more practical, since it makes it easier to implement an ECC that can cope with the bit errors introduced by reconstruction errors.

We notice that the numerical results for the *perfect enrollment* and *identical conditions* noise models are very similar. This gives some indication (but does not prove) that mistakes in the estimation of the parameters σ_X, σ_R do not have serious consequences. This question is left for future work. Another topic for future work is applying the numerical optimisation to different source distributions.

APPENDIX

We list expressions for the BER (26) in terms of the $\Delta_{\hat{s}|s}$ probabilities (27), when the Gray code of Table I is used. We assume a symmetric source pdf f and symmetric noise. As a result the optimal p_α values have the symmetry $p_{N-1-\alpha} = p_\alpha$, and there is a large number of symmetries between the Δ-values, $\Delta_{N-1-\hat{s}|N-1-s} = \Delta_{\hat{s}|s}$.

N	$N\cdot$BER
3	$2p_0(\Delta_{1\|0} + 2\Delta_{2\|0}) + 2p_1\Delta_{2\|1}$
4	$2p_0(\Delta_{1\|0} + \Delta_{3\|0} + 2\Delta_{2\|0})$ $+2p_1(\Delta_{0\|1} + \Delta_{2\|1} + 2\Delta_{3\|1})$
5	$2p_0(\Delta_{1\|0} + \Delta_{3\|0} + 2\Delta_{2\|0} + 2\Delta_{4\|0})$ $+2p_1(\Delta_{0\|1} + \Delta_{2\|1} + 2\Delta_{3\|1} + 3\Delta_{4\|1})$ $+2p_2(\Delta_{1\|2} + 2\Delta_{0\|2})$
6	$2p_0(\Delta_{1\|0} + \Delta_{3\|0} + 2\Delta_{2\|0} + 2\Delta_{4\|0} + 3\Delta_{5\|0})$ $+2p_1(\Delta_{0\|1} + \Delta_{2\|1} + 2\Delta_{3\|1} + 2\Delta_{5\|1} + 3\Delta_{4\|1})$ $+2p_2(\Delta_{1\|2} + \Delta_{3\|2} + 2\Delta_{0\|2} + 2\Delta_{4\|2})$

The p-index in this table runs only to $\lceil N/2 \rceil - 1$ because of the $\alpha \leftrightarrow N - 1 - \alpha$ symmetry; this also gives rise to the factor 2 in front of each p_α. Inside the parentheses, the numerical factor in front of each Δ indicates the number of bit flips that occur due to that specific transition.

REFERENCES

[1] C. H. Bennett, G. Brassard, C. Crépeau, and M.-H. Skubiszewska, "Practical quantum oblivious transfer," in *Proc. CRYPTO*, 1991, pp. 351–366.
[2] C. Böhm and M. Hofer, *Physical Unclonable Functions in Theory and Practice*. New York, NY, USA: Springer, 2013.
[3] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *J. Comput. Syst. Sci.*, vol. 18, no. 2, pp. 143–154, Apr. 1979.
[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.
[5] J. de Groot, B. Škorić, N. de Vreede, and J.-P. Linnartz, "Quantization in zero leakage helper data schemes," *EURASIP J. Adv. Signal Process.*, vol. 2016, p. 54, Dec. 2015. [Online]. Available: https://eprint.iacr.org/2012/566
[6] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, vol. 38, no. 1, pp. 97–139, 2008.
[7] Y. Dodis, K. Pietrzak, and D. Wichs, "Key derivation without entropy waste," in *Proc. EUROCRYPT*, 2014, pp. 93–110.
[8] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Proc. EUROCRYPT*, vol. 3027. 2004, pp. 523–540.
[9] B. Gassend, "Physical random functions," M.S. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2003.
[10] J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "FPGA intrinsic PUFs and their use for IP protection," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 4727. Berlin, Germany: Springer, 2007, pp. 63–80.
[11] O. Günlü and O. İşcan, "DCT based ring oscillator physical unclonable functions," in *Proc. IEEE Int. Conf. Acoustic, Speech Signal Process. (ICASSP)*, May 2014, pp. 8198–8201.

[12] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. ACM Conf. Comput. Commun. Secur. (CCS)*, 1999, pp. 28–36.
[13] J.-P. Kaps, K. Yüksel, and B. Sunar, "Energy scalable universal hashing," *IEEE Trans. Comput.*, vol. 54, no. 12, pp. 1484–1495, Dec. 2005.
[14] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Audio- and Video-Based Biometric Person Authentication*. Berlin, Germany: Springer, 2003.
[15] R. Maes, *Physically Unclonable Functions: Constructions, Properties and Applications*. Heidelberg, Germany: Springer, 2013.
[16] A.-R. Sadeghi and D. Naccache, Eds., *Towards Hardware-Intrinsic Security*, Springer, 2010.
[17] D. R. Stinson, "Universal hashing and authentication codes," *Designs, Codes, Cryptograph.*, vol. 4, no. 3, pp. 369–380, 1994.
[18] P. Tuyls, G.-J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, and R. Wolters, "Read-proof hardware from protective coatings," in *Cryptographic Hardware and Embedded Systems—CHES* (Lecture Notes in Computer Science), vol. 4249. Vienna, Austria: Springer-Verlag, 2006, pp. 369–383.
[19] P. Tuyls, B. Škorić, and T. Kevenaar, *Security With Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. London, U.K.: Springer, 2007.
[20] E. A. Verbitskiy, P. Tuyls, C. Obi, B. Schoenmakers, and B. Škorić, "Key extraction from general nondiscrete signals," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 269–279, Jun. 2010.
[21] B. Škorić and N. de Vreede, "The spammed code offset method," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 5, pp. 875–884, May 2014.

Taras Stanko received the M.Sc. degree in theoretical physics in 2011 from Ivan Franko National University, Lviv, Ukraine. In 2015, he received a Professional Doctorate in Engineering (PDEng) in mathematics for industry. He is currently a Ph.D. student at Eindhoven University of Technology.

Fitria Nur Andini received the M.Sc. degree in computer science from Eindhoven University of Technology in 2015.

Boris Škorić received the Ph.D. degree in theoretical physics from the University of Amsterdam in 1999. From 1999 to 2008, he was a research scientist at Philips Research, working first on display physics and later on security topics. In 2008 he joined the Department of Mathematics and Computer Science at Eindhoven University of Technology, the Netherlands, as Assistant Professor.