Classification of High-Order Boolean Masking Schemes and Improvements of their Efficiency

Houssem MAGHREBI[1], Sylvain GUILLEY[1,2] and Claude CARLET[3] and Jean-Luc DANGER[1,2].

[1] TELECOM-ParisTech, Crypto Group,
37/39 rue Dareau, 75 634 PARIS Cedex 13, France.
[2] Secure-IC S.A.S.,
2 rue de la Châtaigneraie, 35 576 CESSON SEVIGNÉ, France.
[3] LAGA , UMR 7539, CNRS, Department of Mathematics,
University of Paris XIII and University of Paris VIII,
2 rue de la liberté, 93526 Saint-Denis Cedex, France. Email:
claude.carlet@univ-paris8.fr

Abstract. This article provides an in-depth study of high-order (HO) Boolean masking countermeasure against side-channel attacks. We introduce the notion of HO-CPA immunity as a metric to characterize a leakage function. We show that this notion intervenes to assess both the resistance against HO-CPA attacks and the amount of leakage. Namely, the HO-CPA immunity, denoted $\mathsf{HCI} \in \mathbb{N}^*$, coincides with the lowest order of a successful HO-CPA and gives the dependence of leakage behavior with the noise's variance σ^2 (according to $\mathcal{O}(1/\sigma^{2\times\mathsf{HCI}})$ in Landau notation). Then, we introduce the technique of leakage squeezing. It is an optimization of the straightforward masking where masks are recoded relevantly by bijections. Our main contribution is to show that the HO-CPA immunity of a masking countermeasure can be incremented by one or even by two at virtually no added cost. Indeed, the bijections (and inverse bijections) can be incorporated in tables that are often found in cryptographic algorithms (e.g. substitution boxes).

Keywords: High-Order Masking, High-Order Correlation Power Analysis (HO-CPA), High-Order CPA Immunity (HCI), Mutual Information Metric (MIM).

1 Introduction

Masking [10, Chp. 9] is a countermeasure against observation attacks, also known as side-channel attacks (SCA), that is suitable for both hardware and software cryptographic implementations. Indeed, it consists in changing the representation of variables into randomized shares [6,18], and can thus be qualified as a logical countermeasure. Notably, masking does not rely on specific hardware properties (as opposed to dual-rail protection [10, Chp. 7], that demands some physical indiscernibility).

Nonetheless, masked implementations can always be attacked, since all the shares [4] or a judicious combination [15] of them do unambiguously leak information about the sensitive variable. However, it is possible to give a formal definition for a high-order masking scheme: a d-order masking scheme involves $d+1$ shares. The security is reached at order d provided any combination of d variables conveys no information about the sensitive variable.

In fact, many purported solutions have been defeated [3,13,16]. One sound solution has been put forward recently in [17]. This solution can either be applied on software or hardware, because the unitary operation evaluation order is indifferent.

In this article, we illustrate the hardware implementation of d-order masking schemes. By hardware, we mean FPGA or ASIC circuits where identical resources (indiscernibility) can be instantiated and operated at the same time (parallelism). Some specificities of hardware simplify the design of masking countermeasures. We list our hypotheses below:

1. Regarding sequential resources (*i.e.* registers), the sum of their activity is leaked, and regarding the combinational logic, operations can be hidden in memories (executing precomputed masked tables).
2. In integrated devices manufactured with a fine minimal feature size, it is difficult to insulate the consumption of the various components; therefore, the attacker cannot distinguish one share from the other.
3. The noise is large, because of the parallelism. The variables unrelated to the attack act as independent noise sources, that can be modeled as a binomial law. If we take the example of the DES, a maximum of 4-bits is used as oracle. Consequently the noise consists in at least $64-4$ random variables, whose *independent* activity adds up. Hence a very good approximation of the algorithmic noise is a normal law, of variance $\sigma^2 \approx 60 \gg 1$.

In software, the computation basically unfolds the same. The difference with hardware is that variables are evaluated sequentially instead of in parallel. The attacker can thus choose adequate "combination function" using several measurements [14].

In the present paper, we will focus on hardware masking implementations. So, the only combination function available to the attacker is the arithemitic sum of individual leakages of the shares. This is done physically because of the parallelism in hardware devices execution.

The rest of the paper is structured as follows. The notion of HO-CPA immunity is first introduced (even independently of any masking

scheme) in Sec. 2. The security evaluation of hardware masking is studied in Sec. 3. An optimization of the HO-CPA immunity, thanks to a "leakage squeezing" enhancement of plain hardware Boolean masking, is explained in Sec. 4. Further improvements for specific bijections involved in the leakage squeezing are disclosed in this Sec. 4, where a summary of the results obtained in the paper is also given. Finally, section 5 concludes the paper and opens some further research perspectives.

2 HO-CPA Attacks and HO-CPA Immunity

In the context of a side-channel attack (possibly high-order [12]), we denote by L the leakage observations. Furthermore, we assume that the attacker can partition the observations according to a sensitive variable Z (that can be deduced from either the plaintext or the ciphertext through few sub-keys hypotheses). In our analysis, it is convenient to see L and Z as random variables (RVs). Their realization is denoted by small letters (l and z). In this section, L and Z can be arbitrary. In Sec. 3, we will explicit some expressions for them.

There are two ways to address the security evaluation of a countermeasure [20]:

1. Success of attacks (using metrics such as the success rate or the guessing entropy). Basically, there are two kinds of high-order attacks:
   (a) CPA [1], for which the optimal attack (at high orders) is defined in [15];
   (b) Information theoretic attacks, like the MIA [5], stochastic [19] or template [2].
2. Leakage estimation with information theoretic metrics, such as the mutual information between the leakage (observations) and the model.

CPA attacks are studied in Sec. 2.1. Information theoretic attacks and leakage metrics are jointly covered in Sec. 2.3.

2.1 HO-CPA Attacks

In a high-order CPA context, the attacker can compute $\rho((\mathcal{C}(L), f(Z))$ for every function $\mathcal{C} : L \mapsto (L - \mathbb{E}(L))^i$ for a given order i, where $\rho(\,\cdot\,,\,\cdot\,)$ is the Pearson correlation coefficient and f is the prediction function according to some assumptions on the device leakage model. Prouff *et al.* [15] have shown that the function f that maximizes the advantage of the attacker is $f_{\text{opt}}(z) = \mathbb{E}((L - \mathbb{E}(L))^i \mid Z = z)$ in an HO-CPA of order i. In this formula,

the symbol $\mathbb{E}$ represents the expectation. Incidentally, this quantity is also known in statistics as the central moment of order i of probability density function (pdf) $L \mid Z = z$, that we denote by $\mu_i(L \mid Z = z)$.

Remark 1. In hardware implementation, we mean by an attack of order i any attack that aims at exploiting the leakage central moment of order i. The difference with software is that the attacker cannot combine the shares; the sum of their activity L is leaked due to the parallelism in hardware devices execution.

We recall some notation which will be useful in the sequel. We denote by μ_z and σ_z^2 the mean and the variance of $L \mid Z = z$. We call $\mu_{\text{tot}} = \sum_z \mathsf{P}[Z = z]\mu_z$ the mean of L. The total variance σ_{tot}^2 of L decomposes into the sum of inter- and intra-class variance, denoted by σ_{inter}^2 and σ_{intra}^2. Those quantities are defined as: $\sigma_{\text{inter}}^2 \doteq \sum_z \mathsf{P}[Z = z](\mu_z - \mu_{\text{tot}})^2$ and $\sigma_{\text{intra}}^2 \doteq \sum_z \mathsf{P}[Z = z]\sigma_z^2$. We also introduce the cumulants $k_i(X)$ of the random variable X that are defined as: $\ln(\mathbb{E}(\exp(t \cdot X))) = \sum_{i=0}^{+\infty} k_i(X)\frac{t^i}{i!}$, for $t \in \mathbb{R}$.

In the presence of countermeasures, the central moment $\mu_i(L \mid Z = z)$ of order i can be independent of z. In practice, the attacker will typically try to compute the moments starting from low orders, because their estimation is less affected by the measurement noise.

2.2 HO-CPA Immunity

In this section, we define the notion of HO-CPA immunity (denoted by HCI) to quantify the difficulty of an attack.

Definition 1. *The HO-CPA immunity of the random variable $L \mid Z$ is equal to the minimal value $i \in \mathbb{N}^*$ such that $\forall j \in [\![0, i-1]\!]$, the (central) moment of order j of the distribution $L \mid Z = z$ is constant with respect to z.*

This notion is always defined, because the moments of order 0 of any distribution X are equal to 1 (the integral of a pdf). Thus, the minimal value of the HO-CPA immunity is 1 when the distributions do not have the same mean. Said differently, $\mathsf{HCI} = 1$ if $\mu_1(L \mid Z = z)$ depends on z. This is the case of unprotected circuits, for which a first-order CPA [1] works.

The HO-CPA immunity is larger than or equal to 2 when the distributions are balanced (*i.e.* $\forall z, \mu_z = \mu_{\text{tot}}$). In this case, the inter-class

variance is null and the total variance σ^2_{tot} is equal to the intra-class variance $\sigma^2_{\text{intra}} = \sum_z \mathsf{P}[Z = z] \times \mu_2(L \mid Z = z)$. If the $\mu_2(L \mid Z = z)$ are not all equal, then $\mathsf{HCI} = 2$ and a second-order CPA using f_{opt} of order 2 (or a variance-based attack [9]) is possible.

The motivation of the HO-CPA immunity definition is thus straightforward. As noted in Sec. 2.1, all HO-CPA using f_{opt} of order $i < \mathsf{HCI}$ will fail, because the optimal function $f_{\text{opt}}(z)$ is independent of z. Thus the HO-CPA immunity is equal to the smallest order of f_{opt} for which an HO-CPA attack can be successful.

2.3 Link Between $\mathsf{I}(L + N; Z)$ and the HO-CPA Immunity

The HO-CPA reveals linear dependencies between L and Z. Unless the random variables $L \mid Z = z$ are identically distributed for every z, the mutual information $\mathsf{I}(L; Z)$ will be non-zero.

This means that there is no such notion of "order" for MIA or leakage metrics. Nonetheless, it is interesting to compare the value of $\mathsf{I}(L; Z)$ for different leakage functions L. To do so, we notice that in real measurements, the observations L are noisy. We assume that in the presence of noise, the observations are added an Additive White Gaussian Noise (AWGN) $N \sim \mathcal{N}(0, \sigma^2)$. The Gaussian model assumption is both very usual in the side channel literature and fairly realistic in practice (see for instance [10, §IV]). The interpretation of the link between HO-CPA immunity and the HO-CPA success given in Sec. 2.2 is not affected by N if it is independent of L and Z, because in this case:

Lemma 1. $\forall i \in [\![0, \mathit{HCI}[\![, \rho((L - \mathbb{E}(L))^i, Z) = 0 \Rightarrow \forall i \in [\![0, \mathit{HCI}[\![, \rho((L + N - \mathbb{E}(L + N))^i, Z) = 0$.

Proof. The proof is given in Appendix A.1. □

In the case of the mutual information, the impact of the noise N is quantified by Theorem 1.

Theorem 1. *Under the Gaussian assumption,* $\mathsf{I}(L+N; Z) = \mathcal{O}\left(\sigma^{-2\times\mathsf{HCI}}\right)$ *asymptotically when* $\sigma \to +\infty$.

Proof (of Theorem 1 for $\mathsf{HCI} \in \{1, 2\}$*).* If $\mathsf{HCI} \leq 2$, we can use the Gaussian assumption, since the distributions have either different means ($\mathsf{HCI} = 1$) or variances ($\mathsf{HCI} = 2$). Adding N to either $L \mid Z = z$ does not change their mean. The impact on their variance is merely to add σ^2 (because those variables are independent).

Now the mutual information $\mathsf{I}(L+N;Z)$ can be expressed in terms of Kullback-Leibler divergence [22]: $\mathsf{I}(L+N;Z) = \mathbb{E}_Z \mathrm{D}_{\mathrm{KL}}(L+N \parallel L+N \mid Z)$. Under the Gaussian assumption, $L+N \sim \mathcal{N}(\mu_{\mathrm{tot}}, \sigma^2_{\mathrm{tot}} + \sigma^2)$ and $L+N \mid Z=z \sim \mathcal{N}(\mu_z, \sigma_z^2+\sigma^2)$. The Kullback-Leibler divergence of two normal laws has an analytical expression, from which it can be derived the following result:

$$\mathsf{I}(L+N;Z) = -\frac{1}{2} \times \frac{\sigma^2_{\mathrm{inter}}}{\sigma^2_{\mathrm{tot}}+\sigma^2} + \frac{1}{2\ln 2} \sum_z \mathsf{P}[Z=z] \ln \frac{1+\sigma_z^2/\sigma^2}{1+\sigma^2_{\mathrm{tot}}/\sigma^2} \,. \quad (1)$$

If $\mathsf{HCI}=1$, then $\sigma^2_{\mathrm{inter}} \neq 0$. The first order Taylor expansion $\ln(1+\epsilon) = \epsilon + \mathcal{O}(\epsilon)$ on $\epsilon = 1/\sigma^2$ yields: $\mathsf{I}(L+N;Z) = -\left(\frac{1}{2\sigma^2_{\mathrm{tot}}+2\sigma^2} + \frac{1}{2\sigma^2 \ln 2}\right)\sigma^2_{\mathrm{inter}} + \mathcal{O}(\frac{1}{\sigma^2})$, which is about equivalent to $\mathcal{O}(\frac{1}{\sigma^2})$ when σ^2 increases ($\sigma^2 \gg \sigma^2_{\mathrm{tot}}$).

If $\mathsf{HCI}=2$, then $\sigma^2_{\mathrm{inter}}=0$, but $\sigma^2_{\mathrm{tot}} = \sigma^2_{\mathrm{intra}} \neq 0$. Thus, by developing the logarithm at order 2 in $\epsilon = 1/\sigma^2$, we get $\mathsf{I}(L+N;Z) = \sum_z \frac{1}{2\sigma^4} \sum_z \mathsf{P}[Z=z]\left(\sigma^4_{\mathrm{tot}} - \sigma_z^4\right) + \mathcal{O}(\frac{1}{\sigma^4})$. Now, it remains to prove that $\sum_z \mathsf{P}[Z=z]\left(\sigma^4_{\mathrm{tot}} - \sigma_z^4\right) \neq 0$. This is actually true, by the application of the Cauchy-Schwarz theorem on $x_z = \sqrt{\mathsf{P}[Z=z]}$ and $y_z = \sqrt{\mathsf{P}[Z=z]}\sigma_z^2$. Indeed, we have: $\left(\sum_z x_z \cdot y_z\right)^2 \leq \left(\sum_z x_z^2\right) \cdot \left(\sum_z y_z^2\right)$, with equality if and only if x_z and y_z, seen as vectors, are colinear. As at least one σ_z is different from the others, the inequality is strict. Thus: $\left(\sum_z \mathsf{P}[Z=z]\sigma_z^2\right)^2 < \sum_z \sqrt{\mathsf{P}[Z=z]}^2 \cdot \sum_z \left(\sqrt{\mathsf{P}[Z=z]}\sigma_z^2\right)^2 = \sum_z \mathsf{P}[Z=z]\sigma_z^4$. □

Before proving Theorem 1 for $\mathsf{HCI} > 2$, let us introduce the following useful lemma.

Lemma 2. *If L has HCI immunity, then $\forall i \in [\![0, \mathsf{HCI}[\![$, $\forall z, k_i(L \mid Z=z) = k_i(L)$.*

Proof. The proof is given in Appendix A.2. □

We continue hereafter the proof of Theorem 1.

Proof (of Theorem 1 for $\mathsf{HCI} > 2$). If $\mathsf{HCI} > 2$, then all the σ_z^2 are equal, and σ^2_{tot} is equal to them too. The Gaussian assumption leads to the conclusion that $\mathsf{I}(L+N;Z)=0$ (refer to Eqn. 1). Therefore, we refine the density probabilities expressions thanks to an Edgeworth expansion. We reuse the result shown in Lemma 2 of [8] (extended to laws of common variance $\sigma^2_{\mathrm{tot}} + \sigma^2 \neq 1$):

$$\mathsf{I}(L+N;Z) = \sum_{i=3}^{+\infty} \frac{1}{2 \cdot i!} \sum_z \mathsf{P}[Z=z] \frac{\left(k_i(L \mid Z=z) - k_i(L)\right)^2}{\left(\sigma^2_{\mathrm{tot}} + \sigma^2\right)^i} \,. \quad (2)$$

In this equation, we use the fact that the cumulants of order i of two independent RVs are the sums of their individual cumulants. Now, the moments of N are all null but for $k_2(N) = \sigma^2$. This is why $\forall i \geq 3, k_i(L + N \mid Z = z) = k_i(L \mid Z = z)$ and $k_i(L + N) = k_i(L)$. Then, according to Lemma 2, the first non-zero term in this summation is at index $i = \mathsf{HCI}$, which achieves to prove Theorem 1. □

Our main interest in Theorem 1 is that it gives the dependence of leakage behavior with the noise's variance σ^2. The higher HCI (*i.e.* the more statistical moments of $L \mid Z = z$ are constant with respect to z), the less information is leaked by the device.

3 High-Order Hardware Boolean Masking

3.1 Definitions

In a masking scheme with d masks, in addition to the sensitive variable Z, we denote by $M_{i\in[\![1,d]\!]}$ the masks. The $d+1$ shares are $a_0 = Z \oplus \bigoplus_{i=1}^{d} M_i, a_1 = M_1, \cdots, a_d = M_d$. In hardware, all the shares can be manipulated concomitantly. Let us for instance consider the computation of a combinational function $S : x \in \{0,1\}^n \mapsto S(x) \in \{0,1\}^n$, where n denotes the bitwidth of the sensitive word under analysis. Typically, d arbitrary bijective functions $S_{i\in[\![1,d]\!]}$ are used to update the masks for the next clock period, and a function $S_0 : (a_0, a_1, \cdots, a_d) \in \{0,1\}^{n\times(d+1)} \mapsto S(\bigoplus_{i=0}^{d} a_i) \oplus \bigoplus_{i=1}^{d} S_i(a_i)$ achieves the functionally. This scheme is illustrated in Fig. 1(i). It is a generalization from the scheme of first-order $(d = 1)$ protection discussed in [21].

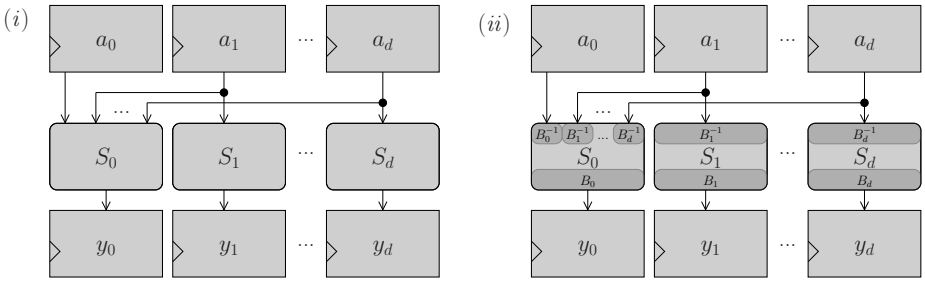

Fig. 1. Hardware d-th order masking of function S, plain (i) and with leakage squeezing (ii).

How the S_i functions are computed is out of the scope of this article. Typically, we warn that the most critical function, namely S_0, must be glitch-free. Indeed, this function receives in input all the shares, and [11] has shown that such configuration can leak even at the first-order. A straightforward implementation in logic cell would certainly make the variable $\bigoplus_{i=0}^{d} a_i$ appear, which is catastrophic since it is exactly the sensitive variable z. Therefore, the general recommendation is to fit S_0 in a RAM, since memories are designed not to glitch [7, §IV.1].

In the sequel, we focus on the security analysis of the registers (of symbol ▷□). According to our hypotheses about hardware (stated in Sec. 1), the registers are equivalent. Therefore, the leakage is invariant by a permutation on the bits. Additionally, we suppose that the attacker cannot specifically distinguish one register from the others. Hence, the leakage function simplifies in a Hamming weight or distance. In Hamming weight:

$$L(Z, M_1, \cdots, M_d) = \mathrm{HW}(Z \oplus \bigoplus_{i=1}^{d} M_i) + \sum_{i=1}^{d} \mathrm{HW}(M_i) \ . \tag{3}$$

In Hamming distance, we can still stick to the same leakage function, since if we denote by prime letters the future state and by ΔZ the bitwise difference $Z \oplus Z'$, then the leakage function of Eqn. (3) rewrites:

$$\begin{aligned} &L(Z, M_1, \cdots, M_d, Z', M'_1, \cdots, M'_d) = \\ &\mathrm{HW}(Z \oplus \bigoplus_{i=1}^{d} M_i \oplus Z' \oplus \bigoplus_{i=0}^{d} M'_i) + \sum_{i=1}^{d} \mathrm{HW}(M_i \oplus M'_i) = \\ &\mathrm{HW}(\Delta Z \oplus \bigoplus_{i=1}^{d} \Delta M_i) + \sum_{i=1}^{d} \mathrm{HW}(\Delta M_i) = L(\Delta Z, \Delta M_1, \cdots, \Delta M_d) \end{aligned} \tag{4}$$

which is the same as Eqn. (3) where every variable is replaced by the difference of variables.

3.2 Resistance of HO Hardware Boolean Masking Against HO-CPA

In this section, we prove that the Hardware Boolean countermeasure with d masks has HO-CPA immunity $\mathsf{HCI} = d + 1$, and thus protects against $(d+1)^{\text{th}}$-order CPA.This is illustrated for $n = 4$ in Tab. 1 first five groups, that correspond to $d \in [\![0, 4]\!]$. In this table, the number of lines in gray is equal to $\mathsf{HCI} - 1$ (Definition 1).

We start the demonstration by Lemma 3.

Lemma 3. *Let α_i be $d + 1$ natural integers, for $i \in [\![0, d]\!]$. If one α_i is equal to zero, then the expression:*

$$\sum_{m_1, \cdots, m_d} HW^{\alpha_0}(z \oplus \bigoplus_{i=1}^{d} m_i) \cdot \prod_{i=1}^{d} HW^{\alpha_i}(m_i) \tag{5}$$

Table 1. Statistics about some leakage models on words of $n = 4$ bitwidth, without noise (*i.e.* $\sigma = 0$).

	R.V.	L	$L \mid Z = 0$	$L \mid Z = 1$	$L \mid Z = 2$	$L \mid Z = 3$	$L \mid Z = 4$
Plain hardware implementation (Eqn. (3)) with $d = 0$ mask (unprotected reference).							
HCI = 1	$\mu_1 = \mathbb{E}(\cdot)$	2.000	0.000	1.000	2.000	3.000	4.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	1.000	0.000	0.000	0.000	0.000	0.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	0.000	0.000	0.000	0.000	0.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	2.500	0.000	0.000	0.000	0.000	0.000
	Entropy [bit]	2.031	0.000	0.000	0.000	0.000	0.000
Plain hardware implementation (Eqn. (3)) with $d = 1$ mask.							
HCI = 2	$\mu_1 = \mathbb{E}(\cdot)$	4.000	4.000	4.000	4.000	4.000	4.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	2.000	4.000	3.000	2.000	1.000	0.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	0.000	0.000	0.000	0.000	0.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	11.000	40.000	21.000	8.000	1.000	0.000
	Entropy [bit]	2.544	2.031	1.811	1.500	1.000	0.000
Plain hardware implementation (Eqn. (3)) with $d = 2$ masks.							
HCI = 3	$\mu_1 = \mathbb{E}(\cdot)$	6.000	6.000	6.000	6.000	6.000	6.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	3.000	3.000	3.000	3.000	3.000	3.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	-3.000	-1.500	0.000	1.500	3.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	25.500	25.500	25.500	25.500	25.500	25.500
	Entropy [bit]	2.839	1.762	1.822	1.836	1.822	1.762
Plain hardware implementation (Eqn. (3)) with $d = 3$ masks.							
HCI = 4	$\mu_1 = \mathbb{E}(\cdot)$	8.000	8.000	8.000	8.000	8.000	8.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	4.000	4.000	4.000	4.000	4.000	4.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	0.000	0.000	0.000	0.000	0.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	46.000	52.000	49.000	46.000	43.000	40.000
	Entropy [bit]	3.047	2.044	2.047	2.046	2.043	2.031
Plain hardware implementation (Eqn. (3)) with $d = 4$ masks.							
HCI = 5	$\mu_1 = \mathbb{E}(\cdot)$	10.000	10.000	10.000	10.000	10.000	10.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	5.000	5.000	5.000	5.000	5.000	5.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	0.000	0.000	0.000	0.000	0.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	72.500	72.500	72.500	72.500	72.500	72.500
	Entropy [bit]	3.208	2.207	2.208	2.208	2.208	2.207
Leakage squeezing hardware implementation with $d = 1$ mask and $B = \overline{I_4}$ (Eqn. (13)).							
HCI = 4	$\mu_1 = \mathbb{E}(\cdot)$	4.000	4.000	4.000	4.000	4.000	4.000
	$\mu_2 = \mathbb{E}((\cdot - \mu_1)^2)$	2.000	2.000	2.000	2.000	2.000	2.000
	$\mu_3 = \mathbb{E}((\cdot - \mu_1)^3)$	0.000	0.000	0.000	0.000	0.000	0.000
	$\mu_4 = \mathbb{E}((\cdot - \mu_1)^4)$	11.000	32.000	11.000	8.000	11.000	8.000
	Entropy [bit]	2.544	0.669	1.544	1.500	1.544	1.500

does not depend on z. In Eqn. (5), $HW^{\alpha}(m)$ is the α^{th} power of $HW(m)$, with the convention that $\forall m, HW^{0}(m) = 1$.

Proof. If $\alpha_0 = 0$, then, by definition, Eqn. (5) does not depend on z. Let us assume $\alpha_i = 0$ with $i > 0$. If we replace the summation variable m_i with $m_i \oplus z$, then we come back to the previous case. □

As a corollary to Lemma 3, for the leakage function defined in Eqn. (3), the HO-CPA immunity HCI is strictly greater than d. Indeed, the development of L^j will imply terms like Eqn. (5) such that $\sum_{i=0}^{d} \alpha_i \leq j$. Thus, to have all α_i non-zero, j must be strictly greater than d. In application of the interpretation of HCI in Sec. 2.2, the first HO-CPA able to extract the correct key uses f_{opt} of order $\mathsf{HCI} = d + 1$.

Theorem 2. *The optimal prediction function for $d+1$-order CPA on the hardware countermeasure with d masks is the Hamming weight of the sensitive variable.*

Proof. When $j = d + 1$, all the terms in the decomposition of L^j are independent of z but the one corresponding to the case for which all the α_i are equal to 1. In this case, the exact expression for Eqn. (5) is:

$$\mathbb{E}\left(\mathrm{HW}(Z \oplus \bigoplus_{i=1}^{d} M_i) \cdot \prod_{i=1}^{d} \mathrm{HW}(M_i) \mid Z = z\right) = \left(-\tfrac{1}{2}\right)^d \left(\mathrm{HW}(z) + \tfrac{n}{2}\left((-n)^d - 1\right)\right) \quad . \tag{6}$$

This equality relies on this relationship: $\mathbb{E}(\mathrm{HW}(Z \oplus M) \times \mathrm{HW}(M) \mid Z = z) = \frac{n^2+n}{4} - \frac{1}{2}\mathrm{HW}(z)$, proved in Eqn. (19) of [15]. If at all orders $1 \leq i \leq d-1$, Eqn. (5) writes $\frac{n^2+n}{4}u_i + v_i\mathrm{HW}(z)$, then this is also true at order d. Moreover $u_d = \frac{n}{2}u_{d-1} + u_1 v_{d-1}$ and $v_d = v_{d-1}v_1$. As such expressions exist for $d = 1$ ($u_1 = 1$ and $v_1 = -1/2$), the induction shows it is true at all order $d \geq 1$. Basic arithmetic shows that $v_d = (-1/2)^d$ and that $u_d = \sum_{i=0}^{d-1}(n/2)^i(-1/2)^{d-1-i} = (-1/2)^{d-1}\sum_{i=0}^{d-1}(-n)^i = (-1/2)^{d-1}\frac{1-(-n)^d}{1+n}$. Now, the correlation with a linear combination of $z \mapsto \mathrm{HW}(z)$ is the same as with merely $\mathrm{HW}(z)$, which achieves to prove that the Hamming weight of the sensitive variable is the best model in HO-CPA against the hardware masking countermeasure. □

3.3 Leakage Estimation of Hardware Boolean Masking

The most accurate leakage estimation requires no model, as opposed to HO-CPA. Instead, all the 2^n values of sensitive variable Z are kept for the partitioning to be the most relevant. Nevertheless, due to the invariance

of the assumed leakage model in the bits order, the partitions degenerate in classes indexed by $\mathrm{HW}(Z)$ or $\mathrm{HW}(\Delta Z)$, belonging to $[\![0, n]\!]$. This means that $\mathsf{I}(L+N;Z) = \mathsf{I}(L+N;\mathrm{HW}(Z))$. Thus, the leakage function in hardware implementation (Eqn. (3)) has also $\mathsf{HCI} = d+1$.

The mutual information at low σ (e.g. $\sigma = 0$) is of little interest for hardware. Indeed, as already argued, hardware implementations being parallel, they have a high algorithmic noise level. Furthermore, regarding applications programmed in FPGA, the FPGA itself is very noisy. Eventually, it is not costly to add noise generators (such as a free running LFSRs) in FPGAs since it always remains a few LUTs (Look-Up Tables) that can be filled this way. So the tendency when $\sigma \to +\infty$ is interesting, and thus, the result of Theorem 1 applies: with d masks, the ratio $\log(\mathsf{I}(L+N;\mathrm{HW}(Z))/\log(\sigma^2)$ is asymptotically equal to $-\mathsf{HCI} = -d-1$. This is confirmed by numerical computations, shown in Fig. 2.

4 Proposed Masking Method for "Leakage Squeezing"

4.1 Principle

The goal of the leakage squeezing is to transform the intermediate variables so as to break the too strong link between the shares. For this purpose, $GF(2)^n \to GF(2)^n$ functions B_i are applied to the $d+1$ shares, as illustrated in Fig. 1(*ii*). It is important that the B_i are bijections, otherwise the entropy of the masks is reduced. Also, the transformation B_i must be inverted to recover the functional values. The leakage function L_{LS} for this countermeasure can be expressed for hardware approach by:

$$L_{\mathrm{LS}}(Z, M_1, \cdots, M_d) = \mathrm{HW}(B_0(Z \bigoplus_{i=1}^{d} M_i)) + \sum_{i=1}^{d} \mathrm{HW}(B_i(M_i)) \qquad (7)$$

4.2 Security Evaluation of the Countermeasure for One Mask

To simplify the study, we focus on $d = 1$ mask, and we also adopt $B_0 = I_n$ (the identity function) and search for a good bijection B_1, denoted simply by B. This section starts with a security analysis of the leakage squeezing countermeasure in order to determinate the efficiency of various distinguishers in exploiting the leakage of (Eqn. (7)). So, the leakage is:

$$L_{\mathrm{LS}}(Z, M) = \mathrm{HW}(Z \oplus M) + \mathrm{HW}(B(M)) \ , \qquad (8)$$

thus, the optimal function for attacks of order i with leakage squeezing is:

$$\begin{aligned} f_{\text{opt}}(z) &= \mathbb{E}_M\left([\text{HW}(z \oplus M) + \text{HW}(B(M)) - \mathbb{E}_M(\text{HW}(z \oplus M) + \text{HW}(B(M)))]^i\right) \\ &= \mathbb{E}_M\left((\text{HW}(z \oplus M) + \text{HW}(B(M)) - n)^i\right) \\ &= \sum_{k=0}^{i}\sum_{l=0}^{k}\binom{i}{k}\binom{k}{l}(-n)^{i-k}\mathbb{E}_M\left(\text{HW}(z \oplus M)^l \text{HW}(B(M))^{k-l}\right) \quad . \end{aligned} \tag{9}$$

Remarkable conclusions of the computation of f_{opt} are:

1. If the bijection B is randomly generated over $GF(2)^n$ – *i.e.* it is a RV – then the optimal function is constant for the different combining functions. Hence, the correlation attacks will not succeed since ρ_{opt} will be null. This method nevertheless requires additional resources (*e.g.* more memories in a FPGA to generate the random bijections using Random Number Generators).
2. If we use one sole bijection B (a constant hardwired transformation), then it should satisfy:

$$fopt_{B,p,q}(z) \doteq \mathbb{E}_M(\text{HW}(z \oplus M)^p \cdot \text{HW}(B(M)))^q = \text{constant} \quad , \tag{10}$$

for some $p, q \in \mathbb{N}$. This solution is the case of study for the rest of this paper.

4.3 Constructive Search of B: Boolean Theory

Eqn. (5) for all $\alpha_i = 1$ also writes as a convolution: $\bigotimes_{i=0}^{d} \text{HW}(z)$. If we restrict to $d = 2$, we similarly notice that $fopt_{B,p,q}$ in Eqn. (10) rewrites $\text{HW}^p \otimes (\text{HW} \circ B)^q$.

Case where $p = 1$ and $q = 1$ We found that some linear functions satisfy Eqn. (10). There are two strong incentives to research for linear bijections B:

1. There is a theory, which helps guide the search (as will be shown below).
2. The leakage squeezing obtained this way still applies both to Hamming weight (Eqn. (3)) and distance (Eqn. (4) remains equivalent to Eqn. (3), since $\Delta B(M) = B(\Delta M)$).

More precisely, these functions are detailed in Theorem 3.

Theorem 3. *The linear functions $B : GF(2)^n \to GF(2)^n$, represented by $m \mapsto B(m) = mA^\mathsf{T}$, with A an invertible $n \times n$ matrix over $GF(2)$, such that all lines have strictly more than one 1, are verifying Eqn. (10) for $p = q = 1$.*

Proof. The demonstration uses the Walsh-Hadamard transform: it maps a bijection f into $\widehat{f}$, defined as $\widehat{f}(u) \doteq \sum_{x \in GF(2)^n} f(x)(-1)^{x \cdot u}$. An appealing property of the Walsh-Hadamard transform is that it turns a convolution into a product:

$$\begin{aligned} fopt_{B,1,1}(z) = \text{cst} &\Leftrightarrow \widehat{fopt_{B,1,1}}(a) \propto \delta(a) \text{ [where } \propto \text{ means "is proportional to"]} \\ &\Leftrightarrow \widehat{\mathrm{HW}}(a) \times \widehat{\mathrm{HW} \circ B}(a) = \left(n \times 2^{n-1}\right)^2 \times \delta(a) \\ &\Leftrightarrow \forall a \neq 0, \widehat{\mathrm{HW}}(a) = 0 \text{ or } \widehat{\mathrm{HW} \circ B}(a) = 0\,. \end{aligned} \tag{11}$$

Now, if we denote by e_i the lines of the identity matrix I_n of size $n \times n$,

$$\begin{aligned} \widehat{\mathrm{HW}}(a) &= \sum_{z \in GF(2)^n} \frac{1}{2} \sum_{i=1}^{n} \left(1 - (-1)^{z_i}\right) (-1)^{a \cdot z} \\ &= n \cdot 2^{n-1} \delta(a) - \tfrac{1}{2} \textstyle\sum_{z \in GF(2)^n} \sum_{i=1}^{n} (-1)^{(a \oplus e_i) \cdot z} \\ &= \begin{cases} n \cdot 2^{n-1} & \text{if } a = 0, \\ -2^{n-1} & \text{if } \exists i \in [\![1, n]\!], \text{ such that } a = e_i, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{12}$$

Thus, the problem comes down to finding a function B such that: $\widehat{\mathrm{HW} \circ B}(a) = 0$ for $a = e_i$. This is the case of B nonzero linear, as long as $m \mapsto B_i(m) = \alpha_i \cdot m$ satisfies $\nexists j$ such that $\alpha_i = e_j$. □

Such matrices exist if n is even and greater or equal to 4. For instance, $\overline{I_n}$ (the complement of the identity I_n) satisfies Theorem 3. Indeed, the matrix $\overline{I_n}$ (abridged $\overline{I}$ in this §) has trivially more than one 1 per line. In addition, it is invertible (in the case $n \in 2\mathbb{N}^* \backslash \{2\}$), because it is orthogonal. We recall that the product of two binary matrices $A_{i,j}$ and $B_{i,j}$, where $i, j \in [\![0, n[\![$ is defined as $(A \cdot B)_{i,j} = \bigoplus_k A_{i,k} \cdot B_{k,j}$. Now we note that $\overline{I}$ is symmetric, because $\overline{I}^\mathsf{T} = \overline{I}$. Also, the elements of $\overline{I}$ are $\overline{I}_{i,j} = \delta(i \neq j)$,

where δ is the Kronecker symbol. Thus:

$$\begin{aligned}\left(\overline{I}^{\mathsf{T}} \cdot \overline{I}\right)_{i,j} &= \bigoplus_{k=0}^{n-1} \delta(i \neq k) \cdot \delta(k \neq j) \\ &= \begin{cases} \bigoplus_{k=0}^{n-1} \delta(i \neq k) = \underbrace{\bigoplus_{k \neq i} 1}_{\text{Odd number of 1s}} \oplus 0 = 1 & \text{if } i = j, \\ \underbrace{\bigoplus_{k \neq i \text{ and } k \neq j} 1}_{\text{Even number of 1s}} \oplus \bigoplus_{k=i \text{ or } k=j} 0 = 0 \oplus 0 = 0 & \text{if } i \neq j; \end{cases} \\ &= \delta(i = j) = (I)_{i,j} \quad . \end{aligned} \tag{13}$$

Note: The application of $\overline{I}$ on x is not equivalent to computing $\overline{x}$. For example, if $n = 4$, the linear function defined by the application of $\overline{I_4}$ maps $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f\}$ to $\{0, e, d, 3, b, 5, 6, 8, 7, 9, a, 4, c, 2, 1, f\}$.

For $n = 8$, we can also use the circulant matrix involved in the AES:

$$A8 = \begin{pmatrix} 1&0&0&0&1&1&1&1 \\ 1&1&0&0&0&1&1&1 \\ 1&1&1&0&0&0&1&1 \\ 1&1&1&1&0&0&0&1 \\ 1&1&1&1&1&0&0&0 \\ 0&1&1&1&1&1&0&0 \\ 0&0&1&1&1&1&1&0 \\ 0&0&0&1&1&1&1&1 \end{pmatrix}. \tag{14}$$

For n odd, some circulant matrices can also be used. For instance, when $n = 5$, there is the couple:

$$A5 = \begin{pmatrix} 1&1&1&0&0 \\ 0&1&1&1&0 \\ 0&0&1&1&1 \\ 1&0&0&1&1 \\ 1&1&0&0&1 \end{pmatrix} \quad \text{and} \quad A5^{-1} = \begin{pmatrix} 0&1&1&0&1 \\ 1&0&1&1&0 \\ 0&1&0&1&1 \\ 1&0&1&0&1 \\ 1&1&0&1&0 \end{pmatrix}.$$

Case where p is arbitrary and $q = 1$ The same functions derived in the previous case also work, because Eqn. (12) remains null on the same values of a:

$$\begin{aligned}\widehat{\mathrm{HW}^p}(a) &= \sum_{z \in GF(2)^n} \frac{1}{2^p} \left(\sum_{i=1}^{n} 1 - \sum_{i=1}^{n} (-1)^{z_i} \right)^p (-1)^{a \cdot z} \\ &= \sum_{z \in GF(2)^n} \frac{1}{2^p} \left(\sum_{j=0}^{p} \binom{p}{j} n^{p-j} \left(- \sum_{i=1}^{n} (-1)^{z_i} \right)^j \right) (-1)^{a \cdot z} . \end{aligned} \tag{15}$$

In general, it is difficult to say more. But if we consider $p = 2$, then $(\sum_{i=1}^{n}(-1)^{z_i})^j$ either:

- is a constant (when $j = 0$), which contributes a non-zero value only if $a = 0$, or
- is identical as in previous Eqn. (12) (when $j = 1$), which contributes a non-zero value only if $\exists i, a = e_i$, or
- implies terms $(-1)^{z_i \oplus z_j}$ (when $j = 2$), which contributes a non-zero value only if $\exists (i, j)$ such that $a = e_i \oplus e_j$.

Therefore, the constraints on $\widehat{\mathrm{HW} \circ B}(a)$ are more stringent. This quantity must be equal to zero on all those n^2 values for a:

- the n vectors $e_{i, i \in [\![0,n[\![}$, as previously when $p = q = 1$, but also
- the $\frac{n(n-1)}{2}$ vectors $e_i \oplus e_j$, for $i \neq j$, in which exactly two bits are set.

However, in special cases, such as the matrix $A8$ involved in Eqn. (14), the lines have 5 ones. Thus, the $n^2 = 8^2$ conditions are met.

Actually, because of the special shape of the matrix defined in Eqn. (14), the functions $fopt_{B:z \mapsto zA8^\intercal, p, q=1}$ are constant for $p \in \{0, 1, 2, 3, 4\}$. Starting from $p = 5$, some combinations of basis vectors e_i will match one line of the matrix, therefore breaking the non-intersection of the set where $\widehat{\mathrm{HW}}(a)$ is null and the set where $\widehat{\mathrm{HW} \circ B}(a)$ is null (outside $a = 0$).

Now, the other way around, for the special matrix of Eqn. (14), we also have the condition that $fopt_{B:z \mapsto zA8^\intercal, p=1, q=2}$ is constant, because:

- none of the rows are equal to one e_i (indeed, the Hamming weight of the rows is 5, whereas e_i is of Hamming weight 1),
- none of the XOR between two arbitrary rows is equal to e_i.

This property is interesting in terms of security, because it shows that using the matrix defined in Eqn. (14), the squeezing leakage resists also 3rd-order CPA attacks.

General Case Theorem 3 can be extended to any (p, q). Eqn. (10) is satisfied if and only if the vectorial space spanned by any p lines of I_n intersects the vectorial space spanned by any q lines of B only in 0. As an example, this extension shows that $\overline{I_4}$ actually satisfies Eqn. (10) for any $p + q < 3$, thus reaching $\mathsf{HCI} = 4$. This is illustrated for $n = 4$ in Tab. 1 the last group.

Number of masks (d)	No LS	LS
$d = 1$	$\mathsf{HCI} = 2$	$\mathsf{HCI} \geq 3^{\dagger}$
$d = 2$	$\mathsf{HCI} = 3$	$\mathsf{HCI} \geq 4$
$d > 2$	$d + 1$	$\geq d + 2$

†: The value $\mathsf{HCI} = 4$ is obtained for matrix $\bar{I}_4$, as illustrated in Eqn. (13).

Table 2. Values of HCI obtained for plain leakage squeezing (LS) countermeasure, *i.e.* that satisfy Theorem 3.

4.4 Security Analysis

Using the linear bijection defined in the previous study, we summarize in Tab. 2 the security improvement brought by the leakage squeezing.

The results in the table show that the leakage squeezing allows to improve the HO-CPA immunity by at least one unit whatever the masking order. In order to confront the theoretical analysis conducted in the previous sections, we performed several attack experiments. Namely, we applied HO-CPA attacks to test the scheme resistance.

The leakage measurements have been simulated as samples of the random variables L_{LS} defined according to Eqn. (8) with $B = \bar{I}_4$. We assume that the leakage is affected by a gaussian noise $N(0, \sigma^2)$. For all the attacks, the sensitive variable Z was chosen to be a DES S-box output of the form $S(X \oplus k)$ where X represents a varying plaintext and k represents the key to recover. We applied the fourth first order CPA attacks such as described in section. 2.1. Theorem 2 leads us to select the Hamming weight function as a prediction function. Each attack simulation was performed 100 times for several noise standard deviation values. As expected, the CPA attacks of order $\{1, 2, 3\}$ were unsuccessful. The success rates stay under 10% even when using up to 10^7 measurements. This confirms the theoretical predictions of table Tab. 2 where the $\mathsf{HCI} = 4$. On the other hand, the 4^{th}-order CPA attack performs well. Tab. 3 summarizes the number of leakage measurements required to observe a success rate of 90% in retrieving k for the different attacks.

As it can be observed in Tab. 3, the 4^{th}-order CPA recovers the key with a success rate equal to 90% when the noise is low. When the noise is high (namely the case of hardware implementation $\sigma \geq 4$), the attack needs more than 10^6 measurements to succeed. For comparison purpose, we perform the 2O-CPA against the first order masking without leakage squeezing. The results are shown in Tab. 3. Based on these results, we see that when the leakage squeezing is applied, not only the order of the

Masking	Attack	σ						
		0	0.25	0.5	1	2	4	8
first-order masking with LS	4O-CPA	3 300	8 000	14 000	30 000	500 000	$> 10^6$	$> 10^7$
first-order masking without LS	2O-CPA	130	220	1 200	5 000	12 000	25 000	200 000

Table 3. Number of leakage measurements for a 90% success rate.

attack raises, but also the attack requires more measurements to succeed. We can conclude that this technique can be a good alternative to classical Boolean masking schemes.

4.5 Information-Theoretic Evaluation of the Countermeasure

In this section, our purpose is to quantify the amount of information that the countermeasure reveals about the sensitive variable Z. To achieve this goal, we follow the information-theoretic approach introduced in [20]. Namely, we compute the mutual information between the sensitive variable Z and the leakage function $L_{LS} + N$, where N is an AWGN of standard deviation σ. In our simulation, we use the bijection $\overline{I_4}$ (called `ni4`). For comparison purpose, we proceed the same for high-order Boolean masking. The mutual information of the leakage squeezing hardware implementation is represented in Fig. 2.

The following observations can be emphasized:

- With or without the leakage squeezing, without noise, the leakage is the same (when using only one mask).
- Without leakage squeezing, the more masks, the more centered moments are balanced. More precisely, with $d > 0$ masks, all the $\mu_{i \in [\![1,d]\!]}$ are identical.
- With the leakage squeezing and the bijection $\overline{I_4}$, the property holds up to $\mathsf{HCI}+2$. Put differently, with one mask, μ_1, μ_2 and μ_3 are balanced.

This first analysis allows us to observe that the gain is high when the leakage squeezing is applied, because the mutual information leaked is less than without the countermeasure whatever the SNR. Our simulations confirm theoretical predictions of Theorem 1. As a corollary, MIM $= \mathcal{O}(1/\sigma^8)$ for first order masking with leakage squeezing ($\mathsf{HCI} = 4$), whereas MIM $= \mathcal{O}(1/\sigma^4)$ for first order masking without ($\mathsf{HCI} = 2$).

Taking advantage of the leakage squeezing principle, the quantity of information leaked with one sole mask is almost the same of the third

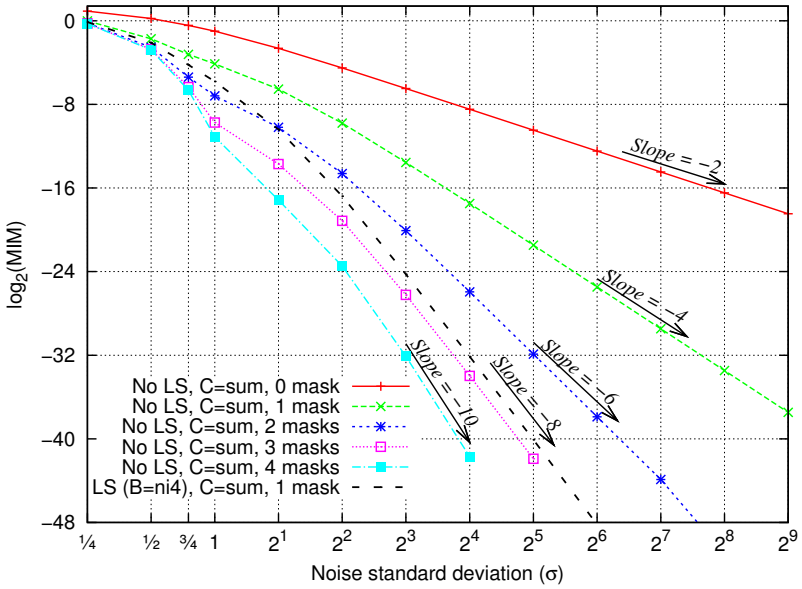

Fig. 2. Leakage metrics for the 4-bit leakage model without and with leakage squeezing (shortened in LS) enhancement, for various number of masks.

order masking without the need of adding extra masks. We conclude that this countermeasure allows to decrease the quantity of the information leaked.

5 Conclusions and Perspectives

In this paper, we have investigated high-order masking countermeasure against side-channel attacks. We have defined the HO-CPA immunity indicator allowing us to assess the resistance against high-order CPA attacks and the amount of leakage. The HO-CPA immunity is equal to the smallest order of a successful optimal HO-CPA attack. Then, we presented a method called leakage squeezing which aims at raising the HO-CPA immunity indicator on hardware masked implementations. This method consists in using bijective encodings which can be implemented in ROMs or LUT networks. Our evaluation analysis shows that this technique provides a great security robustness against HO-CPA. The robustness is corroborated by an information theoretic analysis of the leakage. Indeed, at a given cost and performance level, we show that the leakage squeezing with linear bijections is as efficient as adding one or two other masks. As

a perspective, we intend to extend the formal study of leakage squeezing to non-linear bijections.

References

1. Éric Brier, Christophe Clavier, and Francis Olivier. Correlation Power Analysis with a Leakage Model. In *CHES*, volume 3156 of *LNCS*, pages 16–29. Springer, August 11–13 2004. Cambridge, MA, USA.
2. Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template Attacks. In *CHES*, volume 2523 of *LNCS*, pages 13–28. Springer, August 2002. San Francisco Bay (Redwood City), USA.
3. Jean-Sébastien Coron, Emmanuel Prouff, and Matthieu Rivain. Side Channel Cryptanalysis of a Higher Order Masking Scheme. In *CHES*, volume 4727 of *LNCS*, pages 28–44. Springer, September 10-13 2007. Vienna, Austria.
4. Benedikt Gierlichs, Lejla Batina, Bart Preneel, and Ingrid Verbauwhede. Revisiting Higher-Order DPA Attacks: Multivariate Mutual Information Analysis. In *CT-RSA*, volume 5985 of *LNCS*, pages 221–234. Springer, March 1-5 2010. San Francisco, CA, USA.
5. Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis. In *CHES, 10th International Workshop*, volume 5154 of *LNCS*, pages 426–442. Springer, August 10-13 2008. Washington, D.C., USA.
6. Louis Goubin and Jacques Patarin. DES and Differential Power Analysis. In *CHES*, LNCS, pages 158–172. Springer, Aug 1999. Worcester, MA, USA.
7. ChangKyun Kim, Martin Schläffer, and SangJae Moon. Differential Side Channel Analysis Attacks on FPGA Implementations of ARIA. *ETRI Journal*, 30(2):315–325, 2008. DOI: 10.4218/etrij.08.0107.0167.
8. Thanh-Ha Le and Mael Berthier. Mutual Information Analysis under the View of Higher-Order Statistics. In Isao Echizen, Noboru Kunihiro, and Ryôichi Sasaki, editors, *IWSEC*, volume 6434 of *LNCS*, pages 285–300. Springer, 2010.
9. Yang Li, Kazuo Sakiyama, Lejla Batina, D. Nakatsu, and Kazuo Ohta. Power Variance Analysis breaks a masked ASIC implementation of AES. In *DATE*, pages 1059–1064. IEEE, March 8-12 2010. Dresden, Germany.
10. Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Springer, December 2006. ISBN 0-387-30857-1, http://www.dpabook.org/.
11. Stefan Mangard and Kai Schramm. Pinpointing the Side-Channel Leakage of Masked AES Hardware Implementations. In *CHES*, volume 4249 of *LNCS*, pages 76–90. Springer, October 10-13 2006. Yokohama, Japan.
12. Thomas S. Messerges. Using second-Order Power Analysis to Attack DPA resistant Software. In *CHES*, volume 1965 of *LNCS*, pages 71–77. Springer, August 17-18 2000. Worcester, MA, USA.
13. Emmanuel Prouff and Robert P. McEvoy. First-Order Side-Channel Attacks on the Permutation Tables Countermeasure. In *CHES*, volume 5747 of *LNCS*, pages 81–96. Springer, September 6-9 2009. Lausanne, Switzerland.
14. Emmanuel Prouff and Matthieu Rivain. A generic method for secure sbox implementation. In Sehun Kim, Moti Yung, and Hyung-Woo Lee, editors, *WISA*, volume 4867 of *Lecture Notes in Computer Science*, pages 227–244. Springer, 2007.
15. Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Transactions on Computers*, 58(6):799–811, 2009.

16. Emmanuel Prouff and Thomas Roche. Attack on a Higher-Order Masking of the AES Based on Homographic Functions. In Guang Gong and Kishan Chand Gupta, editors, *INDOCRYPT*, volume 6498 of *LNCS*, pages 262–281. Springer, 2010.
17. Matthieu Rivain and Emmanuel Prouff. Provably Secure Higher-Order Masking of AES. In Stefan Mangard and François-Xavier Standaert, editors, *CHES*, volume 6225 of *Lecture Notes in Computer Science*, pages 413–427. Springer, 2010.
18. Kai Schramm and Christof Paar. Higher Order Masking of the AES. In David Pointcheval, editor, *CT-RSA*, volume 3860 of *LNCS*, pages 208–225. Springer, 2006.
19. François-Xavier Standaert, François Koeune, and Werner Schindler. How to Compare Profiled Side-Channel Attacks? In Springer, editor, *ACNS*, volume 5536 of *LNCS*, pages 485–498, June 2-5 2009. Paris-Rocquencourt, France.
20. François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In *EUROCRYPT*, volume 5479 of *LNCS*, pages 443–461. Springer, April 26-30 2009. Cologne, Germany.
21. François-Xavier Standaert, Gaël Rouvroy, and Jean-Jacques Quisquater. FPGA Implementations of the DES and Triple-DES Masked Against Power Analysis Attacks. In *proceedings of FPL 2006*. IEEE, August 2006. Madrid, Spain.
22. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual Information Analysis: How, When and Why? In *CHES*, volume 5747 of *LNCS*, pages 429–443. Springer, September 6-9 2009. Lausanne, Switzerland.

A Appendix: Some Proofs

A.1 Proof of lemma 1

Proof. $\forall i \in \mathbb{N}$ we have:

$$\begin{aligned}
&\mathrm{cov}((L+N-\mathbb{E}(L+N))^i, Z) = \\
&\mathrm{cov}(\sum_{j=0}^{i} \binom{i}{j} (L-\mathbb{E}(L))^j (N-\mathbb{E}(N))^{i-j}, Z) = \\
&\sum_{j=0}^{i} \binom{i}{j} \mathrm{cov}((L-\mathbb{E}(L))^j (N-\mathbb{E}(N))^{i-j}, Z) \ .
\end{aligned}$$

Now, let X, Y and Z three RV, such that Y is independent of X and Z. We have for all $a, b \in \mathbb{N}$:

$$\begin{aligned}
&\mathrm{cov}(X^a Y^b, Z) = \\
&\mathbb{E}(X^a Y^b Z) - \mathbb{E}(X^a Y^b)\mathbb{E}(Z) = \\
&\mathbb{E}(X^a Z)\mathbb{E}(Y^b) - \mathbb{E}(X^a)\mathbb{E}(Y^b)\mathbb{E}(Z) = \\
&\mathbb{E}(Y^b)\mathrm{cov}(X^a, Z) \ .
\end{aligned}$$

We now apply this result with $X = L - \mathbb{E}(L)$ and $Y = N - \mathbb{E}(N)$. For all $i \in [\![0, \mathsf{HCI}[\![$,

$$\begin{aligned}
&\mathrm{cov}((L + N - \mathbb{E}(L + N))^i, Z) = \\
&\sum_{j=0}^{i} \binom{i}{j} \mathbb{E}(N - \mathbb{E}(N))^{i-j} \mathrm{cov}((L - \mathbb{E}(L))^j, Z) = \\
&\sum_{j=0}^{i} \binom{i}{j} \mathbb{E}(N - \mathbb{E}(N))^{i-j} \times 0 = 0 \ ,
\end{aligned}$$

because, according to our hypothesis, $\mathrm{cov}((L - \mathbb{E}(L))^j, Z) = 0$ for all $j < \mathsf{HCI}$. □

A.2 Proof of Lemma 2

Proof. First of all, we notice that $\forall i \in [\![0, \mathsf{HCI}[\![$, the cumulants $k_i(L \mid Z = z)$ are equal. The reason is that for any law X, $k_j(X)$ can be expressed as a function of $\mu_i(X)$ for $0 \leq i \leq j$ (and reciprocally). For instance $k_3(X) = \mu_3(X)$, $k_4(X) = \mu_4(X) - 3\mu_2^2(X)$, $k_5(X) = \mu_5(X) - 10\mu_3(X)\mu_2(X)$, *etc.* Now, according to definition 1, if L has HO-CPA immunity HCI, then all $\mu_j(L \mid Z = z)$ for $0 \leq i < \mathsf{HCI}$ are independent of z. Consequently, the same holds for the cumulants of orders $i \in [\![0, \mathsf{HCI}[\![$. Eventually, as $\forall i < \mathsf{HCI}$, $\forall z$, $\mu_i(L \mid Z = z) = \mu_i(L)$, we also have $k_i(L \mid Z = z) = k_i(L)$. □