J. Math. Cryptol. **8** (2014), 249–295
DOI 10.1515/jmc-2012-0018

Leakage squeezing: Optimal implementation and security evaluation

Claude Carlet, Jean-Luc Danger, Sylvain Guilley and Houssem Maghrebi

Communicated by Kaoru Kurosawa

Abstract. Hardware devices can be protected against side-channel attacks by introducing one random mask per sensitive variable. The computation throughout is unaltered if the shares (masked variable and mask) are processed concomitantly, in two distinct registers. Nonetheless, this setup can still be attacked if the side-channel is squared, because this operation causes an interference between the two shares. This more sophisticated analysis is referred to as a zero-offset second-order correlation power analysis (CPA) attack. When the device leaks in Hamming distance, the countermeasure can be improved by the "leakage squeezing". It consists in manipulating the mask through a bijection, aimed at reducing the dependency between the shares' leakage. Thus dth-order zero-offset attacks, that consist in applying CPA on the dth power of the centered side-channel traces, can be thwarted for $d \geq 2$ at no extra cost. We denote by n the size in bits of the shares and call F the transformation function, that is, a bijection of $\mathbb{F}_2^n$. In this paper, we explore the functions F that thwart zero-offset high-order CPA (HO-CPA) of maximal order d. We mathematically demonstrate that optimal choices for F relate to optimal binary codes (in the sense of communication theory). First, we exhibit optimal linear F functions. They are suitable for masking schemes where only one mask is used throughout the algorithm. Second, we note that for values of n for which non-linear codes exist with better parameters than linear ones, better protection levels can be obtained. This applies to implementations in which each mask is randomly cast independently of the previous ones. These results are exemplified in the case $n = 8$, where the optimal F can be identified: it is derived from the optimal rate $1/2$ binary code of size $2n$, namely the Nordstrom–Robinson $(16, 256, 6)$ code. This example provides explicitly with the optimal protection that limits to one mask of byte-oriented algorithms such as AES or AES-based SHA-3 candidates. It protects against all zero-offset HO-CPA attacks of order $d \leq 5$. Eventually, the countermeasure is shown to be resilient to imperfect leakage models, where the registers leak differently than the sum of their toggling bits.

Keywords. First-order masking countermeasure (CM), high-order correlation power analysis (HO-CPA), zero-offset HO-CPA, Hamming distance/Hamming weight leakage models, leakage squeezing, linear and non-linear codes.

2010 Mathematics Subject Classification. 94C10.

1 Introduction

Hardware implementations of block-oriented cryptographic functions are vulnerable to side-channel attacks. Yet their lack of algebraic structure makes them hard to protect efficiently. Additive Boolean masking is one answer to secure them, because it can be adapted to any function implemented. Early masking schemes involved only one mask per data needed to be protected [2]. Nonetheless, straightforward implementations of this "first-order" countermeasure (CM) happened to be vulnerable to zero-offset "second-order" attacks [29,45]. We call a "first-order" CM an implementation where one single mask protects the sensitive data. Using more masks per sensitive data yields CM of higher-order, only provided the masking scheme is *sound* [32, §2.3]. Zero-offset attacks use one sample of side-channel trace, and are thus monovariate. They apply when the masked variable and the mask are consumed simultaneously by the implementation, which is commonplace in hardware. Indeed, this architectural strategy allows to keep the throughput unchanged. Zero-offset second-order attacks consider not the plain observations themselves, but their variance instead. The variance of the leakage function, that involves its squaring (second-order moment), does depend strongly on the sensitive data, which allows for an attack. Consequently, a branch of the research on masking CMs has evolved towards masking schemes with multiple masks [35].

Besides, another direction for improvements consists in the adaptation of the first-order CMs to resist attacks that use high-order moments of one single side-channel observation (commonly referred to as zero-offset HO-CPA, of order $d > 1$). Such result can be obtained by transforming the mask before it is latched in register [10]. Concretely, a bijection F is applied to the mask, in a view to reduce the dependency of its leakage with that of the masked data. This optimization of straightforward masking is called "leakage squeezing". It is of no effect in the hypothetical cases when the device completely leaks the shares (identity leakage function); however, it is useful in the realistic cases when the device leaks a non-injective function of the shares (e.g., the Hamming distance leakage model). The goal of this article is to find bijections F that protect against zero-offset attacks of order d as high as possible. This article is based on the results presented in [21], notably by extending the part about the choice of F using the coding theory.

A proof of concept about leakage squeezing has already been studied in [23]. This article provides overhead figures as well as a qualitative estimation of the security gain offered by the CM. An extremely close CM has been introduced recently in [4]: it consists in encoding the mask and mapping the sensitive data to vectors that are not codewords, as if it was noise. In another article a vectorial Boolean function is used to ensure that one share does not leak whereas the other one leaks the perfectly masked sensitive data [24]. This CM is known as a "first-

order leak-free"; its main drawback is that it is not resilient to any leakage model imperfections.

The rest of the paper is structured as follows. In Section 2, the first-order masking scheme that involves the bijection F is described, and its leakage is explained under the Hamming distance model. We also prove that if the masks are refreshed deterministically, then F must be linear. As a positive side effect, the protection now covers not only the Hamming distance leakage model, but also the Hamming weight leakage model. In Section 3, the best zero-offset HO-CPA is derived for all orders d; also, a necessary and sufficient condition on F for the CM to resist all zero-offset HO-CPA of orders $1, 2, \ldots, d$ is formulated. Based on this formal statement of the problem, optimal solutions for F are researched and given in Section 4. The characterization of some optimal bijections F is conducted in Section 5, where both a security analysis against zero-offset HO-CPA and a leakage analysis with an information theoretic metric are conducted. This analysis is carried out both with a perfect and an imperfect leakage model. The conclusions are in Section 6. To ease the reading of the article, some long proofs, some detailed computations, some secondary results (such as the leakage statistical moments) and some simulation graphs (such as the information leakage in the imperfect model) are given in Appendices A–E. The article is self-contained without those appendices; however, they bring interesting insights to support the article's body.

2 Studied implementation and its leakage

2.1 Additive boolean masking

In side-channel analysis, an attacker is able to recover noisy information from a sensitive data, denoted by X. It is of small size (n bit, where n is typically equal to $4, \ldots, 8$), and depends on the secret key. In a masking CM, the leaked information is randomized via the usage of a mask M "entangled" with the functional computation. The two shares manipulated in a Boolean first-order CM are $(X \oplus M, M)$. Provided the mask is uniformly distributed, neither share leaks information about X. Of course, the joint leakage $(X \oplus M, M)$ does depend on X. Concretely speaking, the dependence can be obtained by combining the leakage of $X \oplus M$ on the one hand and M on the other, when the shares are used at different dates. When the shares are manipulated simultaneously, the trace can be raised at the power $d = 2$ to have the shares interfere in the leakage.

It is worthwhile to precise that additive Boolean masking schemes are designed to withstand chosen plaintext attacks. Indeed, each share $X \oplus M$ and X must be uniformly distributed, which is equivalent to saying that the mask M is uni-

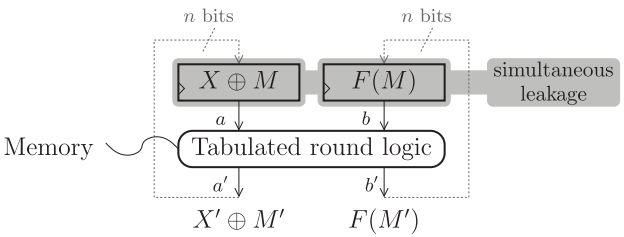

Figure 1. Setup of the first-order masking countermeasure with bijection F.

formly distributed; but no assumption is needed on the distribution of the sensitive variable X. For instance, the security of additive Boolean masking schemes is unaltered if X is for instance fixed to a given value, or somehow biased.

2.2 Leakage squeezing for additive boolean masking

In the "leakage squeezing" CM we study, a bijection F is applied on the mask share. Thus, the shares are now $(X \oplus M, F(M))$. In [23], this CM is called *leakage squeezing*. The schematic of this scheme is illustrated in Figure 1. The variables X and X' are the two consecutive values of the sensitive variable. Similarly, M and M' are the two consecutive values of the mask. This figure highlights two registers able to hold one n-bit word each. The left register hosts the masked data, $X \oplus M$, whereas the register on the right holds $F(M)$, the mask M passed through the bijection F. In this article, we are concerned with the leakage from those two registers only. Indeed, they are undoubtedly the resource that leaks the most. Also, the rest of the logic can be advantageously hidden in tables, thereby limiting their side-channel leakage [36]. It is referred to as "tabulated round logic" in Figure 1. This figure provides an abstract description of the round, since it usually splits nicely into independent datapaths of smaller bitwidth. Typically, an AES can be pipelined to manipulate only bytes (refer to Section 2.5 for a description of the architecture). However, in practice, article [26] (resp. [33]) shows how to handle AES substitution box with 4-bit (resp. 2-bit) non-linear data transformations.

The computation of the bijection F shall not leak. Actually, F can be merged into memories, hence being totally dissolved. Therefore, the two shares $(X \oplus M, F(M))$ remain manipulated concomitantly only once, namely at the clock rising edge. For the sake of illustration, we provide a typical functionality of this combinational logic hidden in memory. If we denote by C the round function and by R the mask refresh function, then the table implements:

- $a' = C(a \oplus F^{-1}(b)) \oplus R(F^{-1}(b))$ and
- $b' = F(R(F^{-1}(b)))$.

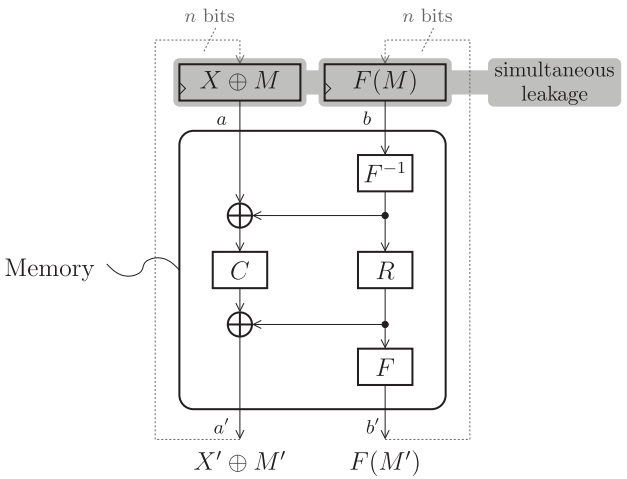

Figure 2. Detail of the function implemented in the tabulated round logic shown in Figure 1.

In the rest of the article, we assume that the mask refresh function R is bijective, so as to maintain the entropy of the mask throughout the computation. The detail of the tabulated round logic is represented in Figure 2. The memory is recomputed statically: at address (a, b), the memory is programmed to return (a', b'), as defined above. So, irrespective of the technology (ASIC, FPGA, etc.) the memory size ($2n \times 2n$) is unchanged. The combinational logic is nonetheless a bit more complex, because the masking must be transformed initially by F and finally by F^{-1}. For example, it is reported in [23] that a regular masking scheme for AES requires 366 adaptative logic modules (ALM) in an Altera Stratix II FPGA, whereas the same architecture with leakage squeezing requires 408 ALMs.

In the context of a side-channel attack against a block cipher, either the first round or the last round is targeted. Thus either the input (plaintext) or the output (ciphertext) is known by the attacker. Hence either X or X' is the sensitive data. We make the assumption that the device leaks in the Hamming distance model. This model is realistic and customarily assumed in the literature related to side-channel analysis [3, 40]. Therefore, the sensitive variable to protect is $X \oplus X'$, denoted by Z. The leakage of the studied hardware (Figure 1) is thus

$$\begin{aligned} &\mathsf{HD}(X \oplus M, X' \oplus M') + \mathsf{HD}(F(M), F(M')) \\ &\quad = \mathsf{HW}(Z \oplus M \oplus M') + \mathsf{HW}(F(M) \oplus F(M')). \end{aligned} \tag{2.1}$$

In this equation, the Hamming distance operator HD and the Hamming weight operator HW are defined as $\mathsf{HD}(w, w') = \mathsf{HW}(w \oplus w') \doteq \sum_{i=1}^{n} (w \oplus w')_i$. The

function F is a constant bijection that will contribute to increase the security of the CM. In addition, F is a public information which we assume to be known by any attacker. It is worthwhile noting that the leakage squeezing CM requires the leakage to be a Hamming distance, like in (2.1), since the optimal bijections will be built based on this assumption. Besides, we insist that it is sane to know, to some extent, how the hardware behaves. Indeed, if the hardware is totally untrusted, no CM can be proved secure. For instance, a hardware Trojan that logs the demasked sensitive data could well be hidden in the circuit. This Trojan will then leak the sensitive information through a functional channel, thereby breaking the overall security. In this context, assuming a Hamming distance model is a minor hypothesis:

- for an ASIC designer, for instance, it is easy to build a circuit that behaves as the model, and
- for an FPGA designer, the CM can be programmed, then potential unbalances are profiled and eventually patched (and this process can be repeated iteratively until the leakage function is as expected).

2.3 Hamming distance vs Hamming weight leakage models

A recent experimental paper [27] has shown that some devices, such as FPGAs, leak mainly in the Hamming distance model, but also – albeit to a lesser extent – in the Hamming weight model. This observation is used to effectively defeat one implementation protected by a masking scheme [24] that assumes a leakage in distance (such as the Hamming distance) but not a leakage in values (such as the Hamming weight). One can thus consider that it is relevant to protect also an implementation against attacks that exploit the Hamming weight. It happens that the leakage squeezing protects as well in both Hamming distance and Hamming weight contexts, if the bijection F is linear. Indeed, in this case, (2.1) is equal to

$$\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(F(M'')) \tag{2.2}$$

(with $M'' \doteq M \oplus M'$). This leakage model has the same expression as the Hamming weight leakage at the input, namely $\mathsf{HW}(X \oplus M) + \mathsf{HW}(F(M))$, and as the Hamming weight leakage at the output, namely $\mathsf{HW}(X' \oplus M') + \mathsf{HW}(F(M'))$; simply, the *dummy variable name* Z is assigned to either the *value* X or X', and M'' to either M or M'. In other words, the expression in (2.2) is an *equation* and not *an equality between values*.

Still, it must be ensured that M, M' and M'' are uniformly distributed. The mask M is indeed chosen uniformly at random at every new encryption. The mask $M' = R(M)$ is also uniformly distributed provided the mask refresh function

R is surjective, i.e. bijective, given the equal cardinality of R input and output sets. The difference of masks $M'' = M \oplus R(M)$ is uniformly distributed if and only if $I \oplus R$ is bijective. Many possible such functions exist: they are called *orthomorphic permutations* (orthomorphic meaning that $I \oplus R$ is bijective). For instance, the linear function R generated by the matrix

$$\mathfrak{R} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

is an orthomorphic permutation. Concluding, when F is linear, it is mathematically equivalent to say that the implementation resists attacks of order $i \in \mathbb{N}$ in Hamming distance and in Hamming weight. Therefore, provided F is linear and R is an orthomorphic permutation, the leakage squeezing scheme does not need to follow a specific architecture.

In addition, this means that the leakage squeezing used with a linear bijection is able to protect against *any combination* of Hamming weight and Hamming distance model attacks.

2.4 Impact of the countermeasure design on leakage scenarios and implications on the linearity (or not) of F

In the rest of the article, we consider two leakages scenarios.

First scenario: Design with a deterministic mask refresh function R

For the design that has been the running example of Section 2, the mask M' is obtained deterministically from mask M via an orthomorphic permutation R (recall Figure 2). We have seen in the previous Section 2.3 that it is advantageous to take F linear, so as to support leakage models in Hamming weight and distance. This is our first leakage scenario.

Second scenario: Design with a non-deterministic mask refresh mechanism

Generally speaking, it is customary in the implementation of masking schemes to employ different masks each time the sensitive variables are updated. This means

in particular that

- the variable X is masked with M as $X \oplus M$, and
- after the application of the round function C, the random variable $X' = C(X)$ is masked with a new mask M', as $X' \oplus M'$. This new mask M' is preferentially independent from the previous value of the mask M.

Many design options can lead to the independence between M and M'. For instance, it is possible to add in the schematic of Figure 2 a third input that would be M' (independent from M). Alternatively, a heuristic method to obtain a similar effect is to use a tactic explained by Güneysu and Moradi [18]. While sticking at the schematic of Figure 2, the $R : \mathbb{F}_2^n \to \mathbb{F}_2^n$ orthomorphic permutation is *updated* on a frequent basis, with a new one randomly chosen. Technically speaking, this can be achieved for instance by using a double-bank memory instead of only one. While one bank is used for the functional computation, the other bank is reprogrammed with a new fresh orthomorphic permutation R. Now, any other implementation that leaks as per (2.1) is eligible.

In such case, any function F can be used (linear or not). If the scheme is derived from that of Figure 2, then R must always be a orthomorphic permutation.

2.5 Architecture of AES based on byte updates

This section describes how a full-fledged AES can be computed from the "variable update" scheme presented in Figure 1. The term "variable update" (or "byte update" for $n = 8$) is related to the distance computed between one register current value and its future value. AES has been designed to work on machines using bytes (8-bit registers) as well as words (32-bit registers).

On 32-bit machines, the substitution box (S-box, named `SubBytes`) is not computed alone, but already combined with the diffusion layer (linear operation on columns, named `MixColumns`). This demands the computation of so-called T-boxes [14] (there are four of them, one for each column). For example, in encryption mode, the first T-box computes

$$A \in \mathbb{F}_2^8 \mapsto (\texttt{xtime}(A), A, A, \texttt{xtime}(A) \oplus A).$$

Such T-box fits in a memory of 256 words. The `xtime` operation computes the multiplication by X in the field $\mathbb{F}_{2^8} \equiv \mathbb{F}_2[X]/X^8 + X^4 + X^3 + X + 1$.

On 8-bit machines, `SubBytes` and `xtime` are typically computed sequentially, i.e. one after the other. The memory requirement is then only two memories of 256 bytes, at the expense of a slower computation.

So, in particular, the AES block cipher can be computed by a sequence of

(i) byte updates, and

(ii) byte moves (from register to register).

As mentioned, byte updates are operations like

- $X \leftarrow \texttt{SubBytes}(X)$,
- $X \leftarrow \texttt{xtime}(X)$

(cf. the definition in the AES standard [28]). Both operations can be captured by the byte update operation, by setting C to `SubBytes` or `xtime` in Figure 2.

The XOR between two different shared values $(X_1 \oplus M_1, F(M_1))$ and $(X_2 \oplus M_2, F(M_2))$ can be done independently on each share, if F is linear.

Concluding, the setup presented in Figure 1 can serve as building block for a construction of AES.

3 Optimal function in zero-offset dth-order CPA

3.1 Optimal prediction function f_{opt} definition

Prouff et al. have shown in [31] that an attacker can optimize a CPA [3] against a device leaking L by computing the correlation between the random variables L and $f_{\text{opt}}(Z)$, where Z is the sensitive variable. The function $f_{\text{opt}}(\cdot)$ is called the "optimal prediction function", and is defined as $f_{\text{opt}}(z) = \mathbb{E}[L - \mathbb{E}[L] \mid Z = z]$. This function depends on the leakage model L which is thus assumed to be known (even imperfectly) by the attacker. Put differently, the optimal prediction function is device-specific. In this definition, the capital letters denote random variables, and $\mathbb{E}$ is the expectation operator. If $z \mapsto f_{\text{opt}}(z)$ is constant (i.e. $f_{\text{opt}}(Z)$ is deterministic), then [31] shows that the correlation coefficient of the attack is null, which means that the attack fails.

This result can be applied on the studied leakage function of (2.1), without F (i.e. with F equal to the identity function Id). The leakage function therefore simplifies in $\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(M'')$, where $M'' \doteq M \oplus M'$ is a uniformly distributed random variable in $\mathbb{F}_2^n$.

- In a zero-offset first-order attack, the attacker uses

$$f_{\text{opt}}(Z) = \mathbb{E}\big[\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(M'') - n \mid Z\big] = 0,$$

which is deterministic,

- whereas in a zero-offset second-order attack, the attacker uses

$$f_{\text{opt}}(Z) = \mathbb{E}\big[(\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(M'') - n)^2 \mid Z\big] = n - \mathsf{HW}(Z),$$

which depends on Z. This result is easily obtained by developing the square. The only non-trivial term in this computation is $\mathbb{E}[\mathsf{HW}(z \oplus M'') \times \mathsf{HW}(M'')]$, which is proved to be equal to $\frac{n^2+n}{4} - \frac{1}{2}\mathsf{HW}(z)$ in [31, (19)].

In summary, without F, a first-order attack is thwarted, but a second-order zero-offset attack will succeed. In the sequel, when mentioning HO-CPA attacks, we implicitly mean "zero-offset HO-CPA", i.e. a mono-variate attack that uses a high-order moment of the traces instead of the raw traces. Nonetheless, as explained in [45], this second-order attack requires more traces than a first-order attack on an unprotected version that do not use any mask. Indeed, the noise is squared and thus its effect is exacerbated. More generally, the higher the order d of a HO-CPA attack, the greater the impact of the noise. Thus, attacks are still possible for small d, but get more and more difficult when d increases. Therefore, our objective is to improve the masking CM so that the zero-offset HO-CPA fails for orders $[\![1, d]\!]$, with d being as high as possible. This translates in terms of $f_{\text{opt}}(Z)$ by having

$$\mathbb{E}\big[(\mathsf{HW}(Z \oplus M \oplus M') + \mathsf{HW}(F(M) \oplus F(M')) - n)^d \mid Z\big]$$

deterministic (i.e. independent of random variable Z) for the highest possible values of the integer d. Thus, when developing the sum raised at the power d, we are led to study terms of the form

$$\mathsf{Term}[p, q](f_{\text{opt}})(z) \doteq \mathbb{E}\big[\mathsf{HW}^p(z \oplus M \oplus M') \times \mathsf{HW}^q(F(M) \oplus F(M'))\big], \quad (3.1)$$

where p and q are two positive integers. If either p or q is null, then trivially, $\mathsf{Term}[p, q](f_{\text{opt}})$ is constant. We are thus interested more specifically in p and q values that are strictly positive. We note that in order to resist dth order zero-offset HO-CPA, $\mathsf{Term}[p, q](f_{\text{opt}})(z)$ must not depend on z for all p and q that satisfy $p + q \leq d$.

Equation (3.1) can be rewritten as follows:

$$\mathsf{Term}[p, q](f_{\text{opt}})(z) = \mathbb{E}\big[\mathsf{HW}^p(z \oplus M'') \times \mathsf{HW}^q(F(M) \oplus F(M \oplus M''))\big]. \quad (3.2)$$

Indeed:

(i) In the first scenario of Section 2.4, $M' = R(M)$ and F is linear, so $M'' = (I \oplus R)(M)$ is a random variable that is balanced (owing to the orthomorphicity of R and to the balancedness of M). The expectation is taken only on M'', since M cancels due to the linearity of F: $F(M) \oplus F(M \oplus M'') = F(M'')$.

(ii) In the second scenario, M and M' are independent and uniformly distributed, hence so is $M'' = M \oplus M'$.

Incidentally, we notice that the same condition would hold if the two shares were

- $F_0(X \oplus M)$ and $F_1(M)$, where F_0 and F_1 are two bijections, with F_0 linear,
- instead of merely $X \oplus M$ and $F(M)$, as in Figure 2.

This new setting is more general, since by choosing $F_0 = \mathsf{Id}$ (linear bijection) and $F_1 = F$ (arbitrary bijection), it comes down to that of Figure 2. Because F_0 is linear, the generalization of (3.2) is

$$\begin{aligned}
&\mathbb{E}\big[\mathsf{HW}^p(F_0(z \oplus M \oplus M')) \times \mathsf{HW}^q(F_1(M) \oplus F_1(M'))\big] \\
&\quad = \mathbb{E}\big[\mathsf{HW}^p(F_0(z) \oplus F_0(M) \oplus F_0(M')) \times \mathsf{HW}^q(F_1(M) \oplus F_1(M'))\big] \\
&\quad = \mathbb{E}\big[\mathsf{HW}^p(\tilde{z} \oplus \tilde{M} \oplus \tilde{M}') \times \mathsf{HW}^q\big(F_1(F_0^{-1}(\tilde{M})) \oplus F_1(F_0^{-1}(\tilde{M}'))\big)\big], \quad (3.3)
\end{aligned}$$

where $\tilde{z} \doteq F_0(z)$, $\tilde{M} \doteq F_0(M)$ and $\tilde{M''} \doteq F_0(M'')$. The random variable $\tilde{M}$ (resp. $\tilde{M''}$) is uniformly distributed, because M (resp. M'') is also uniformly distributed and F_0 (resp. F_1) is a bijection. Thus, the more general setting is secure if (3.3) does not depend on $\tilde{z}$ for all $p + q \leq d$, which is equivalent to having the setting of Figure 2 secure with $F = F_1 \circ F_0^{-1}$. For this reason, we simply consider in the sequel that only one bijection is applied to the mask – the masked sensitive data being manipulated plain.

3.2 Sequential leakage

If the shares $Z \oplus M$ and $F(M)$ are manipulated at different dates (i.e. not simultaneously as in Figure 2), then the attacker could typically attempt to combine their leakage. The paper [31] precisely covers this topic: it proves that the best combination tool is the centered product. Thus, the attacker's strategy remains in line with that discussed on parallel leakage: terms such as $\mathsf{Term}[p, q](f_{\text{opt}})(z)$ (recall their definition in (3.2)) are checked for dependence in z. So, the results discussed in this paper also apply to "software" implementations that handle the shares sequentially. Indeed, the central processing unit (CPU) executes the cryptographic code with masking instruction by instruction, and thus processes one share after the other.

3.3 Condition on F for the resistance against 2nd-order CPA

In a zero-offset second-order CPA, the attacker squares the side-channel leakage of (2.1) and considers its correlation with Z. Thus, to resist this attack, the term in (3.2) must be constant for $p + q \leq 2$. As just mentioned, the cases $(p, q) = (2, 0)$ and $(0, 2)$ are trivial. This subsection thus focuses on the case where $p = q = 1$.

The term $F(m) \oplus F(m \oplus m'')$ is also known as the value at m of the derivative of F in the direction m'', and is denoted by $D_{m''}F(m)$. This notion is for instance defined in [7, §8.2.2, Definition 8.2]. When F is linear (first scenario presented in Section 2.4), we have $D_{m''}F(m) = F(m'')$ irrespective of m. It can be observed that (3.2) can also be written as a convolution product:

$$\mathsf{Term}[p,q](f_{\mathrm{opt}})(z) = \frac{1}{2^n}\big(\mathsf{HW} \otimes \mathbb{E}[\mathsf{HW}(D_{(\cdot)}F(M))]\big)(z).$$

In this expression, $\mathbb{E}[\mathsf{HW}(D_{(\cdot)}F(M))]$ designates the function

$$\mathbb{E}[\mathsf{HW}(D_{(\cdot)}F(M))] : \mathbb{F}_2^n \to \mathbb{Z},$$

$$m'' \mapsto \mathbb{E}[\mathsf{HW}(D_{m''}F(M))] = \frac{1}{2^n}\sum_m \mathsf{HW}(D_{m''}F(m)).$$

The Fourier transform of a function $f : \mathbb{F}_2^n \to \mathbb{Z}$ is defined as

$$\hat{f} : \mathbb{F}_2^n \to \mathbb{Z}, \quad z \mapsto \sum_{y \in \mathbb{F}_2^n} f(y)(-1)^{y \cdot z}.$$

An appealing property of this Fourier transform is that it turns a convolution into a product. So, we have[1]:

$$\begin{aligned} f_{\mathrm{opt}}(z) = \mathrm{cst} &\iff \widehat{f_{\mathrm{opt}}}(a) \propto \delta(a) \\ &\iff \widehat{\mathsf{HW}}(a) \times \widehat{\mathbb{E}[\mathsf{HW} \circ D_{(\cdot)}F(M)]}(a) = (n \times 2^{n-1})^2 \times \delta(a) \\ &\iff \forall a \neq 0, \widehat{\mathsf{HW}}(a) = 0 \text{ or } \widehat{\mathbb{E}[\mathsf{HW} \circ D_{(\cdot)}F(M)]}(a) = 0, \quad (3.4) \end{aligned}$$

where $\propto$ means "is proportional to" and $\delta(\cdot)$ is the Kronecker symbol.

To prove the second line, we note that on the one hand

$$\widehat{\mathsf{HW}}(0) = \sum_z \mathsf{HW}(z) \cdot (-1)^{0 \cdot z} = \frac{n}{2}2^n$$

and on the other hand

$$\begin{aligned} \widehat{\mathbb{E}[\mathsf{HW} \circ D_{(\cdot)}F(M)]}(0) &= \sum_z \mathbb{E}[\mathsf{HW}(D_z F(M))(-1)^{0 \cdot z}] \\ &= \mathbb{E}\Big[\sum_z \mathsf{HW}(F(M) \oplus F(M \oplus z))\Big] \\ &= \mathbb{E}\Big[\sum_{z'} \mathsf{HW}(z')\Big] = \mathbb{E}\Big[\frac{n}{2}2^n\Big] = \frac{n}{2}2^n. \end{aligned}$$

The third equality holds because $z \mapsto F(m) \oplus F(m \oplus z)$ is bijective for all m.

[1] The letters a and b used in the sequel are elements of $\mathbb{F}_2^n$ that do not represent the input ports of the ROM in Figure 1 and 2.

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

$$\begin{aligned}
\widehat{\mathsf{HW}}(a) &= \sum_z \frac{1}{2} \sum_{i=1}^{n} \big(1 - (-1)^{z_i}\big)(-1)^{a \cdot z} \\
&= n \cdot 2^{n-1} \delta(a) - \frac{1}{2} \sum_z \sum_{i=1}^{n} (-1)^{(a \oplus e_i) \cdot z} \\
&= \begin{cases} n \cdot 2^{n-1} & \text{if } a = 0, \\ -2^{n-1} & \text{if there exists } i \in [\![1, n]\!] \text{ such that } a = e_i, \\ 0 & \text{otherwise.} \end{cases} \qquad (3.5)
\end{aligned}$$

Thus, the problem comes down to finding a function F such that

$$\widehat{\mathbb{E}[\mathsf{HW} \circ D_{(\cdot)} F(M)]}(a) = 0 \quad \text{for all } a = e_i .$$

This condition can be rewritten as

$$\sum_{z,m} \mathsf{HW}\big(F(m) \oplus F(m \oplus z)\big)(-1)^{a \cdot z} = 0 \quad \text{for all } a = e_i . \qquad (3.6)$$

Let $a \neq 0$. Then

$$\begin{aligned}
&\sum_{z,m} \mathsf{HW}\big(F(m) \oplus F(m \oplus z)\big)(-1)^{a \cdot z} \\
&= \sum_{z,m} \frac{1}{2} \sum_{i=1}^{n} \big(1 - (-1)^{F_i(m) \oplus F_i(m \oplus z)}\big)(-1)^{a \cdot z} \\
&= \cancel{n 2^{2n-1} \delta(a)} - \frac{1}{2} \sum_{i=1}^{n} \sum_{z,m} (-1)^{F_i(m) \oplus F_i(m \oplus z) \oplus a \cdot z} \\
&= -\frac{1}{2} \sum_{i=1}^{n} \sum_{m} (-1)^{F_i(m)} \sum_{z} (-1)^{a \cdot z \oplus F_i(m \oplus z)} \\
&= -\frac{1}{2} \sum_{i=1}^{n} \sum_{m} (-1)^{F_i(m)} \sum_{z} (-1)^{a \cdot (z \oplus m) \oplus F_i(z)} \qquad (z \leftarrow z \oplus m) \\
&= -\frac{1}{2} \sum_{i=1}^{n} \sum_{m} (-1)^{a \cdot m \oplus F_i(m)} \sum_{z} (-1)^{a \cdot z \oplus F_i(z)} \\
&= -\frac{1}{2} \sum_{i=1}^{n} \Big(\sum_{m} (-1)^{a \cdot m \oplus F_i(m)} \Big)^2 = -\frac{1}{2} \sum_{i=1}^{n} \big(\widehat{(-1)^{F_i}}(a)\big)^2 .
\end{aligned}$$

Thus, this quantity is null if and only if

$$\widehat{(-1)^{F_i}}(a) = 0 \quad \text{for all } i \in [\![1, n]\!].$$

Thus, if we generalize the Fourier transform on vectorial Boolean functions (by applying the transformation component-wise), and use the notation f_χ for the sign function of f (also component-wise), then (3.6) is equivalent to $\widehat{F_\chi}(a) = 0$ for all $a = e_i$. The Fourier transform of a sign function is also known as the Walsh–Hadamard transform. (Both notions are linked through the relationship $\widehat{F_\chi}(a) = 2^n\delta(a) - 2\widehat{F}(a)$ for all a.) Now, as F is balanced (since bijective), $\widehat{F_\chi}(a) = 0$ also holds for $a = 0$. By definition, a Boolean function g is d-resilient if its Walsh–Hadamard transform $\widehat{g_\chi}(a)$ is null for all a such as $\mathsf{HW}(a) \leq d$. Thus every coordinate of F is 1-resilient. Constructions for such functions exist, as explained in [6, Section 8.7].

In the next subsection, we use P-resilient functions F: according to the definition, they are balanced when up to P input bits are fixed.

3.4 Condition on F for the resistance against dth-order CPA

A generalization of the previous result for arbitrary $p, q \in \mathbb{N}^\star \doteq \mathbb{N}\backslash\{0\}$ is presented in this section. We have the following theorem, whose proof is given in Appendix A.

Theorem 3.1. *Let P and Q be two positive integers, and F a bijection of $\mathbb{F}_2^n$. Equation* (3.2) *is constant for all $p \in [\![0, P]\!]$ and $q \in [\![0, Q]\!]$ if and only if*

$$\widehat{(b \cdot F)_\chi}(a) = 0 \tag{3.7}$$

whatever $a, b \in \mathbb{F}_2^n$ with $0 < \mathsf{HW}(a) \leq P$ and $0 \leq \mathsf{HW}(b) \leq Q$.

An (n, m)-function is defined as a vectorial Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$.

Proposition 3.2. *The condition expressed in* (3.7) *of Theorem* 3.1 *can be reformulated as follows. Every restriction of the bijective (n, n)-function F to Q components is an (n, Q)-function that is P-resilient.*

4 Existence of bijections meeting the condition expressed in (3.7)

In this section, we find bijections that meet (3.7).

4.1 Three conditions on optimal bijections F

Condition in terms of Walsh–Hadamard transform

The condition expressed in (3.7) can be rewritten as follows:

$$\widehat{(b \cdot F)}_{\chi}(a) = 0 \tag{4.1}$$

for all $b \in \mathbb{F}_2^{n\star} \doteq \mathbb{F}_2^n \backslash \{0\}$ and $a \in \mathbb{F}_2^n$ with $\mathsf{HW}(a) \leq d - \mathsf{HW}(b)$.

Condition in terms of correlation-immunity

Given any (n,n)-function F, let $C \doteq \{(x, F(x));\ x \in \mathbb{F}_2^n\}$ be the graph of F. The indicator 1_C of C is the Boolean function

$$1_C : \xi \in \mathbb{F}_2^{2n} \mapsto \begin{cases} 1 & \text{if } \xi \in C, \\ 0 & \text{otherwise.} \end{cases}$$

The condition on F given in (4.1) is satisfied if and only if the indicator of the graph C of F is dth order correlation-immune (see definition in [5]); this result comes from the characterization of correlation-immune functions by their Fourier transform available in [46].

By the definition of correlation-immune functions, we also have this characterization on F. For all subsets I of $\{1, \ldots, 2n\}$ of cardinality $|I|$ at most d, and for all $a \in \mathbb{F}_2^I$, there are $\frac{|C|}{2^{|I|}}$ codewords of C whose coordinates of indices $i \in I$ coincide with those of a. This means that, by fixing some coordinates of x and some coordinates of $F(x)$, it is impossible to bias $C = \{(x, F(x));\ x \in \mathbb{F}_2^n\}$ if the number of fixed coordinates does not exceed d.

Condition in terms of code

Given any (n,n)-function F, we define the weight enumerator of the code C by

$$W_C(X,Y) \doteq \sum_{x \in \mathbb{F}_2^n} X^{2n - \mathsf{HW}(x,F(x))} Y^{\mathsf{HW}(x,F(x))}$$

and the distance enumerator by

$$D_C(X,Y) \doteq \frac{1}{|C|} \sum_{x,y \in \mathbb{F}_2^n} X^{2n - \mathsf{HW}(x \oplus y, F(x) \oplus F(y))} Y^{\mathsf{HW}(x \oplus y, F(x) \oplus F(y))}.$$

We have

$$W_C(X+Y, X-Y) = \sum_{a,b \in \mathbb{F}_2^n} \Big(\sum_{x \in \mathbb{F}_2^n} (-1)^{b \cdot F(x) + a \cdot x} \Big) X^{2n - \mathsf{HW}(a,b)} Y^{\mathsf{HW}(a,b)}$$

and

$$D_C(X+Y,X-Y)=\frac{1}{|C|}\sum_{a,b\in\mathbb{F}_2^n}\Big(\sum_{x\in\mathbb{F}_2^n}(-1)^{b\cdot F(x)\oplus a\cdot x}\Big)^2 X^{2n-\mathsf{HW}(a,b)}Y^{\mathsf{HW}(a,b)}.$$

Hence $d+1$ is exactly the minimum value of the non-zero exponents of Y with non-zero coefficients in $D_C(X+Y,X-Y)$, called the dual distance of C in the sense of Delsarte [11, 20].

In summary, our goal can be restated as follows: we seek to find a bijection F such that the code C equal to the graph of F has the largest possible dual distance.

4.2 Optimal linear bijections

The bijection F can be chosen linear. This choice suits both design scenarios presented in Section 2.4. All linear (n,n)-functions are of the form

$$F(x)=(x\cdot v_1,\ldots,x\cdot v_n),$$

where v_i are elements of $\mathbb{F}_2^n$. F is bijective if and only if $(v_1,\ldots,v_n)$ is a basis of $\mathbb{F}_2^n$. We have

$$\begin{aligned}
\widehat{(b\cdot F)}_\chi(a)=0 &\iff \sum_x(-1)^{b\cdot F(x)\oplus x\cdot a}=0\\
&\iff \sum_x(-1)^{\bigoplus_{i=1}^n b_i(x\cdot v_i)\oplus x\cdot a}=0\\
&\iff \sum_x(-1)^{x\cdot\bigoplus_{i=1}^n(b_i v_i)\oplus x\cdot a}=0\\
&\iff \sum_x(-1)^{x\cdot(\bigoplus_{i=1}^n(b_i v_i)\oplus a)}=0\\
&\iff \bigoplus_{i=1}^n b_i v_i\neq a.
\end{aligned}$$

As this is true for all a such that $\mathsf{HW}(a)\le d-\mathsf{HW}(b)$, we have the following necessary and sufficient condition for all $b\neq 0$:

$$\mathsf{HW}\Big(\bigoplus_{i=1}^n b_i v_i\Big)>d-\mathsf{HW}(b). \tag{4.2}$$

We notice that the set of ordered pairs

$$C'\doteq\Big\{\Big(b,\bigoplus_{i=1}^n b_i v_i\Big);\ b\in\mathbb{F}_2^n\Big\}$$

forms a vector subspace of $\mathbb{F}_2^{2n}$. Therefore, it defines a $[2n,n,\delta]$ binary linear

code, where δ is its minimum (direct) distance. Because of (4.2), the necessary and sufficient condition becomes merely $\delta > d$.

The codes C and C' have rate $1/2$; and F being bijective, each of these codes admits the two information sets $[\![1, n]\!]$ and $[\![n+1, 2n]\!]$ (recall that a set of indices I is called an information set of a code if every possible tuple occurs in exactly one codeword within the specified coordinates x_i; $i \in I$). More generally, a rate $1/2$ code which admits two complementary information sets is called a "complementary information set" code, or CIS code in short. These codes are studied in [8,9]. From any such linear CIS code, it is possible to deduce a linear bijection F. Indeed, by permuting the coordinates, these two information sets can be respectively available at coordinates of indices $[\![1, n]\!]$ and $[\![n+1, 2n]\!]$ in the codewords. The $[2n, n, \delta]$ binary linear code can thus be spawned by a generator matrix $(A \ B)$, where A and B are two $n \times n$ invertible matrices. A left-hand side multiplication by the inverse of A turns the generic generator matrix into the systematic representation of the code, namely $(I_n \ G)$, where $G \doteq A^{-1} \times B$. This corresponds to a code $\{(x, F(x));\ x \in \mathbb{F}_2^n\}$ where F is bijective because G is invertible (in the general case of $1/2$ rate codes, the systematic representation can also be written as $(I_n \ G)$, but G in not necessarily invertible). It also corresponds to a code $C' = \{(b, \bigoplus_{i=1}^{n} b_i v_i);\ b \in \mathbb{F}_2^n\}$ giving a bijection F.

Now, $[2n, n, \delta]$ binary linear codes have been well studied. They are also referred to as $1/2$-rate codes in the literature. Their greatest minimal distance $\delta_{\max}(n)$ is known for all lengths $2n$ up to 36 (see the BKLC function of Magma [41]). From the condition $\delta > d$, we deduce that the best achievable d using a linear bijection F is $\delta_{\max}(n) - 1$. Consequently, $d \leq n$, and this bound is met if and only if C is *maximum distance separable* (MDS), which is equivalent to saying that F is a multipermutation [42]. However, binary MDS codes exist only if the code dimension is equal to 0, 1, the code length or the code length minus 1. Thus, they do not exist if $n > 1$, hence the bound $d \leq n - 1$.

The greatest minimal distance $\delta_{\max}(n)$ of rate $1/2$ binary linear codes is known (see, e.g., [17]); corresponding codes are called "optimal". For some practical values of n, they are recalled in Table 1. For instance, when $n = 4$, the optimal code is the Hamming code $[8, 4, 4]$, and when $n = 8$, the optimal code is $[16, 8, 5]$, a subcode of the BCH (Bose–Chaudhuri–Hocquenghem) code $[17, 9, 5]$.

In particular, this result proves that with linear F, it is possible to protect[2]

- DES against all zero-offset HO-CPA of order $d \leq 3$, and
- AES against all zero-offset HO-CPA of order $d \leq 4$.

[2] The block cipher MISTY1 is special in that it has in its design both substitution boxes of input size 7 and 9 bits; hence, *at minima*, it protects against all zero-offset HO-CPA attacks of order $d \leq \min\{4, 6\} = 4$, i.e. as good as the weakest substitution box.

Sboxes of Algorithm [19]	**DES, CAST-128, HIGHT**	**n/a**	**n/a**	**MISTY1**	**AES, Camellia, SEED**	**MISTY1**
Value of n	4	5	6	7	8	9
Value of $2n$	8	10	12	14	16	18
Value of $\delta_{\max}(n)$	4	4	4	4	5	6

Table 1. Minimal distance of some binary optimal linear rate $1/2$ codes.

Indeed, there exist optimal linear codes that also enjoy the CIS property; as mentioned in [9], this notion is not trivial, since for instance there exists a $[34, 17, 8]$ code which is not CIS. Now, for $n = 4$, the matrix $G = G3'$ is shown to be invertible in (5.4) (notice that $\overline{I_4}^{-1} = \overline{I_4}$). For $n = 8$, the matrix $G = B4$ is shown to be invertible in Appendix B.

For $n = 4$, the bound $d \leq n - 1$ is met, but not for $n = 8$, since the best $d = 4$ is at distance 3 from $n - 1 = 7$.

We recall that the leakage squeezing (hence the announced security) applies to the leakage model given in (2.1) that corresponds to the leakage of the state register. The rest of the algorithm consists in the round logic that shall be implemented in such a way it does not leak sensitive information. In this article, we suggest to tabulate the round logic (as for instance explained in [13] for AES). By chance, it happens that this strategy leads to efficient implementations in FPGAs. On ASICs, the use of memories in lieu of combinational logic is more costly but definitely secure against glitches (that are known to reduce the security order [25]).

We note that C' is a permuted code of the dual $C^\perp$ of C, obtained by swapping the leftmost half of the codewords (i.e. b) with the rightmost half (i.e. $\bigoplus_{i=1}^{n} b_i v_i$). Indeed, C and C' have the same dimension n, hence $C^\perp$ and C' have the same dimension n, and the scalar product between $(\bigoplus_{i=1}^{n} b_i v_i, b)$ and $(x, F(x))$ is equal to

$$\Big(\bigoplus_{i=1}^{n} b_i v_i\Big) \cdot x \oplus b \cdot F(x) = b \cdot F(x) \oplus b \cdot F(x) = 0.$$

Thus, finding the largest dual distance of C is equivalent to finding the largest minimal distance of C'.

Incidentally, when no bijection is used (as, e.g., for the genuine masking [45]), F is the identity (hence a linear function) and C' is the repetition code. This code is autodual ($C'^\perp = C'$) and furthermore $C = C'$, because the generating matrix $(I_n \ I_n)$ is invariant under left-right halves exchange. Its minimal distance is $\delta = 2$, and thus the maximal resistance order is $d = 2 - 1 = 1$, as expected.

Now, the minimal order d for leakage squeezing with CIS codes (linear or not) is $\delta = 2$. Indeed the distance between two different codewords $(x, F(x))$ and $(y, F(y))$ is $\mathsf{HW}(x \oplus y) + \mathsf{HW}(F(x) \oplus F(y)) \geq 1 + 1 = 2$, because the code has two complementary information sets. As a consequence, the minimal order d for CIS codes is $\delta_{\min} \geq 2$. As it is equal to 2 for the identity, the protection order is exactly at least 1. This worst case for the security is thus attained when the leakage squeezing is not used, which positively motivates for its usage.

4.3 Optimal non-linear bijections

We recall that the bijection F can be chosen non-linear when the leakage model in Hamming distance involves two independent masks M and M' (second scenario of Section 2.4). Under some circumstances, a non-linear bijection F allows to reach better performances. There is no non-linear code for $n = 4$ that has a better dual distance than linear codes of the same length and size, but there are some for $n = 8$. A non-linear optimal code for $n = 8$ is the Nordstrom–Robinson $(16, 256, 6)$ code (that is also CIS, as discussed in details in [9, Example III.4]). With these parameters, this code coincides with Preparata and Kerdock codes [38] and has same minimum distance and dual distance (namely $d = d^{\perp} = 6$). Some codewords, as obtained from Golay code in standard form [15], are listed in Table 2.

It happens that the code cannot be trivially split into two halves that each fill exactly $\mathbb{F}_2^n$. Indeed, if the codewords are partitioned with bits $[\![15, 8]\!]$ on the one hand, and bits $[\![7, 0]\!]$ on the other, then

- $(11111111)_2$ (also denoted by `0xff` in the sequel) is present (at least) twice in the first half (from the high byte of codewords $x = 3$ and $x = 7$), and
- $(00000000)_2$ (also denoted by `0x00` in the sequel) is present (at least) twice in the second half (from the low byte of codewords $x = 0$ and $x = 7$).

We tested all the $\binom{16}{8}$ partitionings. For 2760 of them, the code can be cut into two bijections F_{high} and F_{low} of $\mathbb{F}_2^8$. This means that if $x \in \mathbb{F}_2^8$ denotes the codewords' index in Table 2, the Nordstrom–Robinson $(16, 256, 6)$ code can be written as $F_{\text{high}}(x) \parallel F_{\text{low}}(x)$. The codewords can be reordered according to the first column, so that the code can be rewritten as $x \parallel F_{\text{low}}(F_{\text{high}}^{-1}(x))$; see [9]. So the bijection F can be chosen equal to $F_{\text{low}} \circ F_{\text{high}}^{-1}$. For example, when F_{high} consists in bits $[\![15, 9]\!] \cup \{7\}$ of the code (and F_{low} in bits $\{8\} \cup [\![6, 0]\!]$), F takes the values tabulated in Appendix C. Thus byte-oriented cryptographic implementations can be protected with this code against all zero-offset HO-CPA of order $d \leq 5$.

Bit index	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
$x = 0$	0	0	0	0	0	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
$x = 1$	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
$x = 2$	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
$x = 3$	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	1	1	1	1	1	1	1	1
$x = 4$	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
$x = 5$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
$x = 6$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$x = 7$	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
$x = 8$	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x = 254$	1	0	1	1	0	0	1	0	1	0	0	0	0	0	0	1
$x = 255$	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1

Table 2. Some codewords of the Nordstrom–Robinson $(16, 256, 6)$ code.

4.4 Optimality in terms of cost of the leakage squeezing

The leakage squeezing CM can be generalized to any injective (n, m)-function F, where $m \geq n$. The corresponding hardware architecture is depicted in Figure 3. For instance, the first-order leakage-free CM presented in [24] also uses a mask size greater than that of the sensitive variable to protect.

In terms of codes, relaxing F from a bijection to an injection means that codes of rates smaller than $1/2$ are also eligible. For linear codes (i.e. linear F functions), this is tantamount to saying that there exists an information set I such that the restriction of the code to the complement of I is of same dimension as the code.

Unfortunately, this strategy does not bring any improvement. Indeed, codes $\{(x, F(x));\ x \in \mathbb{F}_2^n\} \subseteq \mathbb{F}_2^{n+m}$ can have a greater minimal distance when m increases (for instance by padding the code with new columns), but in the meantime their dual distance decreases. Consequently the cost of the CM increases with m, while the security of the masking scheme decreases.

The best situation is thus to have m minimal, i.e. $m = n$. The leakage squeezing CM initially presented (in Figure 2) is thus optimal.

5 Security and leakage evaluations of the optimal linear and non-linear bijections

As argued in [39], the robustness evaluation of a CM encompasses two dimensions: its resistance to specific attacks, and its amount of leakage irrespective of

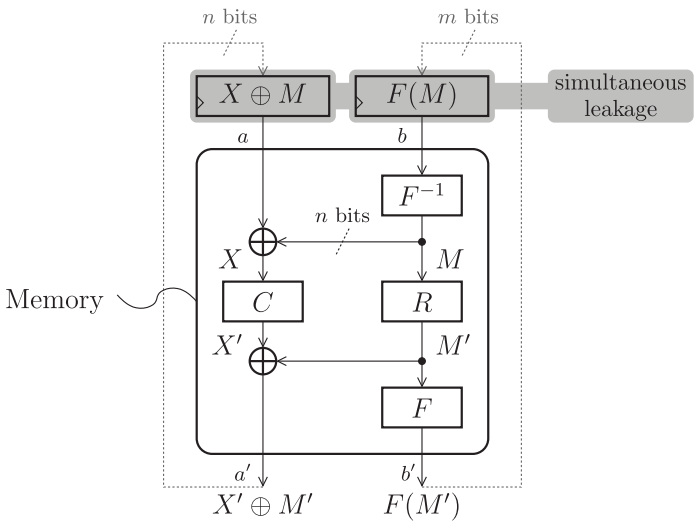

Figure 3. Generalization of the leakage squeezing countermeasure to any injective function $n \to m$.

any attack strategy. Indeed, a CM could resist some attacks, but still be vulnerable to others. For instance, in our study, we have focused on zero-offset HO-CPA, but we have disregarded other attacks, such as mutual information analysis (MIA [1]) or attacks based on generic side-channel distinguishers [44]. Therefore, in addition to a security evaluation conducted in Section 5.1, we will also estimate the leakage of the CM in Section 5.2.

5.1 Verification of the security for $n = 8$

In this section, we illustrate the efficiency of the identified bijection from a zero-offset HO-CPA point of view. We focus more specifically on the $n = 8$ bit case, because of its applicability to AES. We compute the values of $f_{\text{opt}}(z)$ for the centered leakage raised at power $1 \leq d \leq 6$ for four linear bijections (denoted by $F1$, $F2$, $F3$ and $F4$) and the non-linear bijection given in Section 4.3 (denoted by $F5$). The linear functions are defined from their matrix:

- $G1$ is the identity I_8, i.e. the Boolean masking function without F;
- $G2$ is a matrix that allows second-order resistance and is found without method;
- $G3$ is the circulant matrix involved in the AES block cipher;
- $G4$ is non-systematic half of the $[16, 8, 5]$ code matrix (see Appendix B).

The $G2$, $G3$ and $G4$ matrices are:

$$G2 = \begin{pmatrix} 0&0&0&0&0&1&1&1\\ 0&0&0&1&1&0&1&1\\ 1&0&1&0&0&0&1&1\\ 0&0&1&1&1&0&0&0\\ 1&1&1&0&0&0&0&0\\ 1&1&1&1&1&1&1&1\\ 1&0&1&0&1&1&0&0\\ 0&1&0&1&0&1&1&0 \end{pmatrix}, \quad G3 = \begin{pmatrix} 1&0&0&0&1&1&1&1\\ 1&1&0&0&0&1&1&1\\ 1&1&1&0&0&0&1&1\\ 1&1&1&1&0&0&0&1\\ 1&1&1&1&1&0&0&0\\ 0&1&1&1&1&1&0&0\\ 0&0&1&1&1&1&1&0\\ 0&0&0&1&1&1&1&1 \end{pmatrix},$$

$$G4 = \begin{pmatrix} 1&0&0&1&1&1&1&0\\ 0&1&0&0&1&1&1&1\\ 1&1&0&0&1&1&0&0\\ 0&1&1&0&0&1&1&0\\ 0&0&1&1&0&0&1&1\\ 1&1&1&1&0&0&1&0\\ 0&1&1&1&1&0&0&1\\ 1&1&0&1&0&1&1&1 \end{pmatrix}.$$

It can be checked that they are invertible. Namely, their inverses are

$$G2^{-1} = \begin{pmatrix} 0&1&1&1&0&0&0&0\\ 1&1&1&0&0&1&0&0\\ 1&0&0&1&1&1&0&0\\ 0&0&1&0&1&1&1&0\\ 1&0&1&0&0&0&1&0\\ 0&1&0&0&1&1&0&0\\ 1&0&0&0&0&1&1&1\\ 0&1&0&0&1&0&1&1 \end{pmatrix}, \quad G3^{-1} = \begin{pmatrix} 0&0&1&0&0&1&0&1\\ 1&0&0&1&0&0&1&0\\ 0&1&0&0&1&0&0&1\\ 1&0&1&0&0&1&0&0\\ 0&1&0&1&0&0&1&0\\ 0&0&1&0&1&0&0&1\\ 1&0&0&1&0&1&0&0\\ 0&1&0&0&1&0&1&0 \end{pmatrix},$$

$$G4^{-1} = \begin{pmatrix} 1&1&1&0&1&0&1&1\\ 1&0&0&1&1&1&1&0\\ 0&1&0&0&1&1&1&1\\ 1&1&0&0&1&1&0&0\\ 0&1&1&0&0&1&1&0\\ 0&0&1&1&0&0&1&1\\ 1&1&1&1&0&0&1&0\\ 0&1&1&1&1&0&0&1 \end{pmatrix}.$$

Table 4, in Appendix D, reports some values of the optimal functions. The lines represented in gray are those for which the $f_{\text{opt}}(z)$ are the same for all the values of the sensitive variable $z \in \mathbb{F}_2^n$. For the sake of clarity, we represent only $n+1$ values of z, i.e. one per value of $\mathsf{HW}(z)$. But we are aware that unlike in the case where $F = \mathsf{Id}$, the optimal functions are not invariant in the bits reordering of x. If the line d is represented in gray, then a dth order zero-offset HO-CPA cannot succeed. The table shows that amongst the linear functions, $F4 : x \mapsto G4 \times x$ is indeed the best, since it protects against zero-offset HO-CPA of orders 1, 2, 3 and 4. It can also be seen that the non-linear function $F5$ further protects against 5th order zero-offset HO-CPA, as announced in Section 4.3.

5.2 Verification of the leakage of the identified bijections

As a complement to the security analysis carried out in Section 5.1, the leakage of the CM using the bijections $F1$, $F2$, $F3$, $F4$ and $F5$ is computed. It consists in the mutual information metric (MIM), defined as

$$\mathsf{I}[\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(F(M) \oplus F(M \oplus M'')) - n + N; Z].$$

The random variable N is an additive noise that follows a normal law of variance σ^2. The result of the MIM computation is shown in Figure 4. In the ordinate, the smaller the MIM, the more secure the CM. Now, there are at least two ways to interpret the abscissa:

(i) *In terms of attacker budget:* an attacker who is able to develop advanced denoising filters and who can buy accurate side-channel probes will be placed in the low noise areas (i.e. at the left-hand side of the graph).

(ii) *In terms of defender budget:* the designer can integrate more logic so as to increase the algorithmic noise, or he can even add artificial noise sources [18]; but the more noise the designer wishes to inject in a view to obscure the leakage (i.e. at the right-hand side of the graph), the more area and power are required.

It appears that the leakage agrees with the strength of the CM against HO-CPA: the greater the order of resistance against HO-CPA, the smaller the mutual information, at least for a reasonably large noise $\sigma \geq 1$. This simulated characterization validates (in the particular scheme of Figure 2) the relevance of choosing F based on a HO-CPA criterion. The explanation for this observation is given in [22, §2.3]. Basically, the idea is that in the MIM expression, the mutual information can be decomposed as a Taylor series in the cumulants of the conditional leakage that are equal to zero if and only if an HO-CPA does not succeed.

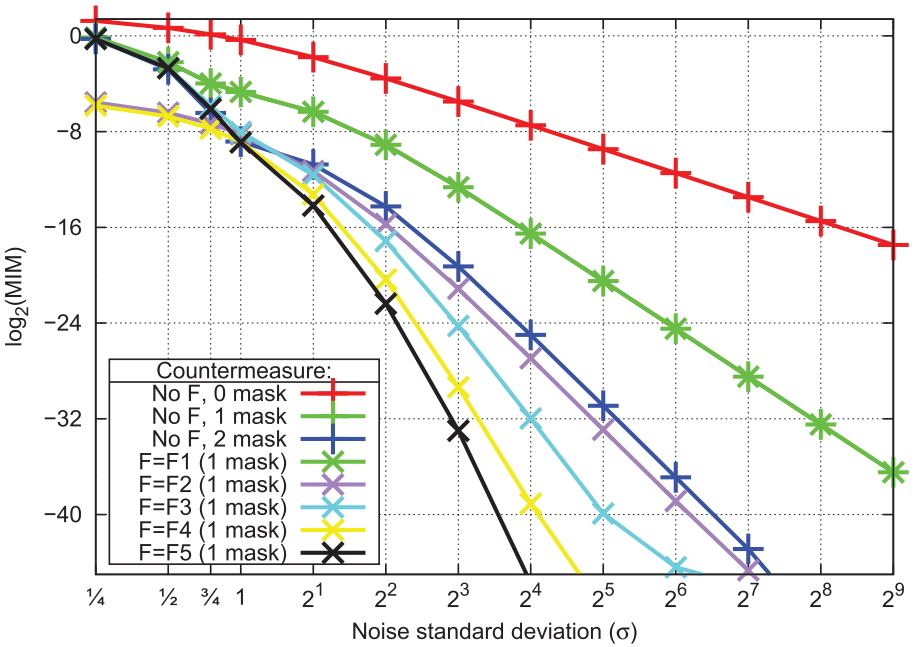

Figure 4. Mutual information of the leakage with the sensitive variable Z for $n = 8$ bit.

Furthermore, Figure 4 represents the leakage of a similar CM, where more than two shares would be used. More precisely, the shares would be the triple $(x \oplus m_1 \oplus m_2, m_1, m_2)$, where the independent masks m_i are not transformed by bijections. This CM is obviously more costly than our proposal of keeping one single mask, but passed through F. We notice that all the proposed bijections (suboptimal $F2$ and $F3$, optimal linear $F4$ and optimal non-linear $F5$) perform better, in that they leak less irrespective of σ. Therefore, in the context of the Hamming distance leakage model, the leakage squeezing is proven to have both a smaller implementation and a smaller leakage than the dth-order masking scheme that makes use of d distinct shares.

5.3 Results in imperfect models

Masking schemes randomize more or less properly the leakage. In the straightforward example studied in this paper (equation (2.1) with $F = \mathsf{Id}$), when the sensitive variable z has all its bits equal to '1' (i.e. $Z = \texttt{0xff}$), then the mask has no effect whatsoever on the leakage. Indeed, this is due to a well-known property

of the Hamming weight function:

$$\mathsf{HW}(\texttt{0xff} \oplus M'') + \mathsf{HW}(M'') = \mathsf{HW}(\overline{M''}) + \mathsf{HW}(M'') = n$$

for all $M'' \in \mathbb{F}_2^n$. To avoid this situation, the proposed CM based on the bijection F consists in tuning the leakage, so that the masks indeed dispatch randomly the leakage for most (if not all [24]) values of the sensitive data. The working factor of this improvement is the introduction of a specially crafted Boolean function F aiming at weakening the link between the data to protect and the leakage function.

This technique has been shown to be very effective in the previous sections. Now, the analysis assumes a perfect leakage model. But the Hamming distance leakage model is in practice an idealization of the reality. Indeed, the assumption that all the bits leak identically, and without interfering, does not hold in real hardware [43]. Also, it has been shown that with specific side-channel capturing systems the attacker can distort the measurement. For instance, in [30], Peeters, Standaert and Quisquater show that with a home-made magnetic coil probing the circuit at a crucial location, the rising edges can be forced to dissipate 17% more than the falling edges.

Therefore, we study how the CM is resilient to imperfections of the leakage model. To do so, we define a general model that depends on random variables. The variability is quantified in units of the side-channel dissipation of a bit-flip. The model is affected by small imperfections (due to process variation, or small cross-coupling) when the variability is about 10%. We also consider the 20% case, that would reflect a distortion of the leakage due to measurements in weird conditions. Eventually, the cases of a 50% and of a 100% deviation indicate that the designer has few or no a priori knowledge about the device leakage's model.

More precisely, the leakage model is written as a multivariate polynomial in $\mathbb{R}[X_1, \ldots, X_n, X'_1, \ldots, X'_n]$ of degree less than or equal to $\tau \in [\![1, 2n]\!]$, where $X = (X_{i \in [\![1,n]\!]})$ and $X' = (X'_{i \in [\![1,n]\!]})$ are the initial and final values of the sensitive variable. It takes the form

$$L \doteq P(X_1, \ldots, X_n, X'_1, \ldots, X'_n) = \sum_{\substack{(u,v) \in \mathbb{F}_2^n \times \mathbb{F}_2^n, \\ \mathsf{HW}(u) + \mathsf{HW}(v) \leq \tau}} A_{(u,v)} \cdot \prod_{i=1}^{n} X_i^{u_i} X'^{v_i}_i, \quad (5.1)$$

where the $A_{(u,v)}$ are real coefficients. This leakage formulation is similar to that of the high-order stochastic model [34]. For example, it is shown in [31, (3)] that $P(X_1, \ldots, X_n, X'_1, \ldots, X'_n)$ is equal to $\mathsf{HW}(X \oplus X')$ when the coefficients

$A_{(u,v)} \doteq a^{\mathrm{HD}}_{(u,v)}$ satisfy

$$a^{\mathrm{HD}}_{(u,v)} = \begin{cases} +1 & \text{if } \mathsf{HW}(u) + \mathsf{HW}(v) = 1, \\ -2 & \text{if } \mathsf{HW}(u) = 1 \text{ and } v = u, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

This property is derived from the equality

$$\mathsf{HW}(X \oplus X') = \mathsf{HW}(X) + \mathsf{HW}(X') - 2\mathsf{HW}(X \wedge X').$$

In the following experiments, we compute the mutual information between L and $Z = X \oplus X'$ when $\tau \leq 1, 2, 3$ and when the coefficients $A_{(u,v)}$ deviate randomly from those of (5.2) or are completely random (i.e. deviate from a NULL model). More precisely, the coefficients $A_{(u,v)}$ are respectively drawn at random from one of these laws:

$$A^{\mathrm{HD}}_{(u,v)} \sim a^{\mathrm{HD}}_{(u,v)} + \mathcal{U}\Big(\Big[-\frac{\delta}{2}, +\frac{\delta}{2}\Big]\Big), \quad A^{\mathrm{NULL}}_{(u,v)} \sim 0 + \mathcal{U}\Big(\Big[-\frac{\delta}{2}, +\frac{\delta}{2}\Big]\Big). \tag{5.3}$$

The randomness lays in the uniform law $\mathcal{U}([-\frac{\delta}{2}, +\frac{\delta}{2}])$ that we parametrize by the deviation $\delta \in \{0.1, 0.2, 0.5, 1.0\}$. The exact choice of the random law is actually irrelevant for our simulations; rather, the variance of the law ($\delta^2/12$ in our case[3]) is interesting as it quantifies the amount of imperfection. The mutual information $\mathsf{I}[L; Z]$ is computed ten times for ten different randomized models. Four bit variables (case useful for DES) are considered because the computation time for the MI would have been too long for $n = 8$. The study is conducted on three bijections:

$F1'$: the identity (Id) that acts as a reference,

$F2'$: one bijection that cancels the first-order leakage but not the second-order,

$F3'$: another bijection that cancels both first- and second-orders.

They are linear, that is, we can write $Fi'(x) = Gi' \times x$, where the generating

[3] The variance of a uniform law of amplitude δ is indeed equal to $\mathsf{Var}(\mathcal{U}([-\delta/2, +\delta/2])) = \frac{1}{\delta}\int_{-\delta/2}^{+\delta/2}(u-0)^2\,\mathrm{d}u = \big[\frac{u^3}{3\delta}\big]_{u=-\delta/2}^{u=+\delta/2} = \frac{\delta^2}{12}$.

matrices Gi' are

$$G1' = I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad G2' = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad G3' = \overline{I_4} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}. \tag{5.4}$$

In this section, we use bijections Fi' from $\mathbb{F}_2^4$ to $\mathbb{F}_2^4$ (written with a prime) to mark the difference with the bijections $Fi : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ that were studied in Section 5.1 and 5.2.

The results are plotted in Tables 5, 6 and 7 for the randomized HD model and in Tables 8, 9 and 10 for the randomized NULL model. In all those tables, the left columns show the leakage of the proposed countermeasure (leakage squeezing), whereas the right columns show the leakage of the leak-free countermeasure (see [25]). Notice that the smaller the mutual information, the better the countermeasure. The curves are represented for a noise standard deviation σ living in the interval $[0, 5]$. We insist that the comparison between the different curves shall only be done when the noise is larger than the leakage of one sensitive bit, i.e. for $\sigma \geq 1$. This recommendation also applies to the interpretation of Figure 4. Indeed, a noise that is too low is not realistic in practice. Furthermore, if it was indeed technologically feasible to record measurements with low σ, the attacker would not compute statistics on the conditional distributions of the leakage (as in this paper). Instead, she would directly analyze the distributions, and for example, spot the apparition of particular values. One simple example can better explicit this point. Assume a sensitive bit $X \in \mathbb{F}_2^n$ is masked with a uniformly distributed mask $M \in \mathbb{F}_2$. In the absence of noise, the leakage is $L = \mathsf{HW}(X \oplus M, M)$.

- When $X = 0$, L takes two values, namely 0 and 2, with probability $\frac{1}{2}$;
- when $X = 1$, L is deterministic, actually always equal to 1.

So, it is trivial for an attacker to discover the value x of X by a single measurement of l of L; the deduction is the following: if $l = 0$ or $l = 2$, then $x = 0$, otherwise $x = 1$. In terms of information theory, this means that the mutual information between L and X is maximal, i.e. equal to $\mathsf{H}[X] = 1$ bit. Of course, in the presence of noise, it becomes more chancy to distinguish singularities in distributions based on a single leakage value; indeed, there is little information to be recovered on x

when $l = 1/2$ or $l = 3/4$. Thus, the attacker would rather compute statistics. The means of $L \mid X = 0$ and $L \mid X = 1$ are the same, but not their variance. From this point we see that the attack criteria defined in Section 3 makes sense.

In Tables 5, 6 and 7, it can be seen that despite the HD model degradation, the leakage of the CM remains

- ordered ($F3'$ leaks less than $F2'$, and $F2'$ in turn leaks less than $F1'$), as well as
- low, irrespective of δ.

The average leakage is unchanged, and the leakage values are simply getting slightly scattered. In particular, this means that if the leakage model is not the Hamming weight but a linear combination of the shares' bits, then the leakage squeezing continues to improve the security at its specified order. The reason for this resilience comes from the rationale of the leakage squeezing CM: the masked value and the mask are decorrelated as much as possible. The dispatching is guided by a randomized pigeon-hole of the values in the image of the leakage function. The CM thus looses efficiency only in the case where two different values of leakage become similar due to the imperfection. This can happen for some variables, but it is very unlikely that it occurs coherently for all variables at the same time. Rather, given the way the imperfect model is built (see (5.3)), it is almost as likely that two classes get nearer or further away. This explains why, in average, the leakage is not affected: the model noise acts as a random walk that has an impact on the variance but not on the average. Of course, some samples (with a degraded model) will be weaker than the others (because the variance of the MIA increases with the variance $\delta^2/12$ of the model).

It is interesting to contrast the leakage squeezing with the first-order leak-free CM presented in [24]. This CM aims at leaking no information when the HD leakage model is perfect. A study for model imperfection has also been conducted (see the right column of Tables 5, 6 and 7). It appears that this CM is much less robust to deviations from the ideal model. Indeed, the working factor of the CM is to have one share that leaks nothing. But as soon as there is some imperfection, the very principle of the CM is violated, and it starts to function less well. Concretely the leaked information increases with the model variance, up to a point where the CM is less efficient than the straightforward first-order Boolean masking (starting from $\delta > 50\%$).

For the sake of comparison, we also computed the same curves when the unnoised model is a constant one (called NULL model in (5.3)). The simulation results are shown in Tables 8, 9 and 10. The reference leakage (when $\delta = 0$) is null; consequently only the noisy curves are shown. It is noticeable that despite this NULL leakage model is random, the different CMs have clearly distinguishable

efficiencies. This had already been noticed by Doget et al. in [12]. In particular, it appears that our CM (the leakage squeezing) continues to work ($F3$ leaks less than $F2$, and $F2$ in turn leaks less than $F1$), at least for large enough noise standard deviations σ. At the opposite, the leak-free CM is not resilient to this random model: it leaks more than the straightforward masking (i.e. with $F1$).

Eventually, the impact of the leakage degree τ can be studied. Results are computed for $\tau \leq \tau_{\max}$, with $\tau_{\max} \in \{1, 2, 3\}$. In all the cases, τ does not impact the general conclusions.

Regarding the deviation from the HD model, the greater the multivariate degree τ, the more possible deviations from the genuine ideal model. Indeed, the number of random terms in (5.1) is increasing with τ (and is equal to $\sum_{t=0}^{\tau} \binom{2n}{t}$). This explains the greatest variability in the mutual information results. In the meantime, the argumentation for the robustness of the CM against the model deviation still holds, which explains why the average leakage is roughly unchanged. Nonetheless, as the order τ of the imperfection increases, some combinations are already done within the leakage model before any attack. This explains why the slope of the mutual information versus noise standard deviation σ becomes slightly less steep when τ is higher. In the NULL model, the greater τ, the less singularities in the leakage. This explains why the mutual information curves get smoother despite the additional noise. But with the greater τ, the more leaking sources (because the more non-zero terms in the polynomial), which explains why the leaked mutual information increases in average with τ.

6 Conclusions

Masking is a CM against side-channel attacks that consists in injecting some randomness in the execution of a computation. The sensitive value is split in several shares; altogether, they allow to reconstruct the sensitive data by an adequate combination [16]. In this article, we focus on a first-order Boolean masking CM that uses two shares, computed concomitantly. Zero-offset HO-CPA attacks can defeat this CM. They consist in computing a correlation with the centered side-channel traces, raised at the power $d \in \mathbb{N}^\star$. We show that when we know that the device leaks in Hamming distance, the highest order d of a successful zero-offset attack can be increased significantly thanks to the "leakage squeezing". Its principle is to store $F(m)$ (the image of m by a bijection F) instead of m in the mask register. Typically, when the data to protect are bytes, the state-of-the-art implementations with one mask could be attacked with HO-CPA of order $d = 2$. We show how to find an optimal linear F that protects against zero-offset HO-CPA of orders 1, 2, 3 and 4. We also show that optimal non-linear functions F pro-

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$
$p=0$	16	0	0	0	0
$p=1$	32	−8	0	0	0
$p=2$	80	−32	8	0	0
$p=3$	224	−116	48	−12	0
$p=4$	680	−416	224	−96	$24^{\dagger}$
$\vdots$	>0	<0	>0	<0	>0

Table 3. Some values of $H(n=4,p,h)$.

tect against zero-offset HO-CPA of orders 1, 2, 3, 4 and 5. The implementations that can benefit from the protection conveyed by such non-linear bijection F are those for which the masks used at each clock cycle are independent and uniformly distributed. This security increase also translates into a leakage reduction. An information-theoretic study reveals that the mutual information between the leakage and the sensitive variable is lower than the same metric computed on a similar CM without F but that uses two masks (instead of one).

A Proof of Theorem 3.1

A.1 First intermediate result for the proof of Theorem 3.1

Theorem A.1. *For all $a \in \mathbb{F}_2^n$ and $p \in \mathbb{N}$,*

$$\widehat{\mathsf{HW}^p}(a) = 0 \iff \mathsf{HW}(a) > p.$$

For $n \in \mathbb{N}^\star$, $p \in \mathbb{N}$ and $h \in [\![0,n]\!]$, let us define the function

$$H(n,p,h) \doteq \sum_{z \in \mathbb{F}_2^n} \mathsf{HW}^p(z)(-1)^{z \cdot \bigoplus_{i=1}^h e_i}.$$

It is tabulated for $n=4$ in Table 3. The value $H(n,n,n)$, indicated by the dagger symbol † in the table, is equal to $(-1)^n n!$.

As the order of the bits of the dummy variable z is indifferent in the term $\sum_z \mathsf{HW}^p(z)(-1)^{a\cdot z}$, we have $\widehat{\mathsf{HW}^p}(a) = H(n,p,\mathsf{HW}(a))$.

Lemma A.2.

$$H(n,p,n) \begin{cases} = 0 & \text{if } p < n, \\ > 0 & \text{if } p \geq n \text{ and } n \text{ is even}, \\ < 0 & \text{if } p \geq n \text{ and } n \text{ is odd}. \end{cases}$$

Proof. We have

$$\begin{aligned} H(n,p,n) &= \sum_z \mathsf{HW}^p(z)(-1)^{z\cdot\bigoplus_{i=1}^n e_i} = \sum_z \mathsf{HW}^p(z)(-1)^{\mathsf{HW}(z)} \\ &= \sum_{j=0}^{n} \binom{n}{j} j^p (-1)^j = (-1)^n \sum_{j=0}^{n} \binom{n}{j} j^p (-1)^{n-j} \\ &= (-1)^n n! \left\{ {p \atop n} \right\}, \end{aligned}$$

where $\left\{ {p \atop n} \right\}$ is a Stirling number of the second kind [37]. More precisely, it is the number of ways of partitioning a set of p elements into n non-empty sets. Consequently, $\left\{ {p \atop n} \right\} = 0$ if $n > p$, because otherwise at least one set would be empty. Also, $\left\{ {p \atop n} \right\} > 0$ if $n \le p$. Now, the sign of $H(n,p,n)$ depends on the parity of n if $n \le p$. It is positive (resp. negative) if n is even (resp. odd). □

Lemma A.3.

$$H(n,p,h) \begin{cases} = 0 & \text{if } p < h, \\ > 0 & \text{if } p \ge h \text{ and } h \text{ is even}, \\ < 0 & \text{if } p \ge h \text{ and } h \text{ is odd}. \end{cases}$$

Proof. This lemma has already been proved in Lemma A.2 if $h = n$. Thus, we assume in the remainder of this proof that $h < n$. For $z \in \mathbb{F}_2^n$, we write $z = (z_L, z_H)$, where $z_L \in \mathbb{F}_2^h$ and $z_H \in \mathbb{F}_2^{n-h}$. Then

$$\begin{aligned} H(n,p,h) &= \sum_{(z_L,z_H)} \mathsf{HW}^p((z_L,0) \oplus (0,z_H))(-1)^{(z_L\cdot\bigoplus_{i=1}^h e_i)\oplus(z_H\cdot 0)} \\ &= \sum_{(z_L,z_H)} (\mathsf{HW}(z_L) + \mathsf{HW}(z_H))^p (-1)^{z_L\cdot\bigoplus_{i=1}^h e_i} \\ &= \sum_{(z_L,z_H)} \sum_{j=0}^{p} \binom{p}{j} \times \mathsf{HW}^j(z_L) \times \mathsf{HW}^{p-j}(z_H)(-1)^{z_L\cdot\bigoplus_{i=1}^h e_i} \\ &= \sum_{j=0}^{p} \binom{p}{j} \sum_{z_L} \mathsf{HW}^j(z_L)(-1)^{z_L\cdot\bigoplus_{i=1}^h e_i} \times \sum_{z_H} \mathsf{HW}^{p-j}(z_H) \\ &= \sum_{j=0}^{p} \binom{p}{j} \times H(h,j,h) \times H(n-h,p-j,0). \end{aligned} \tag{A.1}$$

Now, given Lemma A.2, we have $H(h, j, h) = 0$ for all $j < h$. Thus, if $p < h$, then all the terms $H(h, j, h)$ involved in (A.1) are null, since $j \in [\![0, p]\!]$ is strictly inferior to h.

Besides, for all $j \in [\![0, p]\!]$, $\binom{p}{j}$ and $H(n-h, p-j, 0)$ are strictly positive. If $p \geq h$, the terms $H(h, j, h)$ for $j \leq p$ are

- either all strictly positive if h is even,
- or all strictly negative if h is odd.

Hence, so is the sum in (A.1). □

Proof of Theorem A.1. As already noticed, $\widehat{\mathsf{HW}^p}(a) = H(n, p, \mathsf{HW}(a))$. According to Lemma A.3, this quantity is null if and only if $p < \mathsf{HW}(a)$. □

A.2 Second intermediate result for the proof of Theorem 3.1

For every $X \in \mathbb{F}_2^n$, we have

$$\begin{aligned}\Big(\sum_{i=1}^{n}(-1)^{X\cdot e_i}\Big)^j &= \sum_{i_1,\ldots,i_j\in[\![1,n]\!]^j}\prod_{l=1}^{j}(-1)^{X\cdot e_{i_l}}\\ &= \sum_{i_1,\ldots,i_j\in[\![1,n]\!]^j}(-1)^{X\cdot\bigoplus_{l=1}^{j}e_{i_l}}\\ &= \sum_{\substack{(k_1,\ldots,k_n)\in\mathbb{N}^n,\\ k_1+\cdots+k_n=j}}\binom{j}{k_1,\ldots,k_n}(-1)^{X\cdot(\bigoplus_{i=1}^{n}k_ie_i)}, \qquad (A.2)\end{aligned}$$

where each vector k_ie_i in $\bigoplus_{i=1}^{n}k_ie_i$ is either e_i if k_i is odd or 0 otherwise. The term $\binom{j}{k_1,\ldots,k_n}$ is a multinomial coefficient. Actually, under the form in the second line of (A.2) some terms appear multiple times.

Then, we have

$$\begin{aligned}&\sum_{z,m}\mathsf{HW}^q(F(m)\oplus F(m\oplus z))(-1)^{a\cdot z}\\ &\quad=\frac{1}{2^q}\sum_{z,m}\Big(n-\sum_{i=1}^{n}(-1)^{F_i(m)\oplus F_i(m\oplus z)}\Big)^q(-1)^{a\cdot z}\\ &\quad=\frac{1}{2^q}\sum_{z,m}\sum_{j=0}^{q}\binom{q}{j}n^{q-j}(-1)^j\Big(\sum_{i=1}^{n}(-1)^{F_i(m)\oplus F_i(m\oplus z)}\Big)^j(-1)^{a\cdot z}\end{aligned}$$

$$= \frac{1}{2^q} \sum_{j=0}^{q} \binom{q}{j} n^{q-j} (-1)^j \sum_{k_1+\cdots+k_n=j} \binom{j}{k_1, \ldots, k_n} \times \sum_{z,m} (-1)^{(F(m)\oplus F(m\oplus z))\cdot(\bigoplus_{i=1}^{n} k_i e_i)} (-1)^{a\cdot z}$$

$$= \frac{1}{2^q} \sum_{j=0}^{q} \binom{q}{j} n^{q-j} (-1)^j \times \sum_{k_1+\cdots+k_n=j} \binom{j}{k_1, \ldots, k_n} \widehat{\left(\left(\left(\bigoplus_{i=1}^{n} k_i e_i \right) \cdot F \right)_\chi (a) \right)^2}. \tag{A.3}$$

See (A.2) for the third line of (A.3).

A.3 Complete proof of Theorem 3.1

As requested by Theorem 3.1, we introduce two positive integers P and Q, and a bijection F of $\mathbb{F}_2^n$. With a reasoning close to that of (3.4) for the case $p = q = 1$, we get the following equivalences for all $p \in [\![0, P]\!]$, $q \in [\![0, Q]\!]$ and $a \in \mathbb{F}_2^{n\star}$:

the function f_{opt} defined in (3.2) is constant

$\iff$ $\widehat{\mathsf{HW}^p}(a) = 0$ or $\widehat{\mathbb{E}[\mathsf{HW}^q \circ D_{(\cdot)}F(M)]}(a) = 0$

$\iff$ either $\mathsf{HW}(a) > p$ (see Theorem A.1) or (A.3) of Section A.2 is zero

$\iff$ $\mathsf{HW}(a) \le p \implies$ equation (A.3) is zero

$$\iff \mathsf{HW}(a) \le p \implies \begin{cases} \forall b, \mathsf{HW}(b) \le 1 \implies \widehat{(b \cdot F)_\chi}(a) = 0 & \text{if } q = 1, \\ \forall b, \mathsf{HW}(b) \le 2 \implies \widehat{(b \cdot F)_\chi}(a) = 0 & \text{if } q = 2, \\ \vdots & \\ \forall b, \mathsf{HW}(b) \le Q \implies \widehat{(b \cdot F)_\chi}(a) = 0 & \text{if } q = Q. \end{cases}$$

We provide an explanation for the last part. The terms of (A.3) corresponding to a given j are squares (weighted by quantities of the same sign). Thus, if those terms for $j < q$ are null, then the ones for $j = q$ must also be null, because the complete sum (of squares) is null by hypothesis.

B Optimal linear solution for $n = 8$

As shown in Section 4.2, the optimal linear function in the case $n = 8$ is generated by the non-identity half of the systematic matrix of $[16, 8, 5]$ code. This matrix is[4]

$$\begin{pmatrix} 1&1&1&0&1&0&1&1&1&0&0&0&0&0&0&0 \\ 0&1&1&1&0&1&0&1&1&1&0&0&0&0&0&0 \\ 0&0&1&1&1&0&1&0&1&1&1&0&0&0&0&0 \\ 0&0&0&1&1&1&0&1&0&1&1&1&0&0&0&0 \\ 0&0&0&0&1&1&1&0&1&0&1&1&1&0&0&0 \\ 0&0&0&0&0&1&1&1&0&1&0&1&1&1&0&0 \\ 0&0&0&0&0&0&1&1&1&0&1&0&1&1&1&0 \\ 0&0&0&0&0&0&0&1&1&1&0&1&0&1&1&1 \end{pmatrix} \begin{matrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8. \end{matrix}$$

It is already in row echelon form. Therefore, it can be turned into systematic form with a Gauss–Jordan elimination. It involves the following linear operations on the rows:

$$\begin{aligned} L'_1 &\leftarrow L_1 \oplus L_2 \oplus L_4 \oplus L_7 \\ L'_2 &\leftarrow L_2 \oplus L_3 \oplus L_5 \oplus L_8 \\ L'_3 &\leftarrow L_3 \oplus L_4 \oplus L_6 \\ L'_4 &\leftarrow L_4 \oplus L_5 \oplus L_7 \\ L'_5 &\leftarrow L_5 \oplus L_6 \oplus L_8 \\ L'_6 &\leftarrow L_6 \oplus L_7 \\ L'_7 &\leftarrow L_7 \oplus L_8 \\ L'_8 &\leftarrow L_8 \end{aligned}$$

whose execution yields

$$\begin{pmatrix} 1&0&0&0&0&0&0&0&1&0&0&1&1&1&1&0 \\ 0&1&0&0&0&0&0&0&0&1&0&0&1&1&1&1 \\ 0&0&1&0&0&0&0&0&1&1&0&0&1&1&0&0 \\ 0&0&0&1&0&0&0&0&0&1&1&0&0&1&1&0 \\ 0&0&0&0&1&0&0&0&0&0&1&1&0&0&1&1 \\ 0&0&0&0&0&1&0&0&1&1&1&1&0&0&1&0 \\ 0&0&0&0&0&0&1&0&0&1&1&1&1&0&0&1 \\ 0&0&0&0&0&0&0&1&1&1&0&1&0&1&1&1 \end{pmatrix} \begin{matrix} L'_1 = \texttt{0x80} \parallel \texttt{0x9e} \\ L'_2 = \texttt{0x40} \parallel \texttt{0x4f} \\ L'_3 = \texttt{0x20} \parallel \texttt{0xcc} \\ L'_4 = \texttt{0x10} \parallel \texttt{0x66} \\ L'_5 = \texttt{0x08} \parallel \texttt{0x33} \\ L'_6 = \texttt{0x04} \parallel \texttt{0xf2} \\ L'_7 = \texttt{0x02} \parallel \texttt{0x79} \\ L'_8 = \texttt{0x01} \parallel \texttt{0xd7} \end{matrix}$$

[4] As already mentioned in Section 4.2, this code is a subcode of the BCH $[17, 9, 5]$ code. For more details, we refer to `www.math.colostate.edu/~betten/research/codes/BOUNDS/sub_16_8_5-7_2.code`.

having the expected form $(I_8\ B4)$. The bijection $F4 : x \mapsto B4 \times x$ is the optimal linear bijection for $n = 8$.

C Optimal non-linear solution for $n = 8$

The function $F5$, whose construction is given in Section 4.3, takes the values

$$
\begin{aligned}
&\{F(x);\ x \in \mathbb{F}_2^8\} \\
&= \{\ \texttt{0x00},\ \texttt{0xb3},\ \texttt{0xe5},\ \texttt{0x6a},\ \texttt{0x2f},\ \texttt{0xc6},\ \texttt{0x5c},\ \texttt{0x89},\\
&\quad\ \texttt{0x79},\ \texttt{0xac},\ \texttt{0x36},\ \texttt{0xdf},\ \texttt{0x9a},\ \texttt{0x15},\ \texttt{0x43},\ \texttt{0xf0},\\
&\quad\ \texttt{0xcb},\ \texttt{0x1e},\ \texttt{0xb8},\ \texttt{0x51},\ \texttt{0x72},\ \texttt{0xfd},\ \texttt{0x97},\ \texttt{0x24},\\
&\quad\ \texttt{0xd4},\ \texttt{0x67},\ \texttt{0x0d},\ \texttt{0x82},\ \texttt{0xa1},\ \texttt{0x48},\ \texttt{0xee},\ \texttt{0x3b},\\
&\quad\ \texttt{0x9d},\ \texttt{0x74},\ \texttt{0xd2},\ \texttt{0x07},\ \texttt{0xe8},\ \texttt{0x5b},\ \texttt{0x31},\ \texttt{0xbe},\\
&\quad\ \texttt{0x4e},\ \texttt{0xc1},\ \texttt{0xab},\ \texttt{0x18},\ \texttt{0xf7},\ \texttt{0x22},\ \texttt{0x84},\ \texttt{0x6d},\\
&\quad\ \texttt{0xa6},\ \texttt{0x29},\ \texttt{0x7f},\ \texttt{0xcc},\ \texttt{0x45},\ \texttt{0x90},\ \texttt{0x0a},\ \texttt{0xe3},\\
&\quad\ \texttt{0x13},\ \texttt{0xfa},\ \texttt{0x60},\ \texttt{0xb5},\ \texttt{0x3c},\ \texttt{0x8f},\ \texttt{0xd9},\ \texttt{0x56},\\
&\quad\ \texttt{0x57},\ \texttt{0xd8},\ \texttt{0x8e},\ \texttt{0x3d},\ \texttt{0xb4},\ \texttt{0x61},\ \texttt{0xfb},\ \texttt{0x12},\\
&\quad\ \texttt{0xe2},\ \texttt{0x0b},\ \texttt{0x91},\ \texttt{0x44},\ \texttt{0xcd},\ \texttt{0x7e},\ \texttt{0x28},\ \texttt{0xa7},\\
&\quad\ \texttt{0x6c},\ \texttt{0x85},\ \texttt{0x23},\ \texttt{0xf6},\ \texttt{0x19},\ \texttt{0xaa},\ \texttt{0xc0},\ \texttt{0x4f},\\
&\quad\ \texttt{0xbf},\ \texttt{0x30},\ \texttt{0x5a},\ \texttt{0xe9},\ \texttt{0x06},\ \texttt{0xd3},\ \texttt{0x75},\ \texttt{0x9c},\\
&\quad\ \texttt{0x3a},\ \texttt{0xef},\ \texttt{0x49},\ \texttt{0xa0},\ \texttt{0x83},\ \texttt{0x0c},\ \texttt{0x66},\ \texttt{0xd5},\\
&\quad\ \texttt{0x25},\ \texttt{0x96},\ \texttt{0xfc},\ \texttt{0x73},\ \texttt{0x50},\ \texttt{0xb9},\ \texttt{0x1f},\ \texttt{0xca},\\
&\quad\ \texttt{0xf1},\ \texttt{0x42},\ \texttt{0x14},\ \texttt{0x9b},\ \texttt{0xde},\ \texttt{0x37},\ \texttt{0xad},\ \texttt{0x78},\\
&\quad\ \texttt{0x88},\ \texttt{0x5d},\ \texttt{0xc7},\ \texttt{0x2e},\ \texttt{0x6b},\ \texttt{0xe4},\ \texttt{0xb2},\ \texttt{0x01},\\
&\quad\ \texttt{0xfe},\ \texttt{0x4d},\ \texttt{0x1b},\ \texttt{0x94},\ \texttt{0xd1},\ \texttt{0x38},\ \texttt{0xa2},\ \texttt{0x77},\\
&\quad\ \texttt{0x87},\ \texttt{0x52},\ \texttt{0xc8},\ \texttt{0x21},\ \texttt{0x64},\ \texttt{0xeb},\ \texttt{0xbd},\ \texttt{0x0e},\\
&\quad\ \texttt{0x35},\ \texttt{0xe0},\ \texttt{0x46},\ \texttt{0xaf},\ \texttt{0x8c},\ \texttt{0x03},\ \texttt{0x69},\ \texttt{0xda},\\
&\quad\ \texttt{0x2a},\ \texttt{0x99},\ \texttt{0xf3},\ \texttt{0x7c},\ \texttt{0x5f},\ \texttt{0xb6},\ \texttt{0x10},\ \texttt{0xc5},\\
&\quad\ \texttt{0x63},\ \texttt{0x8a},\ \texttt{0x2c},\ \texttt{0xf9},\ \texttt{0x16},\ \texttt{0xa5},\ \texttt{0xcf},\ \texttt{0x40},\\
&\quad\ \texttt{0xb0},\ \texttt{0x3f},\ \texttt{0x55},\ \texttt{0xe6},\ \texttt{0x09},\ \texttt{0xdc},\ \texttt{0x7a},\ \texttt{0x93},\\
&\quad\ \texttt{0x58},\ \texttt{0xd7},\ \texttt{0x81},\ \texttt{0x32},\ \texttt{0xbb},\ \texttt{0x6e},\ \texttt{0xf4},\ \texttt{0x1d},\\
&\quad\ \texttt{0xed},\ \texttt{0x04},\ \texttt{0x9e},\ \texttt{0x4b},\ \texttt{0xc2},\ \texttt{0x71},\ \texttt{0x27},\ \texttt{0xa8},\\
&\quad\ \texttt{0xa9},\ \texttt{0x26},\ \texttt{0x70},\ \texttt{0xc3},\ \texttt{0x4a},\ \texttt{0x9f},\ \texttt{0x05},\ \texttt{0xec},\\
&\quad\ \texttt{0x1c},\ \texttt{0xf5},\ \texttt{0x6f},\ \texttt{0xba},\ \texttt{0x33},\ \texttt{0x80},\ \texttt{0xd6},\ \texttt{0x59},\\
&\quad\ \texttt{0x92},\ \texttt{0x7b},\ \texttt{0xdd},\ \texttt{0x08},\ \texttt{0xe7},\ \texttt{0x54},\ \texttt{0x3e},\ \texttt{0xb1},\\
&\quad\ \texttt{0x41},\ \texttt{0xce},\ \texttt{0xa4},\ \texttt{0x17},\ \texttt{0xf8},\ \texttt{0x2d},\ \texttt{0x8b},\ \texttt{0x62},\\
&\quad\ \texttt{0xc4},\ \texttt{0x11},\ \texttt{0xb7},\ \texttt{0x5e},\ \texttt{0x7d},\ \texttt{0xf2},\ \texttt{0x98},\ \texttt{0x2b},\\
&\quad\ \texttt{0xdb},\ \texttt{0x68},\ \texttt{0x02},\ \texttt{0x8d},\ \texttt{0xae},\ \texttt{0x47},\ \texttt{0xe1},\ \texttt{0x34},\\
&\quad\ \texttt{0x0f},\ \texttt{0xbc},\ \texttt{0xea},\ \texttt{0x65},\ \texttt{0x20},\ \texttt{0xc9},\ \texttt{0x53},\ \texttt{0x86},\\
&\quad\ \texttt{0x76},\ \texttt{0xa3},\ \texttt{0x39},\ \texttt{0xd0},\ \texttt{0x95},\ \texttt{0x1a},\ \texttt{0x4c},\ \texttt{0xff}\ \}.
\end{aligned}
$$

D Computation of the optimal function $z \mapsto f_{\mathrm{opt}}(z)$ for some bijections F

Some $f_{\mathrm{opt}}(z)$ have been computed in Table 4 for centered traces raised at power $d \in [\![1, 6]\!]$, for some representative bijections, including the optimal linear ($F4$) and non-linear ($F5$) ones. The last column shows the optimal correlation coefficient ρ_{opt} that an attacker can expect (see the definition in [31, (15)]). It can be seen that the first non-zero ρ_{opt} approximately decreases with the CM strength: it is about 25% for $F1$, about 4% for $F2$ and $F3$, and about 2% for $F4$ and $F5$.

E Information leakage in the imperfect model

The information leakage plots are plotted in Tables 5, 6 and 7 for the randomized HD model and in Tables 8, 9 and 10 for the randomized NULL model.

z	0x00	0x01	0x03	0x07	0x0f	0x1f	0x3f	0x7f	0xff	
Bijection $F = F1$ (reference $F1 : x \mapsto I_8 \times x = x$)										
$d = 1$	0	0	0	0	0	0	0	0	0	0.000000
$d = 2$	8	7	6	5	4	3	2	1	0	0.258199
$d = 3$	0	0	0	0	0	0	0	0	0	0.000000
$d = 4$	176	133	96	65	40	21	8	1	0	0.235341
$d = 5$	0	0	0	0	0	0	0	0	0	0.000000
$d = 6$	5888	3787	2256	1205	544	183	32	1	0	0.197908
Bijection $F = F2$ (linear $F2 : x \mapsto G2 \times x$)										
$d = 1$	0	0	0	0	0	0	0	0	0	0.000000
$d = 2$	4	4	4	4	4	4	4	4	4	0.000000
$d = 3$	−1.5	−1.5	−1.5	−1.5	0	0	0	0	1.5	0.036509
$d = 4$	49	49	49	49	49	46	49	46	46	0.015548
$d = 5$	−120	−75	−37.5	−30	7.5	22.5	15	22.5	67.5	0.051072
$d = 6$	1399	1061	949	971.5	971.5	821.5	971.5	821.5	979	0.027247
Bijection $F = F3$ (linear $F3 : x \mapsto G3 \times x$)										
$d = 1$	0	0	0	0	0	0	0	0	0	0.000000
$d = 2$	4	4	4	4	4	4	4	4	4	0.000000
$d = 3$	0	0	0	0	0	0	0	0	0	0.000000
$d = 4$	70	61	52	43	40	37	40	43	46	0.043976
$d = 5$	0	0	0	0	0	0	0	0	0	0.000000
$d = 6$	2584	1684	1144	694	544	484	544	694	664	0.067175
Bijection $F = F4$ (linear $F4 : x \mapsto G4 \times x$)										
$d = 1$	0	0	0	0	0	0	0	0	0	0.000000
$d = 2$	4	4	4	4	4	4	4	4	4	0.000000
$d = 3$	0	0	0	0	0	0	0	0	0	0.000000
$d = 4$	46	46	46	46	46	46	46	46	46	0.000000
$d = 5$	−90	−37.5	−15	15	7.5	−22.5	7.5	7.5	0	0.023231
$d = 6$	1339	956.5	799	799	866.5	821.5	776.5	821.5	844	0.016173
Bijection $F = F5$ (non-linear F tabulated in Section 4.3)										
$d = 1$	0	0	0	0	0	0	0	0	0	0.000000
$d = 2$	4	4	4	4	4	4	4	4	4	0.000000
$d = 3$	0	0	0	0	0	0	0	0	0	0.000000
$d = 4$	46	46	46	46	46	46	46	46	46	0.000000
$d = 5$	0	0	0	0	0	0	0	0	0	0.000000
$d = 6$	2104	1159	844	799	664	799	844	1159	844	0.023258

Table 4. Computation of $f_{\text{opt}}(z)$ (columns 2–10) for centered traces raised at several powers d, and optimal correlation coefficient ρ_{opt} (column 11).

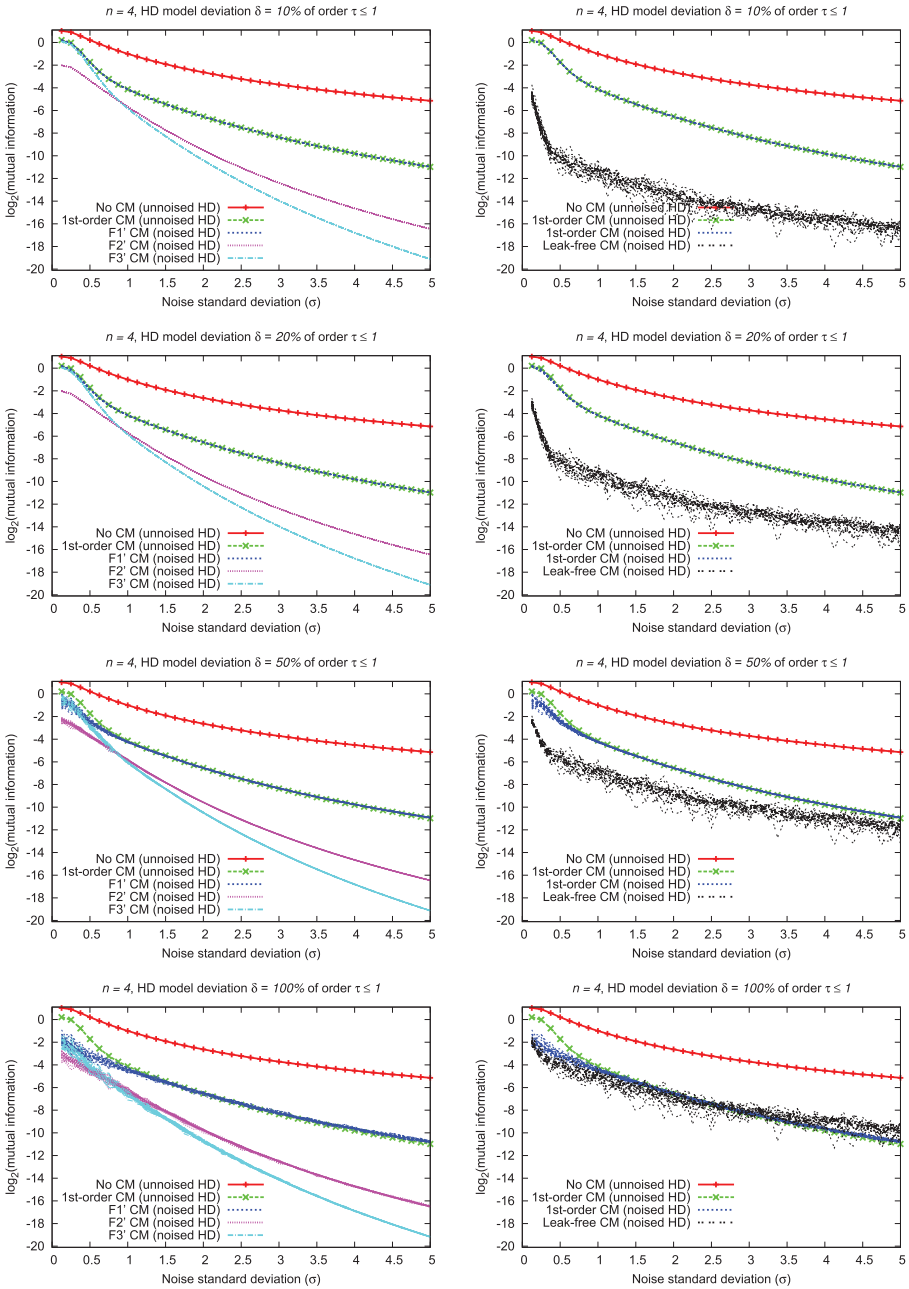

Table 5. Leakage comparison in the imperfect HD leakage model, where imperfections are of multivariate degree $\tau \leq 1$.

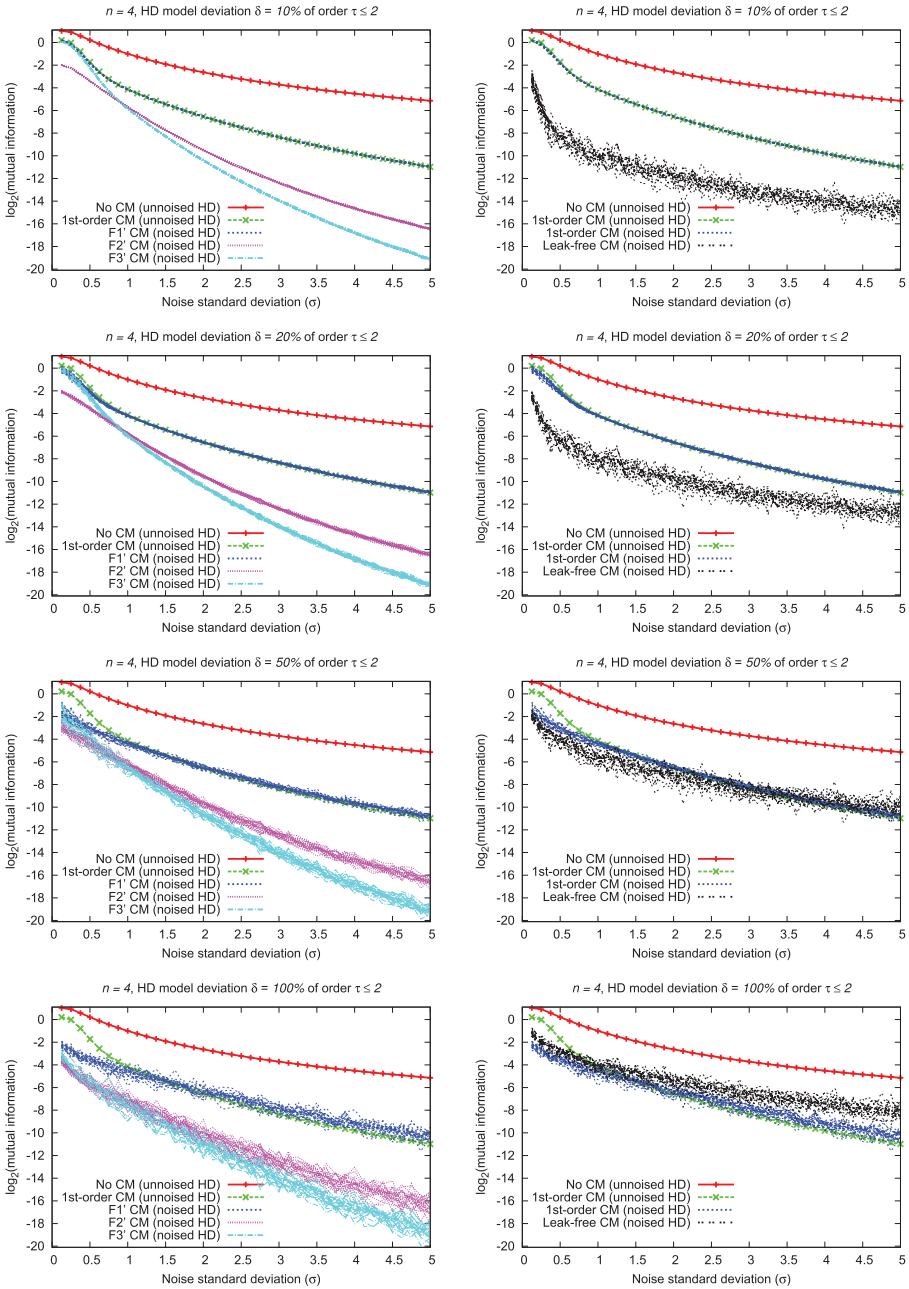

Table 6. Leakage comparison in the imperfect HD leakage model, where imperfections are of multivariate degree $\tau \leq 2$.

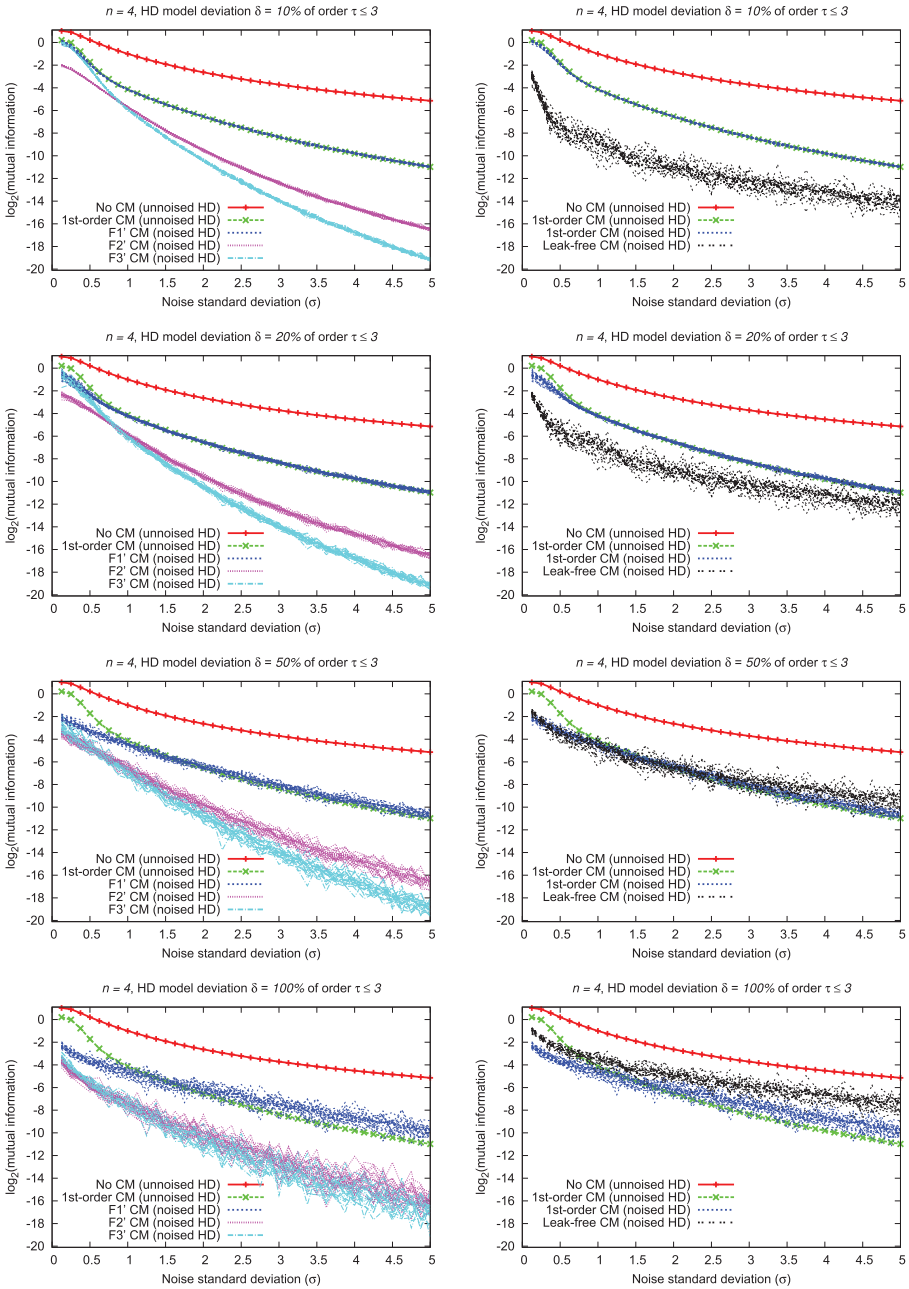

Table 7. Leakage comparison in the imperfect HD leakage model, where imperfections are of multivariate degree $\tau \leq 3$.

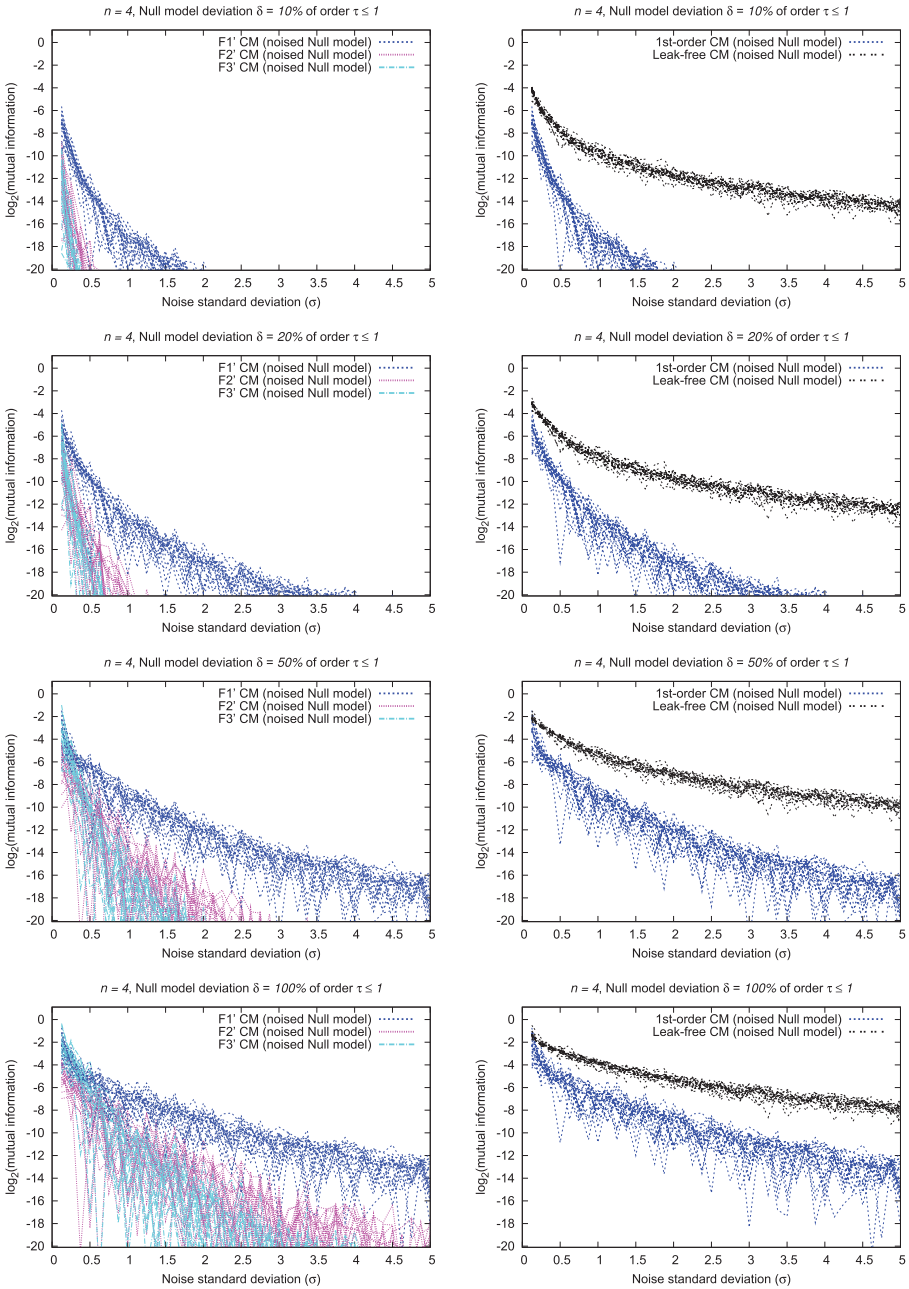

Table 8. Leakage comparison in the imperfect NULL leakage model, where imperfections are of multivariate degree $\tau \leq 1$.

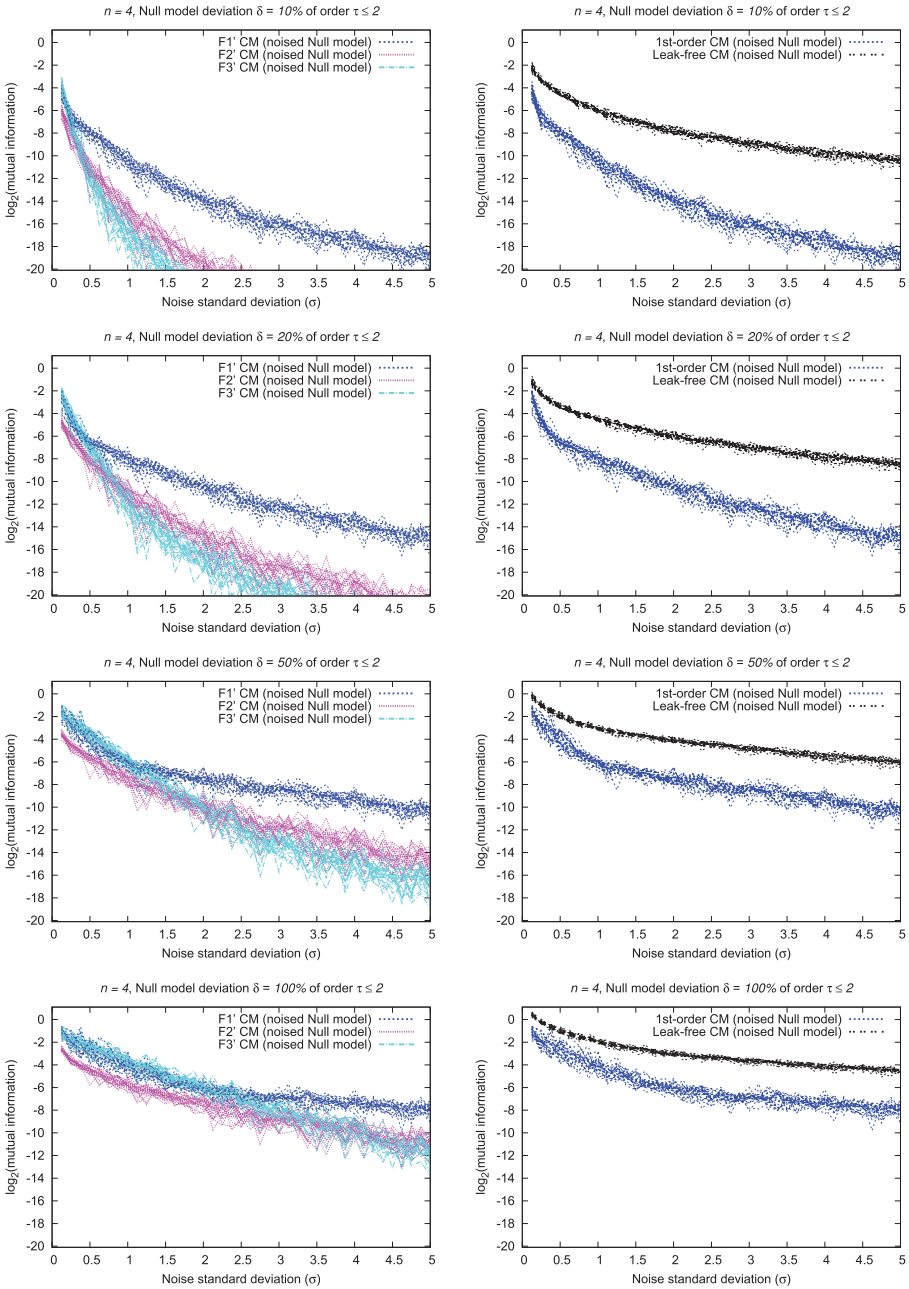

Table 9. Leakage comparison in the imperfect NULL leakage model, where imperfections are of multivariate degree $\tau \leq 2$.

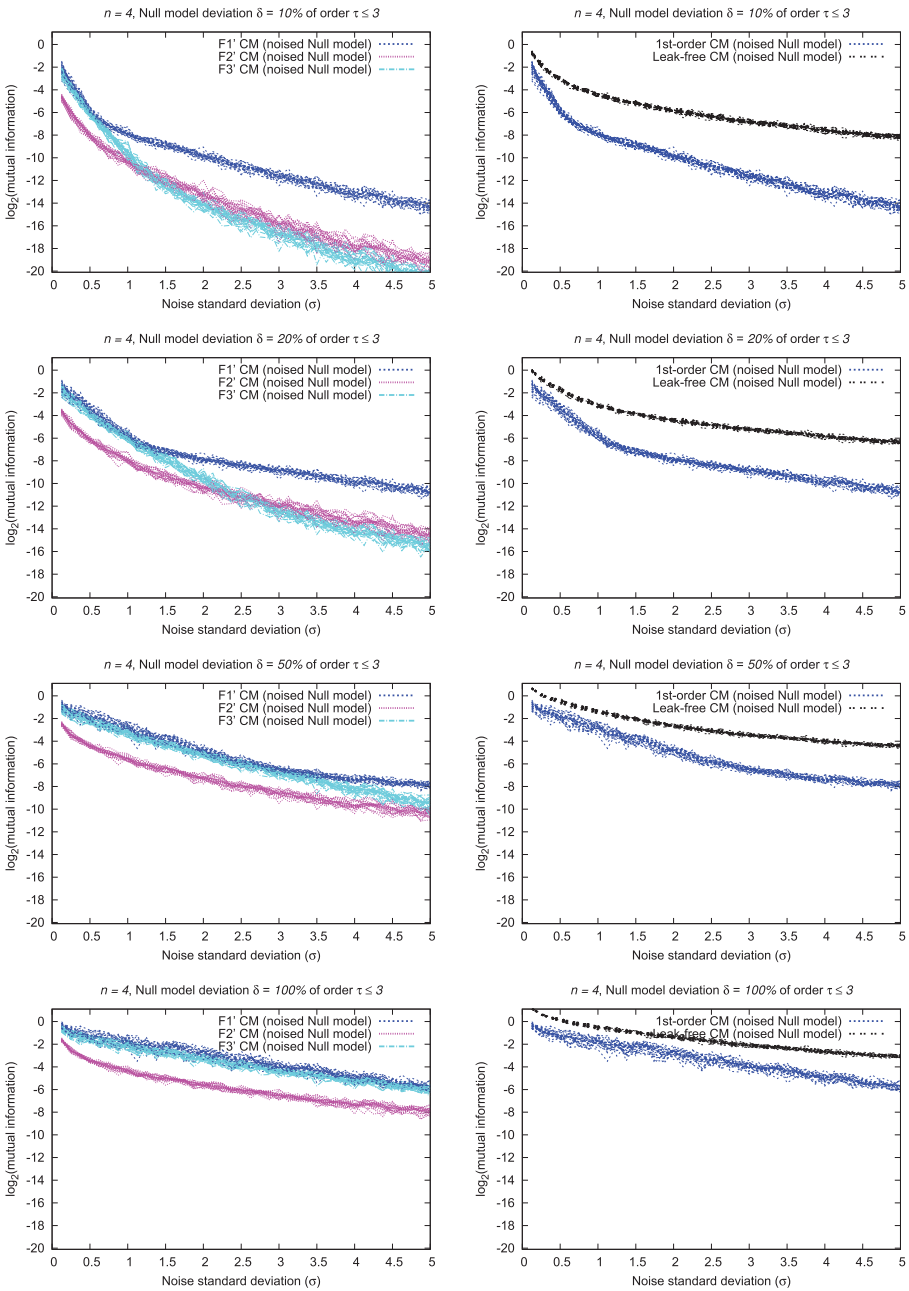

Table 10. Leakage comparison in the imperfect NULL leakage model, where imperfections are of multivariate degree $\tau \leq 3$.

Acknowledgments. The authors are grateful to Sébastien Briais (Secure-IC S.A.S.), M. Abdelaziz Elaabid (Université Paris 8) and Patrick Solé (TELECOM-ParisTech and King Abdulaziz University) for insightful discussions. Besides, we sincerely acknowledge the thorough reviews we have received. Many points, such as the conditions on the mask refresh function for leakage squeezing to work as well in the context of Hamming distance and Hamming weight leakage functions, have been suggested by the reviewers.

Secure-IC and Télécom-ParisTech are funding members, with DOREMI, of the "Secure Compression Lab". Morpho and Télécom-ParisTech are funders of the "Identity & Security Alliance".

Bibliography

[1] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert and N. Veyrat-Charvillon, Mutual information analysis: A comprehensive study, *J. Cryptology* **24** (2011), no. 2, 269–291.

[2] J. Blömer, J. Guajardo and V. Krummel, Provably secure masking of AES, in: *Selected Areas in Cryptography*, Lecture Notes in Comput. Sci. 3357, Springer (2004), 69–83.

[3] É. Brier, C. Clavier and F. Olivier, Correlation power analysis with a leakage model, in: *CHES*, Lecture Notes in Comput. Sci. 3156, Springer (2004), 16–29.

[4] J. Bringer, H. Chabanne and T. Ha Le, Protecting AES against side-channel analysis using wire-tap codes, *J. Cryptographic Engineering* **2** (2012), no. 2, 129–141.

[5] P. Camion, C. Carlet, P. Charpin and N. Sendrier, On correlation-immune functions, in: *CRYPTO*, Lecture Notes in Comput. Sci. 576, Springer (1991), 86–100.

[6] C. Carlet, Boolean functions for cryptography and error correcting codes, in: *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, Cambridge University Press (2010), 257–397.

[7] C. Carlet, Vectorial boolean functions for cryptography, in: *Boolean Models and Methods in Mathematics, Computer Science, and Engineering*, Cambridge University Press (2010), 398–469.

[8] C. Carlet, P. Gaborit, J.-L. Kim and P. Solé, A new class of codes for Boolean masking of cryptographic computations, preprint (2011), `http://arxiv.org/abs/1110.1193`. Early version of [9].

[9] C. Carlet, P. Gaborit, J.-L. Kim and P. Solé, A new class of codes for boolean masking of cryptographic computations, *IEEE Trans. Inform. Theory* **58** (2012), no. 9, 6000–6011.

[10] J.-L. Danger and S. Guilley, Cryptography circuit protected against observation attacks, in: *Particular of a High Order*, September 23, 2010. International patent, published as FR2941342 (A1), WO2010084106 (A1) & (A9), EP2380306 (A1), CA2749961 (A1).

[11] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, PhD thesis, Université Catholique de Louvain, Belgium, 1973.

[12] J. Doget, E. Prouff, M. Rivain and F.-X. Standaert, Univariate side channel attacks and leakage modeling, *J. Cryptographic Engineering* **1** (2011), no. 2, 123–144.

[13] S. Drimer, T. Güneysu and C. Paar, DSPs, BRAMs, and a pinch of logic: Extended recipes for AES on FPGAs, *ACM Trans. Reconfigurable Tech. Syst.* **3** (2010), 1–27.

[14] V. Fischer and M. Drutarovský, Two methods of Rijndael implementation in reconfigurable hardware, in: *CHES*, Lecture Notes in Comput. Sci. 2162, Springer (2001), 77–92.

[15] G. D. Forney Jr., N. J. A. Sloane and M. D. Trott, The Nordstrom–Robinson code is the binary image of the octacode, in: *Coding and Quantization: DIMACS/IEEE Workshop* (1992), 19–26.

[16] L. Goubin and J. Patarin, DES and differential power analysis. The "duplication" method, in: *CHES*, Lecture Notes in Comput. Sci., Springer (1999), 158–172.

[17] T. A. Gulliver and P. R. J. Östergård, Binary optimal linear rate 1/2 codes, *Discrete Math.* **283** (2004), no. 1–3, 255–261.

[18] T. Güneysu and A. Moradi, Generic side-channel countermeasures for reconfigurable devices, in: *CHES*, Lecture Notes in Comput. Sci. 6917, Springer (2011), 33–48.

[19] ISO/IEC 18033-3:2010, Information technology – Security techniques – Encryption algorithms – Part 3: Block ciphers.

[20] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North Holland, Amsterdam, 1977.

[21] H. Maghrebi, C. Carlet, S. Guilley and J.-L. Danger, Optimal first-order masking with linear and non-linear bijections, in: *AFRICACRYPT*, Lecture Notes in Comput. Sci. 7374, Springer (2012), 360–377.

[22] H. Maghrebi, S. Guilley, C. Carlet and J.-L. Danger, Classification of high-order boolean masking schemes and improvements of their efficiency, preprint (2011), `http://eprint.iacr.org/2011/520`.

[23] H. Maghrebi, S. Guilley and J.-L. Danger, Leakage squeezing countermeasure against high-order attacks, in: *WISTP*, Lecture Notes in Comput. Sci. 6633, Springer (2011), 208–223.

[24] H. Maghrebi, E. Prouff, S. Guilley and J.-L. Danger, A first-order leak-free masking countermeasure, in: *CT-RSA*, Lecture Notes in Comput. Sci. 7178, Springer (2012), 156–170.

[25] S. Mangard and K. Schramm, Pinpointing the side-channel leakage of masked AES hardware implementations, in: *CHES*, Lecture Notes in Comput. Sci. 4249, Springer (2006), 76–90.

[26] S. K. Mathew, F. Sheikh, M. Kounavis, S. Gueron, A. Agarwal, S. K. Hsu, H. Kaul, M. A. Anders and R. K. Krishnamurthy, 53 Gbps native GF$(2^4)^2$ composite-field AES-encrypt/decrypt accelerator for content-protection in 45 nm high-performance microprocessors, *IEEE J. Solid-State Circuits* **46** (2011), no. 4, 767–776.

[27] A. Moradi and O. Mischke, How far should theory be from practice? – Evaluation of a countermeasure, in: *CHES*, Lecture Notes in Comput. Sci. 7428, Springer (2012), 92–106.

[28] NIST/ITL/CSD, Advanced Encryption Standard (AES). FIPS PUB 197, November 2001, `http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf`.

[29] É. Peeters, F.-X. Standaert, N. Donckers and J.-J. Quisquater, Improved higher-order side-channel attacks with FPGA experiments, in: *CHES*, Lecture Notes in Comput. Sci. 3659, Springer (2005), 309–323.

[30] É. Peeters, F.-X. Standaert and J.-J. Quisquater, Power and electromagnetic analysis: Improved model, consequences and comparisons, *Integr. VLSI J.* **40** (2007), 52–60.

[31] E. Prouff, M. Rivain and R. Bevan, Statistical analysis of second order differential power analysis, *IEEE Trans. Computers* **58** (2009), no. 6, 799–811.

[32] M. Rivain and E. Prouff, Provably secure higher-order masking of AES, in: *CHES*, Lecture Notes in Comput. Sci. 6225, Springer (2010), 413–427.

[33] A. Satoh, S. Morioka, K. Takano and S. Munetoh, A compact Rijndael hardware architecture with S-box optimization, in: *ASIACRYPT*, Lecture Notes in Comput. Sci. 2248, Springer (2001), 239–254.

[34] W. Schindler, K. Lemke and C. Paar, A stochastic model for differential side channel cryptanalysis, in: *CHES*, Lecture Notes in Comput. Sci. 3659, Springer (2005), 30–46.

[35] K. Schramm and C. Paar, Higher order masking of the AES, in: *CT-RSA*, Lecture Notes in Comput. Sci. 3860, Springer (2006), 208–225.

[36] S. Shah, R. Velegalati, J.-P. Kaps and D. Hwang, Investigation of DPA resistance of block RAMs in cryptographic implementations on FPGAs, in: *ReConFig*, IEEE Computer Society (2010), 274–279.

[37] N. J. A. Sloane (Ed.), The on-line encyclopedia of integer sequences, Sequence A008277: Triangle of Stirling numbers of 2nd kind, $S2(n,k)$, $n \geq 1$, $1 \leq k \leq n$, 2009. `http://oeis.org/A008277`.

[38] S. L. Snover, *The uniqueness of the Nordstrom–Robinson and the Golay binary codes*, PhD thesis, Department of Mathematics, Michigan State University, 1973.

[39] F.-X. Standaert, T. Malkin and M. Yung, A unified framework for the analysis of side-channel key recovery attacks, in: *EUROCRYPT*, Lecture Notes in Comput. Sci. 5479, Springer (2009), 443–461.

[40] F.-X. Standaert, É. Peeters, G. Rouvroy and J.-J. Quisquater, An overview of power analysis attacks against field programmable gate arrays, *Proc. IEEE* **94** (2006), no. 2, 383–394.

[41] University of Sydney, Magma Computational Algebra System, `http://magma.maths.usyd.edu.au/magma/`.

[42] S. Vaudenay, On the need for multipermutations: Cryptanalysis of MD4 and SAFER, in: *FSE*, Lecture Notes in Comput. Sci. 1008, Springer (1994), 286–297.

[43] N. Veyrat-Charvillon and F.-X. Standaert, Mutual information analysis: How, when and why?, in: *CHES*, Lecture Notes in Comput. Sci. 5747, Springer (2009), 429–443.

[44] N. Veyrat-Charvillon and F.-X. Standaert, Generic side-channel distinguishers: Improvements and limitations, in: *CRYPTO*, Lecture Notes in Comput. Sci. 6841, Springer (2011), 354–372.

[45] J. Waddle and D. Wagner, Towards efficient second-order power analysis, in: *CHES*, Lecture Notes in Comput. Sci. 3156, Springer (2004), 1–15.

[46] G.-Z. Xiao and J. L. Massey, A spectral characterization of correlation-immune combining functions, *IEEE Trans. Inform. Theory*, **34** (1988), no. 3, 569–571.

Received July 3, 2012; revised May 2, 2014; accepted May 7, 2014.

Author information

Claude Carlet, LAGA, UMR 7539, CNRS, University of Paris XIII and
University of Paris VIII, 2 rue de la liberté, 93526 Saint-Denis Cedex, France.
E-mail: `claude.carlet@gmail.com`

Jean-Luc Danger, TELECOM-ParisTech, Crypto Group, 37/39 rue Dareau,
75634 Paris Cedex 13; and Secure-IC S.A.S., 80 avenue des Buttes de Coësmes,
35700 Rennes, France.
E-mail: `jean-luc.danger@telecom-paristech.fr`

Sylvain Guilley, TELECOM-ParisTech, Crypto Group, 37/39 rue Dareau,
75634 Paris Cedex 13; and Secure-IC S.A.S., 80 avenue des Buttes de Coësmes,
35700 Rennes, France.
E-mail: `sylvain.guilley@telecom-paristech.fr`

Houssem Maghrebi, TELECOM-ParisTech, Crypto Group, 37/39 rue Dareau,
75634 Paris Cedex 13; and MORPHO, 18 chaussée Jules César, 95520 Osny, France.
E-mail: `houssem.maghrebi@telecom-paristech.fr`