# Optimal First-Order Masking with Linear and Non-Linear Bijections

Anonymous submission to AFRICACRYPT'2012

**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 be attacked by a zero-offset second-order CPA attack. The countermeasure can be improved by manipulating the mask through a bijection $F$, aimed at reducing the dependency between the shares. Thus $d$th-order zero-offset attacks, that consist in applying CPA on the $d$th 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 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. Second, we note that for values of $n$ for which non-linear codes exist with better parameters than linear ones. These results are exemplified in the case $n = 8$, 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.

**Keywords:** First-order masking countermeasure (CM), high-order correlation power analysis (HO-CPA), zero-offset HO-CPA, linear and non-linear codes.

## 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. Boolean masking is one answer to secure them, because it can adapted to any function implemented. Early masking schemes involved only one mask per data to protect [24]. Nonetheless, straightforward implementations of this "first-order" countermeasure (CM) happened to be vulnerable to zero-offset "second-order" attacks [27,15]. We call a "first-order" CM an implementation where one single mask protects the sensitive data. 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. Besides, another improvement direction 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. Concretely, a bijection $F$ is applied to the mask, in a view to reduce its dependency with the masked data. The goal of this article is to find bijections $F$ that protect against zero-offset attacks of order $d$ as high as possible.

The rest of the paper is structured as follows. In Sec. 2, the first-order masking scheme that involves the bijection $F$ is described, and its leakage is explained under the Hamming distance model. In Sec. 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 $1, 2, \cdots, d$ is formulated. Based on this formal statement of the problem, optimal solutions for $F$ are researched and given in Sec. 4. The characterization of some optimal bijections $F$ is conducted in Sec. 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 Sec. 6. To ease the reading of the article, some long proofs, secondary results (such as the leakage statistical moments) and some simulation graphs (such as the information leakage in the imperfect model) have been put in appendix. 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

The sensitive variable is noted $x$ and the mask $m$. The two shares manipulated in a Boolean first-order CM are $(x \oplus m, m)$. In the CM we study, a bijection $F$ is applied on the mask share. Thus, the shares are now $(x \oplus m, F(m))$. The schematic of this scheme is illustrated in Fig. 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 each one $n$-bit word. 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 [20]. It is referred to as "tabulated round logic" in Fig. 1. This figure provides with 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. However, in practice, article [14] (resp. [18]) shows how to handle AES substitution box with 4 bit (resp. 2 bit) non-linear data transformations.

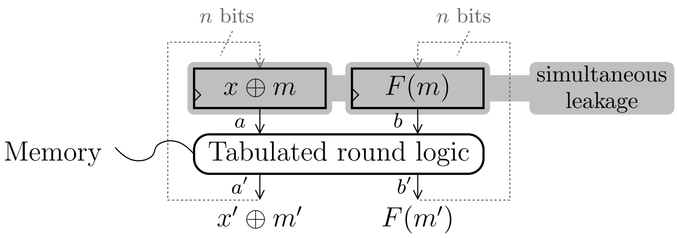



**Fig. 1.** Setup of the first-order masking countermeasure with bijection $F$.

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 with 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
- $b' = F(R(F^{-1}(b)))$.

The detail of the tabulated round logic is represented in Fig. 2.

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 $x$ (plaintext) or the output $x'$ (ciphertext) is known by the attacker. 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 [2,23]. Therefore, the sensitive variable to protect is $x \oplus x'$, noted $z$. The leakage of the studied hardware (Fig. 1) is thus:

$$\begin{aligned} & \mathsf{HD}(x \oplus m, x' \oplus m') + \mathsf{HD}(F(m), F(m')) \\ = \ & \mathsf{HW}(z \oplus m \oplus m') + \mathsf{HW}(F(m) \oplus F(m')) \ . \end{aligned} \tag{1}$$

In this equation, the Hamming distance operator $\mathsf{HD}$ and the Hamming weight operator $\mathsf{HW}$ are defined as $\mathsf{HD}(a, b) = \mathsf{HW}(a \oplus b) \doteq \sum_{i=1}^{n} (a \oplus b)_i$. $F$ is a constant bijection that will contribute to increase the security of the CM. In addition, $F$ is a public information, that we assume known by an attacker.

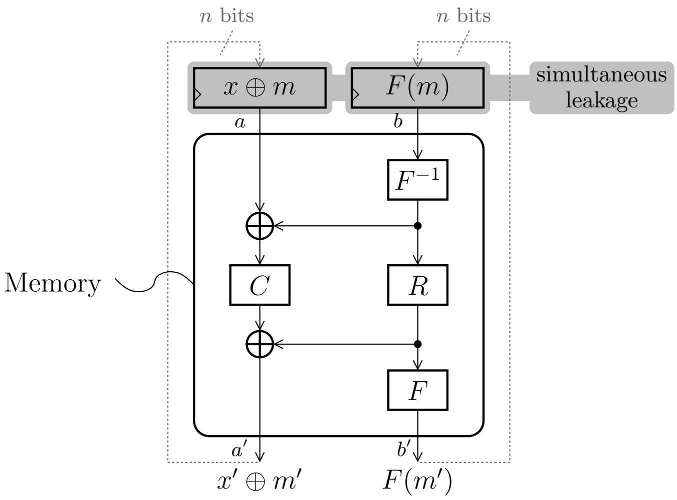



**Fig. 2.** Detail of the function implemented in the tabulated round logic shown in Fig. 1.

# 3 Optimal Function in Zero-Offset $d$th-Order CPA

## 3.1 Optimal Function $f_{\mathrm{opt}}$ Definition

Prouff *et al.* have shown in [17] that an attacker can optimize a CPA [2] against a device leaking $L$ by computing the correlation between the random variables $L$ and $f_{\mathrm{opt}}(Z)$, where $Z$ is the sensitive variable. The function $f_{\mathrm{opt}}(\cdot)$ is called the "optimal function", and is defined as $f_{\mathrm{opt}}(z) = \mathbb{E}[L - \mathbb{E}[L] \mid Z = z]$. In this definition, the capital letters denote random variables, and $\mathbb{E}$ is the expectation operator. If $z \mapsto f_{\mathrm{opt}}(z)$ is constant (*i.e.* $f_{\mathrm{opt}}(Z)$ is deterministic), then [17] 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 Eqn. (1), without $F$ (*i.e.* with $F$ equal to the identity function $\mathsf{Id}$). 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 $f_{\mathrm{opt}}(Z) = \mathbb{E}[\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(M'') - n \mid Z] = 0$, which is deterministic,
- whereas in a zero-offset second-order attack, the attacker uses $f_{\mathrm{opt}}(Z) = \mathbb{E}[(\mathsf{HW}(Z \oplus M'') + \mathsf{HW}(M'') - n)^2 \mid Z] = n - \mathsf{HW}(Z)$, 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 [17, Eqn. (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 [27], 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 $f_{\mathrm{opt}}(Z)$ by having $\mathbb{E}[(\mathsf{HW}(Z \oplus M \oplus M') + \mathsf{HW}(F(M) \oplus F(M')) - n)^d \mid Z]$ 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 this form:

$$\begin{aligned}\mathsf{Term}[p,q](f_{\mathrm{opt}})(z) &\doteq \mathbb{E}[\mathsf{HW}^p[z \oplus M \oplus M'] \times \mathsf{HW}^q[F(M) \oplus F(M')]] \\ &= \mathbb{E}[\mathsf{HW}^p[z \oplus M''] \times \mathsf{HW}^q[F(M) \oplus F(M \oplus M'')]] \ , \quad (2)\end{aligned}$$

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

### 3.2 Condition on $F$ for the Resistance Against 2nd-Order CPA

To resist zero-offset second-order CPA, the term in Eqn. (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 noted $D_{m''}F(m)$. This notion is for instance defined in the Definition 8.2 in §8.2.2 at page 277 of [4]. It can be observed that Eqn. (2) also writes as a convolution product: $\mathsf{Term}[p,q](f_{\mathrm{opt}})(z) = \frac{1}{2^n}\left(\mathsf{HW} \otimes \mathbb{E}[\mathsf{HW}(D_{(\cdot)}F(M))]\right)(z)$. An appealing property of the Walsh-Hadamard transform is that it turns a convolution into a product. So, we have:

$$\begin{aligned} f_{\mathrm{opt}}(z) = \mathrm{cst} &\iff \widehat{f_{\mathrm{opt}}}(a) \propto \delta(a) \quad \begin{array}{l} \text{// where } \propto \text{ means "is proportional to"} \\ \text{// and } \delta(\cdot) \text{ is the Kronecker symbol.} \end{array} \\ &\iff \widehat{\mathsf{HW}}(a) \times \mathbb{E}[\widehat{\mathsf{HW} \circ D_{(\cdot)}}F(M)](a) = \left(n \times 2^{n-1}\right)^2 \times \delta(a) \\ &\iff \forall a \neq 0, \widehat{\mathsf{HW}}(a) = 0 \ \text{ or } \ \mathbb{E}[\widehat{\mathsf{HW} \circ D_{(\cdot)}}F(M)](a) = 0 \,. \quad (3)\end{aligned}$$

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:

$$
\begin{aligned}
& \mathbb{E}[\widehat{\mathsf{HW} \circ D_{(\cdot)}} F(M)](0) \\
&= \textstyle\sum_z \mathbb{E}[\mathsf{HW}(D_z F(M))(-1)^{0 \cdot z}] \\
&= \mathbb{E}[\textstyle\sum_z \mathsf{HW}(F(M) \oplus F(M \oplus z))] \\
&= \mathbb{E}[\textstyle\sum_{z'} \mathsf{HW}(z')] \quad // \text{ Because } \forall m, z \mapsto F(m) \oplus F(m \oplus z) \text{ is bijective} \\
&= \mathbb{E}[\tfrac{n}{2} 2^n] = \tfrac{n}{2} 2^n \ .
\end{aligned}
$$

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

$$
\begin{aligned}
\widehat{\mathsf{HW}}(a) &= \sum_z \frac{1}{2} \sum_{i=1}^{n} \left(1 - (-1)^{z_i}\right) (-1)^{a \cdot z} \\
&= n \cdot 2^{n-1} \delta(a) - \tfrac{1}{2} \textstyle\sum_z \sum_{i=1}^{n} (-1)^{(a \oplus e_i) \cdot z} \\
&= \begin{cases} n \cdot 2^{n-1} & \text{if } a = 0, \\ -2^{n-1} & \text{if } \exists i \in [\![1, n]\!], \text{ such that } a = e_i, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{4}
$$

Thus, the problem comes down to finding a function $F$ such that: $\mathbb{E}[\widehat{\mathsf{HW} \circ D_{(\cdot)}} F(M)](a) = 0$ for all $a = e_i$. This condition rewrites:

$$
\forall a = e_i, \quad \sum_{z,m} \mathsf{HW}(F(m) \oplus F(m \oplus z))(-1)^{a \cdot z} = 0 \ . \tag{5}
$$

Let $a \neq 0$. Then:

$$
\begin{aligned}
& \textstyle\sum_{z,m} \mathsf{HW}(F(m) \oplus F(m \oplus z))(-1)^{a \cdot z} \\
&= \textstyle\sum_{z,m} \frac{1}{2} \sum_{i=1}^{n} \left(1 - (-1)^{F_i(m) \oplus F_i(m \oplus z)}\right) (-1)^{a \cdot z} \\
&= \cancel{n 2^{2n-1} \delta(a)} - \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \sum_{z,m} (-1)^{F_i(m) \oplus F_i(m \oplus z) \oplus a \cdot z} \\
&= -\tfrac{1}{2} \textstyle\sum_{i=1}^{n} \sum_m (-1)^{F_i(m)} \sum_z (-1)^{a \cdot z \oplus F_i(m \oplus z)} \\
&= -\tfrac{1}{2} \textstyle\sum_{i=1}^{n} \sum_m (-1)^{F_i(m)} \sum_z (-1)^{a \cdot (z \oplus m) \oplus F_i(z)} \quad // \ z \leftarrow z \oplus m \\
&= -\tfrac{1}{2} \textstyle\sum_{i=1}^{n} \sum_m (-1)^{a \cdot m \oplus F_i(m)} \sum_z (-1)^{a \cdot z \oplus F_i(z)} \\
&= -\tfrac{1}{2} \textstyle\sum_{i=1}^{n} \left(\sum_m (-1)^{a \cdot m \oplus F_i(m)}\right)^2 \\
&= -\tfrac{1}{2} \textstyle\sum_{i=1}^{n} \left(\widehat{(-1)^{F_i}}(a)\right)^2 \ .
\end{aligned}
$$

Thus, this quantity is null if and only if $\forall i \in [\![1, n]\!]$, $\widehat{(-1)^{F_i}}(a) = 0$. Thus, if we generalize the Walsh-Hadamard transform on vectorial Boolean functions (by applying the transformation component-wise), and use the notation $f_\chi$ for the sign function of $f$ (also component-wise), then Eqn. (5) is equivalent to: $\forall a = e_i, \widehat{F_\chi}(a) = 0$. Now, as $F$ is balanced (since bijective), this equality also holds for

$a = 0$. This means that every coordinate of $F$ is 1-resilient. Constructions exist, as explained in [4, Sec. 8.7].

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

### 3.3 Condition on $F$ for the Resistance Against $d$th-Order CPA

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

**Theorem 1.** *Let $P$ and $Q$ be two positive integers, and $F$ a bijection of $\mathbb{F}_2^n$.*
*Eqn. (2) is constant for all $p \in [\![0, P]\!]$ and $q \in [\![0, Q]\!]$ if and only if:*

$$\forall a, b \in \mathbb{F}_2^n, 0 < \mathsf{HW}(a) \leq P, 0 \leq \mathsf{HW}(b) \leq Q, \quad \widehat{(b \cdot F)}_\chi(a) = 0 \ . \tag{6}$$

**Proposition 1.** *An $(n, m)$-function is defined as a vectorial Boolean function from $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$. The condition expressed in Eqn. (6) of theorem 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 Eqn. (6)

In this section, we find bijections that meet Eqn. (6).

The condition expressed in Eqn. (6) for theorem 1 rewrites: $\forall b \in \mathbb{F}_2^{n*} \doteq \mathbb{F}_2^n\backslash\{0\}$ and $\forall a \in \mathbb{F}_2^n$, if $\mathsf{HW}(a) \leq d - \mathsf{HW}(b)$ then $\widehat{(b \cdot F)}_\chi(a) = 0$.

### 4.1 Optimal Linear Bijections

$F$ can be chosen linear. All linear $(n, n)$-functions write $F(x) = (x \cdot v_1, \cdots, x \cdot v_n)$, where $v_i$ are elements of $\mathbb{F}_2^n$. $F$ is bijective if and only if $(v_1, \cdots, v_n)$ is a basis of $\mathbb{F}_2^n$. We have:

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

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

$$\forall b \neq 0, \quad \mathsf{HW}(\textstyle\bigoplus_{i=1}^n b_i v_i) > d - \mathsf{HW}(b) \ . \tag{7}$$

We notice that the set of ordered pairs $\{(b, \bigoplus_{i=1}^{n} b_i v_i), b \in \mathbb{F}_2^n\}$ forms a vector subspace of $\mathbb{F}_2^{2n}$. Therefore, it defines a $[2n, n, \delta]$ binary linear code, where $\delta$ is its minimum distance. Because of Eqn. (7), the necessary and sufficient condition becomes $\delta > d$. Reciprocally, a $[2n, n, \delta]$ binary linear code (modulo a permutation of its coordinates) can be spawned by a generator matrix $(I_n \quad G)$, where $G$ is an $n \times n$ matrix. This representation is the systematic form of the code; such form is discussed on the $n = 8$ case-study in Appendix B.

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 (refer for instance to [11]); corresponding codes are called "optimal". For some practical values of $n$, they are recalled in Tab. 1.

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

| Sboxes of algorithm | DES | n/a | n/a | n/a | AES |
|---|---|---|---|---|---|
| $2n$ | 8 | 10 | 12 | 14 | 16 |
| $\delta_{\max}(n)$ | 4 | 4 | 4 | 4 | 5 |

Thus, the best achievable $d$ using a linear bijection $F$ is $\delta_{\max}(n) - 1$. In particular, this result proves that with linear $F$, it is possible to protect:

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

### 4.2 Optimal Non-Linear Bijections

Under some circumstances, a non-linear bijection $F$ allows to reach better performances. The condition on $F$ (Eqn. (6)) is equivalent to saying that its graph indicator is $d$-th order correlation immune [3]. Given any $(n, n)$-function $F$, let $C = \{(x, F(x)), x \in \mathbb{F}_2^n\}$. The weight enumerator $W_C(X, Y)$ and distance enumerator $D_C(X, Y)$ of this code are:

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

We have $W_C(X+Y, X-Y) = \sum_{a,b \in \mathbb{F}_2^n} \left( \sum_{x \in \mathbb{F}_2^n} (-1)^{b \cdot F(x) + a \cdot x} \right) X^{2n - \mathsf{HW}(a,b)} Y^{\mathsf{HW}(a,b)}$ and $D_C(X+Y, X-Y) = \frac{1}{|C|} \sum_{a,b \in \mathbb{F}_2^n} \left( \sum_{x \in \mathbb{F}_2^n} (-1)^{b \cdot F(x) \oplus a \cdot x} \right)^2 X^{2n - \mathsf{HW}(a,b)} Y^{\mathsf{HW}(a,b)}$. Hence $d + 1$ is exactly the minimum value of the nonzero exponents of $Y$ with nonzero coefficients in $D_C(X + Y, X - Y)$, called the dual distance of $C$ in the sense of Delsarte [6,12].

There is no non-linear code for $n = 4$ that has a better minimal distance than linear codes, but there are some for $n = 8$. A non-linear optimal code for $n = 8$ is the Nordstrom-Robinson $(16, 256, 6)$ code (see more in [5]). With these parameters, this code is unique and coincides with Preparata and Kerdock codes [21]. Some codewords, as obtained from Golay code in standard form [9], are listed in Tab. 2.

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

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

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 is present (at least) twice in the first half (from the high byte of codewords $x = 3$ and $x = 7$),
- and 00000000 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 in two bijections $F_{\text{high}}$ and $F_{\text{low}}$ of $\mathbb{F}_2^8$. This means that if we note $x \in \mathbb{F}_2^8$ the codewords index in Tab. 2, the Nordstrom-Robinson $(16, 256, 6)$ code writes as $F_{\text{high}}(x) \,||\, F_{\text{low}}(x)$. The codewords can be reordered according to the first column, so that the code rewrites $x \,||\, F_{\text{low}}(F_{\text{high}}^{-1}(x))$ [5]. So the bijection $F$ can be chosen equal to $F = F_{\text{low}} \circ F_{\text{high}}^{-1}$. For example, when $F_{\text{high}}$ consists in bits $[\![15, 9]\!] \cup \{7\}$ of the code (and $F_{\text{low}}$ in bits $\{8\} \cup [\![6, 0]\!]$), $F$ takes the values tabulated as follows: $\{F(x), x \in \mathbb{F}_2^8\} = \{$`0x00`, `0xb3`, `0xe5`, `0x6a`, `0x2f`, `0xc6`, `0x5c`, `0x89`, `0x79`, `0xac`, `0x36`, `0xdf`, `0x9a`, `0x15`, `0x43`, `0xf0`, `0xcb`, `0x1e`, `0xb8`, `0x51`, `0x72`, `0xfd`, `0x97`, `0x24`, `0xd4`, `0x67`, `0x0d`, `0x82`, `0xa1`, `0x48`, `0xee`, `0x3b`, `0x9d`, `0x74`, `0xd2`, `0x07`, `0xe8`, `0x5b`, `0x31`, `0xbe`, `0x4e`, `0xc1`, `0xab`, `0x18`, `0xf7`, `0x22`, `0x84`, `0x6d`, `0xa6`, `0x29`, `0x7f`, `0xcc`, `0x45`, `0x90`, `0x0a`, `0xe3`, `0x13`, `0xfa`, `0x60`, `0xb5`, `0x3c`, `0x8f`, `0xd9`, `0x56`, `0x57`, `0xd8`, `0x8e`, `0x3d`, `0xb4`, `0x61`, `0xfb`, `0x12`, `0xe2`, `0x0b`, `0x91`, `0x44`, `0xcd`, `0x7e`, `0x28`, `0xa7`, `0x6c`, `0x85`, `0x23`, `0xf6`, `0x19`, `0xaa`, `0xc0`, `0x4f`, `0xbf`, `0x30`, `0x5a`, `0xe9`, `0x06`, `0xd3`, `0x75`, `0x9c`, `0x3a`, `0xef`, `0x49`, `0xa0`, `0x83`, `0x0c`, `0x66`, `0xd5`, `0x25`, `0x96`, `0xfc`, `0x73`, `0x50`, `0xb9`, `0x1f`, `0xca`, `0xf1`, `0x42`, `0x14`, `0x9b`, `0xde`, `0x37`, `0xad`, `0x78`, `0x88`, `0x5d`, `0xc7`, `0x2e`, `0x6b`, `0xe4`, `0xb2`, `0x01`, `0xfe`, `0x4d`, `0x1b`, `0x94`, `0xd1`, `0x38`, `0xa2`, `0x77`, `0x87`, `0x52`, `0xc8`, `0x21`, `0x64`, `0xeb`, `0xbd`, `0x0e`, `0x35`, `0xe0`, `0x46`, `0xaf`, `0x8c`, `0x03`, `0x69`, `0xda`, `0x2a`, `0x99`, `0xf3`, `0x7c`, `0x5f`, `0xb6`, `0x10`, `0xc5`, `0x63`, `0x8a`, `0x2c`, `0xf9`, `0x16`, `0xa5`, `0xcf`, `0x40`, `0xb0`, `0x3f`, `0x55`, `0xe6`, `0x09`, `0xdc`, `0x7a`, `0x93`, `0x58`, `0xd7`, `0x81`, `0x32`, `0xbb`, `0x6e`, `0xf4`, `0x1d`, `0xed`, `0x04`, `0x9e`, `0x4b`, `0xc2`, `0x71`, `0x27`, `0xa8`, `0xa9`, `0x26`, `0x70`, `0xc3`, `0x4a`, `0x9f`, `0x05`, `0xec`, `0x1c`, `0xf5`, `0x6f`, `0xba`, `0x33`, `0x80`, `0xd6`, `0x59`, `0x92`, `0x7b`, `0xdd`, `0x08`, `0xe7`, `0x54`, `0x3e`, `0xb1`, `0x41`, `0xce`, `0xa4`, `0x17`, `0xf8`, `0x2d`, `0x8b`, `0x62`, `0xc4`, `0x11`, `0xb7`, `0x5e`, `0x7d`, `0xf2`, `0x98`, `0x2b`, `0xdb`, `0x68`, `0x02`, `0x8d`, `0xae`, `0x47`, `0xe1`, `0x34`, `0x0f`, `0xbc`, `0xea`, `0x65`, `0x20`, `0xc9`, `0x53`, `0x86`, `0x76`, `0xa3`, `0x39`, `0xd0`, `0x95`, `0x1a`, `0x4c`, `0xff`}.

Thus byte-oriented cryptographic implementations can be protected with this code against all zero-offset HO-CPA of order $d \leq 5$.

## 5 Security and Leakage Evaluations of the Optimal Linear and Non-Linear Bijections

As argued in [22], the robustness evaluation of a CM encompasses two dimensions: its resistance to specific attacks, and its amount of leakage irrespective of 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 [26]. Therefore, in addition to a security evaluation conducted in Sec. 5.1, we will also estimate the leakage of the CM in Sec. 5.2.

### 5.1 Verification of the Security for $n = 8$

In this section, we illustrate the efficiency of the identified bijection from an 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 $f_{\mathrm{opt}}(z)$ for the centered leakage raised at power $1 \leq d \leq 6$ for four linear bijections (noted $F1$, $F2$, $F3$ and $F4$) and the non-linear bijection given in Sec. 4.2 (noted $F5$). 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, 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:

$$G2 = \begin{pmatrix} 0&0&0&0&0&1&1&1 \\ 0&0&0&1&1&0&1&1 \\ 1&0&1&0&0&0&1&1 \\ 0&0&1&1&1&0&0&0 \\ 1&1&1&0&0&0&0&0 \\ 1&1&1&1&1&1&1&1 \\ 1&0&1&0&1&1&0&0 \\ 0&1&0&1&0&1&1&0 \end{pmatrix}, G3 = \begin{pmatrix} 1&0&0&0&1&1&1&1 \\ 1&1&0&0&0&1&1&1 \\ 1&1&1&0&0&0&1&1 \\ 1&1&1&1&0&0&0&1 \\ 1&1&1&1&1&0&0&0 \\ 0&1&1&1&1&1&0&0 \\ 0&0&1&1&1&1&1&0 \\ 0&0&0&1&1&1&1&1 \end{pmatrix}, G4 = \begin{pmatrix} 1&0&0&1&1&1&1&0 \\ 0&1&0&0&1&1&1&1 \\ 1&1&0&0&1&1&0&0 \\ 0&1&1&0&0&1&1&0 \\ 0&0&1&1&0&0&1&1 \\ 1&1&1&1&0&0&1&0 \\ 0&1&1&1&1&0&0&1 \\ 1&1&0&1&0&1&1&1 \end{pmatrix}.$$

It can be checked that they are invertible. Their inverses are:

$$G2^{-1} = \begin{pmatrix} 0&1&1&1&0&0&0&0 \\ 1&1&1&0&0&1&0&0 \\ 1&0&0&1&1&1&0&0 \\ 0&0&1&0&1&1&1&0 \\ 1&0&1&0&0&0&1&0 \\ 0&1&0&0&1&1&0&0 \\ 1&0&0&0&0&1&1&1 \\ 0&1&0&0&1&0&1&1 \end{pmatrix}, G3^{-1} = \begin{pmatrix} 0&0&1&0&0&1&0&1 \\ 1&0&0&1&0&0&1&0 \\ 0&1&0&0&1&0&0&1 \\ 1&0&1&0&0&1&0&0 \\ 0&1&0&1&0&0&1&0 \\ 0&0&1&0&1&0&0&1 \\ 1&0&0&1&0&1&0&0 \\ 0&1&0&0&1&0&1&0 \end{pmatrix}, G4^{-1} = \begin{pmatrix} 1&1&1&0&1&0&1&1 \\ 1&0&0&1&1&1&1&0 \\ 0&1&0&0&1&1&1&1 \\ 1&1&0&0&1&1&0&0 \\ 0&1&1&0&0&1&1&0 \\ 0&0&1&1&0&0&1&1 \\ 1&1&1&1&0&0&1&0 \\ 0&1&1&1&1&0&0&1 \end{pmatrix}.$$

Table 4, in Appendix C, reports some values of the optimal functions. The lines represented in gray are those for which the $f_{\mathrm{opt}}(z)$ 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 $d$-th 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 5-th order zero-offset HO-CPA, as announced in Sec. 4.2.

### 5.2 Verification of the Leakage of the Identified Bijections

As a complement to the security analysis carried out in Sec. 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 $\sigma^2$. The result of the MIM computation is shown in Fig. 3.

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 Fig. 2) the relevance of choosing $F$ based on a HO-CPA criterion.

Furthermore, Fig. 3 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 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 $\sigma$.

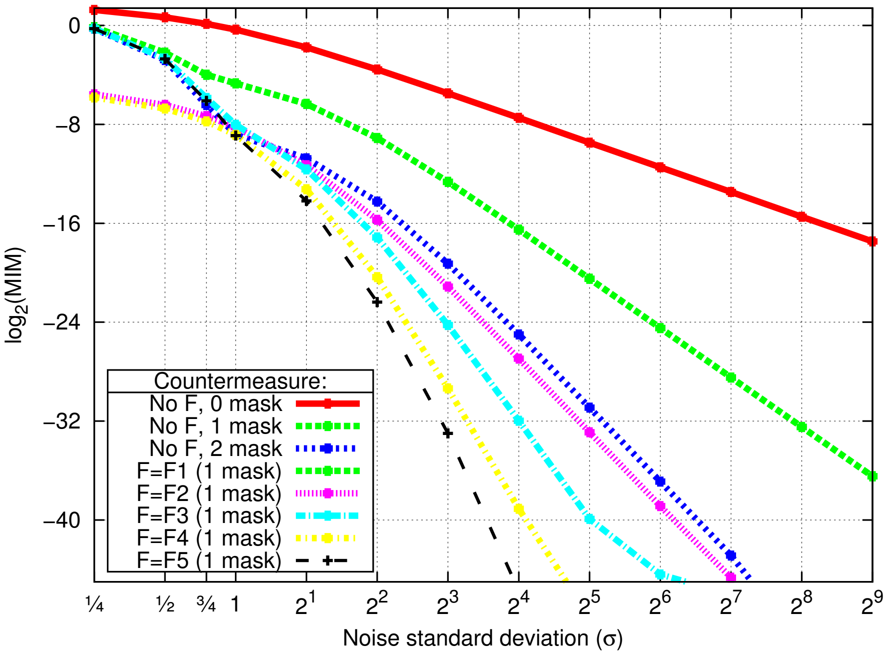



**Fig. 3.** Mutual information of the leakage with the sensitive variable $Z$ for $n = 8$ bit.

### 5.3 Results in Imperfect Models

Masking schemes randomize more or less properly the leakage. In the straightforward example studied in this paper (Eqn. (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: $\forall M'' \in \mathbb{F}_2^n, \mathsf{HW}(\texttt{0xff} \oplus M'') + \mathsf{HW}(M'') = \mathsf{HW}(\overline{M''}) + \mathsf{HW}(M'') = n$. 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 [13]) values of the sensitive data. The working factor of is 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 assumed 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 [25]. Also, it has been shown that with specific side-channel capturing systems the attacker can distort the measurement. For instance, in [16], the authors 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 $\mathbb{R}[X_1, \cdots, X_n, X'_1, \cdots, X'_n]$ of degree less 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 following form:

$$L \doteq P(X_1, \cdots, X_n, X'_1, \cdots, X'_n) = \sum_{\substack{(u,v)\in\mathbb{F}_2^n\times\mathbb{F}_2^n, \\ \mathsf{HW}(u)+\mathsf{HW}(v)\leq\tau}} A_{(u,v)} \cdot \prod_{i=1}^{n} X_i^{u_i} {X'}_i^{v_i} \ , \tag{8}$$

where the $A_{(u,v)}$ are real coefficients. This leakage formulation is similar to that of the high-order stochastic model [19]. For example, it is shown in [17, Eqn. (3)] that $P(X_1, \cdots, X_n, X'_1, \cdots, X'_n)$ is equal to $\mathsf{HW}(X \oplus X')$ when the coefficients $A_{(u,v)} \doteq a^{\mathrm{HD}}_{(u,v)}$ satisfy:

$$a^{\mathrm{HD}}_{(u,v)} = \begin{cases} +1 & \text{if } \mathsf{HW}(u) + \mathsf{HW}(v) = 1 \,, \\ -2 & \text{if } \mathsf{HW}(u) = 1 \text{ and } v = u \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{9}$$

In the following experiments, we compute the mutual information between $L$ and $Z = X \oplus X'$ when $\tau \leq 2$ and when the coefficients $A_{(u,v)}$ deviate randomly from those of (9) 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:

$$\begin{array}{lll} A^{\mathrm{HD}}_{(u,v)} & \sim \ a^{\mathrm{HD}}_{(u,v)} & + \mathcal{U}([-\frac{\delta}{2}, +\frac{\delta}{2}]) \ , \\ A^{\mathrm{NULL}}_{(u,v)} & \sim \ 0 & + \mathcal{U}([-\frac{\delta}{2}, +\frac{\delta}{2}]) \ . \end{array} \tag{10}$$

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 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 ($\mathsf{Id}$), that acts as a reference,
$F2'$ : one bijection that cancels the first-order leakage but not the second-order,
$F3'$ : another that cancels both first- and second-orders.

They are linear, *i.e.* write $Fi'(x) = Gi' \times x$, where the generating matrix $Gi'$ are given below:

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

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

The results are plotted in Tab. 5, 6 & 7 for the randomized HD model and in Tab. 8, 9 & 10 for the randomized "NULL" model.

In Tab. 5, 6 & 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 $F1'$),
- and remains low, irrespective of $\delta$.

The average leakage is unchanged, and the leakage values are simply getting slightly scattered. The reason for this resilience comes from the rationale of the 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 (Eqn. (10)), 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[1] $\delta^2/12$ of the model).

It is interesting to contrast the leakage squeezing with the first-order leak-free CM presented in [13]. This CM aims at leaking no information when the HD leakage model is perfect. A study for model imperfection has also been conducted (see right column of Tab. 5, 6 & 7). It appears that this CM is much less robust to deviation from the ideal model. Indeed, the working factor of the CM is to have one share leak 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 $\delta > 50\%$).

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

For the sake of comparison, we also computed the same curves when the unnoised model is a constant one (called "NULL" model in Eqn. (10)). The simulation results are shown in Tab. 8, 9 & 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 [7]. In particular, it appears that our CM continues to work ($F3$ leaks less than $F2$, that leaks less than $F1$), at least for large enough noise standard deviations $\sigma$. 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 $\tau$ can be studied. Results are computed for $\tau$ in $\{1, 2, 3\}$. In all the cases, $\tau$ does not impact the general conclusions.

Regarding the deviation from the HD model, the greater the multivariate degree $\tau$, the more possible deviations from the genuine ideal model. Indeed, the number of random terms in Eqn. (8) is increasing with $\tau$ (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 unchanged. In the Null model, the greater $\tau$, the less singularities in the leakage. This explains why the mutual information curves get smoother despite the additional noise. But with the greater $\tau$, the more leaking sources (because the more non-zero terms in the polynomial), which explains why the leaked mutual information increases in average with $\tau$.

## 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 [10]. In this article, we focus on a 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 $d \in \mathbb{N}^*$. We show that by storing $F(m)$ (the image of $m$ by a bijection $F$) instead of $m$ in the mask register, the highest order $d$ of a successful zero-offset attack can be increased significantly. 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 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 $F$ protect against zero-offset HO-CPA of orders 1, 2, 3, 4 and 5. 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).

## References


1. Lejla Batina, Benedikt Gierlichs, Emmanuel Prouff, Matthieu Rivain, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. Mutual Information Analysis: a Comprehensive Study. *J. Cryptology*, 24(2):269–291, 2011.
2. Éric Brier, Christophe Clavier, and Francis Olivier. Correlation Power Analysis with a Leakage Model. In *CHES*, volume 3156 of *LNCS*, pages 16–29. Springer, August 11–13 2004. Cambridge, MA, USA.
3. Paul Camion, Claude Carlet, Pascale Charpin, and Nicolas Sendrier. On Correlation-Immune Functions. In Joan Feigenbaum, editor, *CRYPTO*, volume 576 of *Lecture Notes in Computer Science*, pages 86–100. Springer, 1991.
4. Claude Carlet. Boolean Functions for Cryptography and Error Correcting Codes: Chapter of the monography Boolean Models and Methods in Mathematics, Computer Science, and Engineering. pages 257–397. Cambridge University Press, Y. Crama and P. Hammer eds, 2010.
5. Claude Carlet, Philippe Gaborit, Jon-Lark Kim, and Patrick Solé. A new class of codes for Boolean masking of cryptographic computations, October 6 2011. http://arxiv.org/abs/1110.1193.
6. Philippe Delsarte. *An algebraic approach to the association schemes of coding theory*. PhD thesis, Université Catholique de Louvain, Belgium, 1973.
7. Julien Doget, Emmanuel Prouff, Matthieu Rivain, and François-Xavier Standaert. Univariate side channel attacks and leakage modeling. *J. Cryptographic Engineering*, 1(2):123–144, 2011.
8. Neil J. A. Sloane (Ed.). The On-Line Encyclopedia of Integer Sequences, published electronically at http://www.research.att.com/~njas/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.
9. G. David Forney, Jr., Neil J. A. Sloane, and Mitchell D. Trott. The Nordstrom-Robinson Code is the Binary Image of the Octacode. In R. Calderbank Amer. Math. Soc., Jr. G. D. Forney, and N. Moayeri (Eds), editors, *Coding and Quantization: DIMACS / IEEE Workshop*, pages 19–26, October 19-21 1992.
10. Louis Goubin and Jacques Patarin. DES and Differential Power Analysis. The "Duplication" Method. In *CHES*, LNCS, pages 158–172. Springer, Aug 1999. Worcester, MA, USA.
11. T.Aaron Gulliver and Patric R.J. Östergård. Binary optimal linear rate 1/2 codes. *Discrete Mathematics*, 283(1-3):255 – 261, 2004.
12. F. Jessie MacWilliams and Neil J. A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier, Amsterdam, North Holland, 1977. ISBN: 978-0-444-85193-2.
13. Houssem Maghrebi, Emmanuel Prouff, Sylvain Guilley, and Jean-Luc Danger. A First-Order Leak-Free Masking Countermeasure. In *CT-RSA*, LNCS. Springer, February 27 – March 2 2012. San Francisco, CA, USA.
14. 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 $rmGF(2^4)^2$ Composite-Field AES-Encrypt/Decrypt Accelerator for Content-Protection in 45 nm High-Performance Microprocessors. *Solid-State Circuits, IEEE Journal of*, 46(4):767 –776, april 2011.
15. Éric Peeters, François-Xavier Standaert, Nicolas Donckers, and Jean-Jacques Quisquater. Improved Higher-Order Side-Channel Attacks With FPGA Experiments. In *CHES*, volume 3659 of *LNCS*, pages 309–323. Springer-Verlag, 2005. Edinburgh, UK.

16. Éric Peeters, François-Xavier Standaert, and Jean-Jacques Quisquater. Power and electromagnetic analysis: Improved model, consequences and comparisons. *Integration, The VLSI Journal, special issue on* "Embedded Cryptographic Hardware", 40:52–60, January 2007. DOI: 10.1016/j.vlsi.2005.12.013.
17. Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Trans. Computers*, 58(6):799–811, 2009.
18. Akashi Satoh, Sumio Morioka, Kohji Takano, and Seiji Munetoh. A Compact Rijndael Hardware Architecture with S-Box Optimization. In Colin Boyd, editor, *ASIACRYPT*, volume 2248 of *Lecture Notes in Computer Science*, pages 239–254. Springer, 2001.
19. Werner Schindler, Kerstin Lemke, and Christof Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In LNCS, editor, *CHES*, volume 3659 of *LNCS*, pages 30–46. Springer, Sept 2005. Edinburgh, Scotland, UK.
20. Shaunak Shah, Rajesh Velegalati, Jens-Peter Kaps, and David Hwang. Investigation of DPA Resistance of Block RAMs in Cryptographic Implementations on FPGAs. In Viktor K. Prasanna, Jürgen Becker, and René Cumplido, editors, *ReConFig*, pages 274–279. IEEE Computer Society, 2010.
21. Stephen L. Snover. *The uniqueness of the Nordstrom-Robinson and the Golay binary codes*. PhD thesis, Department of Mathematics, Michigan State University, USA, 1973.
22. François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In *EUROCRYPT*, volume 5479 of *LNCS*, pages 443–461. Springer, April 26-30 2009. Cologne, Germany.
23. François-Xavier Standaert, Éric Peeters, Gaël Rouvroy, and Jean-Jacques Quisquater. An Overview of Power Analysis Attacks Against Field Programmable Gate Arrays. *Proceedings of the IEEE*, 94(2):383–394, February 2006. (Invited Paper).
24. François-Xavier Standaert, Gaël Rouvroy, and Jean-Jacques Quisquater. FPGA Implementations of the DES and Triple-DES Masked Against Power Analysis Attacks. In *FPL*. IEEE, August 2006. Madrid, Spain.
25. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual Information Analysis: How, When and Why? In *CHES*, volume 5747 of *LNCS*, pages 429–443. Springer, September 6-9 2009. Lausanne, Switzerland.
26. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Generic Side-Channel Distinguishers: Improvements and Limitations. In Phillip Rogaway, editor, *CRYPTO*, volume 6841 of *Lecture Notes in Computer Science*, pages 354–372. Springer, 2011.
27. Jason Waddle and David Wagner. Towards Efficient Second-Order Power Analysis. In *CHES*, volume 3156 of *LNCS*, pages 1–15. Springer, 2004. Cambridge, MA, USA.


# A Proof of Theorem 1

## A.1 First Intermediate Result for the Proof of Theorem 1

**Theorem 2.** $\forall a \in \mathbb{F}_2^n, \forall p \in \mathbb{N}, \quad \widehat{\mathsf{HW}^p}(a) = 0 \iff \mathsf{HW}(a) > p.$

Let us define the function $H(n,p,h) \doteq \sum_{z\in\mathbb{F}_2^n} \mathsf{HW}^p(z)(-1)^{z\cdot\bigoplus_{i=1}^{h} e_i}$, for $n \in \mathbb{N}^*$, $p \in \mathbb{N}$ and $h \in [\![0,n]\!]$. It is tabulated for $n = 4$ in Tab. 3. The value $H(n,n,n)$, indicated by dagger sign (*i.e.* "†") in the table, is equal to $(-1)^n n!$.

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

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

As the order of the bits of the dummy variable $z$ are 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 1.**

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

*Proof (of Lem. 1.).*

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

where $\left\{ \begin{matrix} p \\ n \end{matrix} \right\}$ is a Stirling number of the second kind [8]. More precisely, it is the number of ways of partitioning a set of $p$ elements into $n$ nonempty sets. Consequently, $\left\{ \begin{matrix} p \\ n \end{matrix} \right\} = 0$ if $n > p$, because otherwise at least one set would be empty. Also, $\left\{ \begin{matrix} p \\ n \end{matrix} \right\} > 0$ if $n \leq p$. Now, the sign of $H(n, p, n)$ depends on the parity of $n$ if $n \leq p$. It is positive (resp. negative) if $n$ is even (resp. odd). $\square$

**Lemma 2.**

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

*Proof (of Lem. 2.).* This lemma has already been proved in Lem. 1 if $h = n$. Thus, we assume in the remainder of this proof that $h < n$. For $z \in \mathbb{F}_2^n$, we note

$z = (z_L, z_H)$, where $z_L \in \mathbb{F}_2^h$ and $z_H \in \mathbb{F}_2^{n-h}$.

$$\begin{aligned}
H(n,p,h) &= \sum_{(z_L,z_H)} \mathsf{HW}^p((z_L,0) \oplus (0,z_H))(-1)^{\left(z_L \cdot \oplus_{i=1}^h e_i\right) \oplus (z_H \cdot 0)} \\
&= \sum_{(z_L,z_H)} (\mathsf{HW}(z_L) + \mathsf{HW}(z_H))^p (-1)^{z_L \cdot \oplus_{i=1}^h e_i} \\
&= \sum_{(z_L,z_H)} \sum_{j=0}^{p} \binom{p}{j} \times \mathsf{HW}^j(z_L) \times \mathsf{HW}^{p-j}(z_H)(-1)^{z_L \cdot \oplus_{i=1}^h e_i} \\
&= \sum_{j=0}^{p} \binom{p}{j} \sum_{z_L} \mathsf{HW}^j(z_L)(-1)^{z_L \cdot \oplus_{i=1}^h e_i} \times \sum_{z_H} \mathsf{HW}^{p-j}(z_H) \\
&= \sum_{j=0}^{p} \binom{p}{j} \times H(h,j,h) \times H(n-h,p-j,0) \ . 
\end{aligned} \tag{11}$$

Now, given Lem. 1, $\forall j < h$, $H(h,j,h) = 0$. Thus, if $p < h$, then all the terms $H(h,j,h)$ involved in Eqn. (11) 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
- or all strictly negative if $h$ is odd.

Hence, so is the sum in Eqn. (11). $\square$

*Proof (of theorem. 2.).* As already noticed, $\widehat{\mathsf{HW}^p}(a) = H(n,p,\mathsf{HW}(a))$. According to Lem. 2, this quantity is null if and only if $p < \mathsf{HW}(a)$. $\square$

### A.2 Second Intermediate Result for the Proof of Theorem 1

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

$$\begin{aligned}
\left(\sum_{i=1}^{n} (-1)^{X \cdot e_i}\right)^j &= \sum_{i_1,\cdots,i_j \in [\![1,n]\!]^j} \prod_{l=1}^{j} (-1)^{X \cdot e_{i_l}} \\
&= \sum_{i_1,\cdots,i_j \in [\![1,n]\!]^j} (-1)^{X \cdot \oplus_{l=1}^j e_{i_l}} \quad \left\{ \begin{array}{l} \text{Under this form,} \\ \text{some terms appear} \\ \text{multiple times.} \end{array} \right. \\
&= \sum_{k_1+\cdots+k_n=j} \binom{j}{k_1,\cdots,k_n} (-1)^{X \cdot (\oplus_{i=1}^n k_i e_i)} \ ,
\end{aligned} \tag{12}$$

where each vector $k_i e_i$ in $\bigoplus_{i=1}^n k_i e_i$ is either $e_i$ if $k_i$ is odd or 0 otherwise. In the Eqn. (12), the term $\binom{j}{k_1,\cdots,k_n}$ is a multinomial coefficient.

Then:

$$
\begin{aligned}
&\sum_{z,m} \mathsf{HW}^q(F(m) \oplus F(m \oplus z))(-1)^{a \cdot z} \\
&= \frac{1}{2^q} \sum_{z,m} \left( n - \sum_{i=1}^{n} (-1)^{F_i(m) \oplus F_i(m \oplus z)} \right)^q (-1)^{a \cdot z} \\
&= \frac{1}{2^q} \sum_{z,m} \sum_{j=0}^{q} \binom{q}{j} n^{q-j} (-1)^j \left( \sum_{i=1}^{n} (-1)^{F_i(m) \oplus F_i(m \oplus z)} \right)^j (-1)^{a \cdot z} \\
&= \frac{1}{2^q} \sum_{j=0}^{q} \binom{q}{j} n^{q-j} (-1)^j \sum_{k_1 + \cdots + k_n = j} \binom{j}{k_1, \cdots, k_n} \sum_{z,m} (-1)^{(F(m) \oplus F(m \oplus z)) \cdot (\oplus_{i=1}^{n} k_i e_i)} (-1)^{a \cdot z} \\
&= \frac{1}{2^q} \sum_{j=0}^{q} \binom{q}{j} n^{q-j} (-1)^j \sum_{k_1 + \cdots + k_n = j} \binom{j}{k_1, \cdots, k_n} \left( \widehat{((\oplus_{i=1}^{n} k_i e_i) \cdot F)_\chi}(a) \right)^2 . \qquad (13)
\end{aligned}
$$

### A.3 Complete Demonstration of Theorem 1

As requested by theorem 1, we introduce $P$ and $Q$, two positive integers, and $F$, a bijection of $\mathbb{F}_2^n$. With a reasoning close to that of Eqn. (3) for the case $p = q = 1$, we get:

$$
\begin{aligned}
& \forall p \in [\![0, P]\!], \forall q \in [\![0, Q]\!], \text{the function } f_{\text{opt}}, \text{ defined in Eqn. (2), is constant} \\
\iff & \forall p \in [\![0, P]\!], \forall q \in [\![0, Q]\!], \forall a \in \mathbb{F}_2^{n*}, \widehat{\mathsf{HW}^p}(a) = 0 \text{ or } \mathbb{E}[\widehat{\mathsf{HW}^q \circ D_{(\cdot)}} F(M)](a) = 0 \\
\iff & \forall p \in [\![0, P]\!], \forall q \in [\![0, Q]\!], \forall a \in \mathbb{F}_2^{n*}, \begin{cases} \text{either } \mathsf{HW}(a) > p \text{ (See theorem 2)} \\ \text{or Eqn. (13) of Sec. A.2 is zero} \end{cases} \\
\iff & \forall p \in [\![0, P]\!], \forall q \in [\![0, Q]\!], \forall a \in \mathbb{F}_2^{n*}, \mathsf{HW}(a) \leq p \Longrightarrow \text{Eqn. (13) is zero} \\
\iff & \begin{cases} \forall p \in [\![0, P]\!], \\ \forall q \in [\![0, Q]\!], \\ \forall a \in \mathbb{F}_2^{n*}, \\ \quad \mathsf{HW}(a) \leq p \end{cases} \Longrightarrow \begin{cases} \underline{q = 1} : \forall b, \mathsf{HW}(b) \leq 1 \Longrightarrow \widehat{(b \cdot F)_\chi}(a) = 0, \\ \underline{q = 2} : \forall b, \mathsf{HW}(b) \leq 2 \Longrightarrow \widehat{(b \cdot F)_\chi}(a) = 0, \\ \quad \vdots \\ \underline{q = Q} : \forall b, \mathsf{HW}(b) \leq Q \Longrightarrow \widehat{(b \cdot F)_\chi}(a) = 0. \end{cases} \qquad (14)
\end{aligned}
$$

We provide with an explanation for the last part of Eqn. (14). The terms of Eqn. (13) corresponding to a given $j$ is a sum of 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 Sec. 4.1, 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[2]:

$$\begin{array}{c}\left(\begin{array}{cccccccc@{\quad}cccccccc}
1&1&1&0&1&0&1&1&1&0&0&0&0&0&0&0\\
0&1&1&1&0&1&0&1&1&1&0&0&0&0&0&0\\
0&0&1&1&1&0&1&0&1&1&1&0&0&0&0&0\\
0&0&0&1&1&1&0&1&0&1&1&1&0&0&0&0\\
0&0&0&0&1&1&1&0&1&0&1&1&1&0&0&0\\
0&0&0&0&0&1&1&1&0&1&0&1&1&1&0&0\\
0&0&0&0&0&0&1&1&1&0&1&0&1&1&1&0\\
0&0&0&0&0&0&0&1&1&1&0&1&0&1&1&1
\end{array}\right)\end{array}
\begin{array}{l}L_1\\L_2\\L_3\\L_4\\L_5\\L_6\\L_7\\L_8\end{array}.$$

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:

$$\left(\begin{array}{cccccccc@{\quad}cccccccc}
1&1&1&0&1&0&1&1&1&0&0&0&0&0&0&0\\
0&1&1&1&0&1&0&1&1&1&0&0&0&0&0&0\\
0&0&1&1&1&0&1&0&1&1&1&0&0&0&0&0\\
0&0&0&1&1&1&0&1&0&1&1&1&0&0&0&0\\
0&0&0&0&1&1&1&0&1&0&1&1&1&0&0&0\\
0&0&0&0&0&1&1&1&0&1&0&1&1&1&0&0\\
0&0&0&0&0&0&1&1&1&0&1&0&1&1&1&0\\
0&0&0&0&0&0&0&1&1&1&0&1&0&1&1&1
\end{array}\right)
\begin{array}{c}L'_1 \leftarrow L_1 \oplus L_2 \oplus L_4 \oplus L_7\\
L'_2 \leftarrow L_2 \oplus L_3 \oplus L_5 \oplus L_8\\
L'_3 \leftarrow L_3 \oplus L_4 \oplus L_6\\
L'_4 \leftarrow L_4 \oplus L_5 \oplus L_7\\
L'_5 \leftarrow L_5 \oplus L_6 \oplus L_8\\
L'_6 \leftarrow L_6 \oplus L_7\\
L'_7 \leftarrow L_7 \oplus L_8\\
L'_8 \leftarrow L_8\end{array},$$

which yields:

$$\left(\begin{array}{cccccccc@{\quad}cccccccc}
1&0&0&0&0&0&0&0&1&0&0&1&1&1&1&0\\
0&1&0&0&0&0&0&0&0&1&0&0&1&1&1&1\\
0&0&1&0&0&0&0&0&1&1&0&0&1&1&0&0\\
0&0&0&1&0&0&0&0&0&1&1&0&0&1&1&0\\
0&0&0&0&1&0&0&0&0&0&1&1&0&0&1&1\\
0&0&0&0&0&1&0&0&1&1&1&1&0&0&1&0\\
0&0&0&0&0&0&1&0&0&1&1&1&1&0&0&1\\
0&0&0&0&0&0&0&1&1&1&0&1&0&1&1&1
\end{array}\right)
\begin{array}{l}L'_1 = \texttt{0x80} \parallel \texttt{0x9e}\\
L'_2 = \texttt{0x40} \parallel \texttt{0x4f}\\
L'_3 = \texttt{0x20} \parallel \texttt{0xcc}\\
L'_4 = \texttt{0x10} \parallel \texttt{0x66}\\
L'_5 = \texttt{0x08} \parallel \texttt{0x33}\\
L'_6 = \texttt{0x04} \parallel \texttt{0xf2}\\
L'_7 = \texttt{0x02} \parallel \texttt{0x79}\\
L'_8 = \texttt{0x01} \parallel \texttt{0xd7}\end{array},$$

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

## C Computation of the Optimal Function $z \mapsto f_{\text{opt}}(z)$ for Some Bijections $F$

Some $f_{\text{opt}}(z)$ have been computed in Tab. 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 $\rho_{\text{opt}}$ that an attacker can expect (See definition in [17, Eqn. (15)]). It can be seen that the first nonzero $\rho_{\text{opt}}$ approximately decreases with the CM strength: it is about 25% for $F1$, about 4% for $F2$ and $F3$, and about 2% for $F4$ and $F5$.

[2] This code is a subcode of the BCH $[17, 9, 5]$ code. For more details, please refer to: http://www.math.colostate.edu/~betten/research/codes/BOUNDS/sub_16_8_5-7_2.code.

**Table 4.** Computation of $f_{\mathrm{opt}}(z)$ for centered traces raised at several powers $d$, and optimal correlation coefficient $\rho_{\mathrm{opt}}$.

| | $f_{\mathrm{opt}}(z)$ | | | | | | | | | $\rho_{\mathrm{opt}}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $z$ | `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 Sec. 4.2)** | | | | | | | | | | |
| $d = 1$ | 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 |

## D Information Leakage in the Imperfect Model

The information leakage plots are plotted in Tab. 5, 6 & 7 for the randomized HD model and in Tab. 8, 9 & 10 for the randomized "Null" model.

**Table 5.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect HD leakage model.
*Nota bene: The smaller the mutual information, the better the countermeasure.*

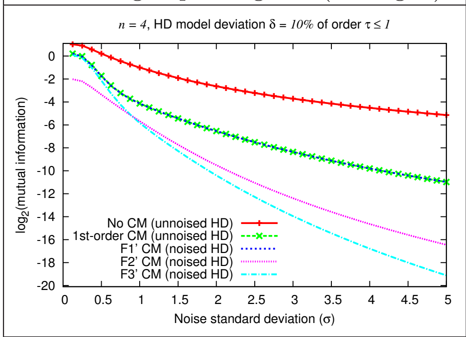
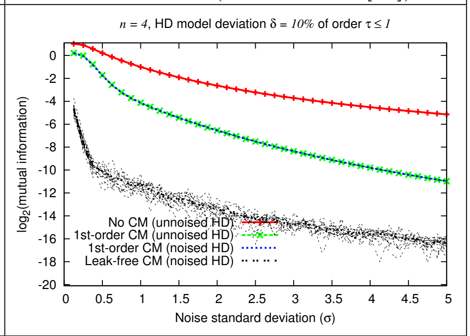
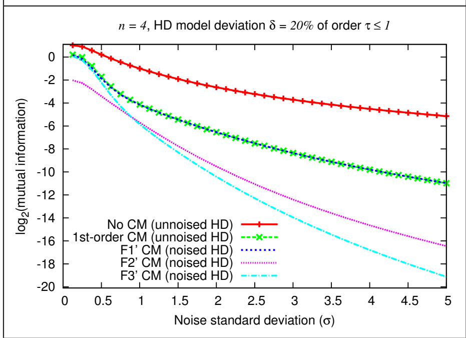
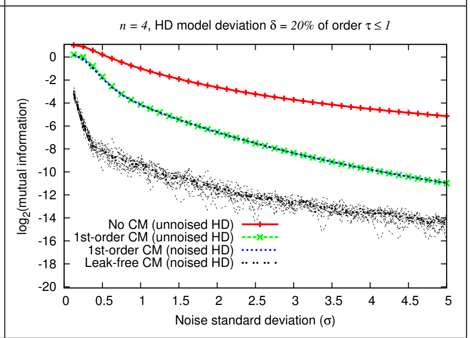
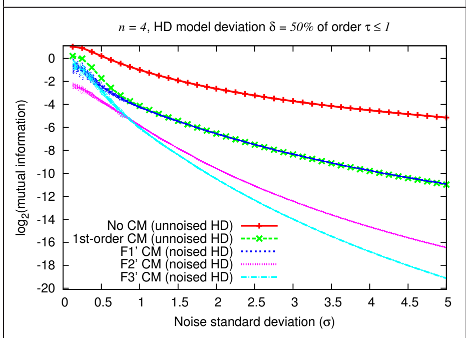
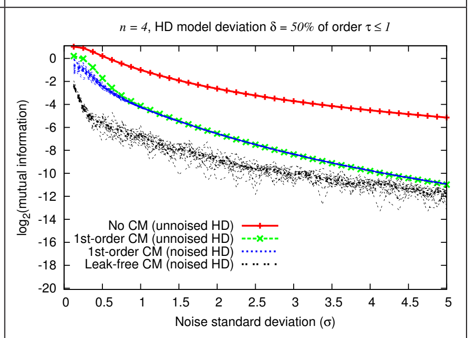
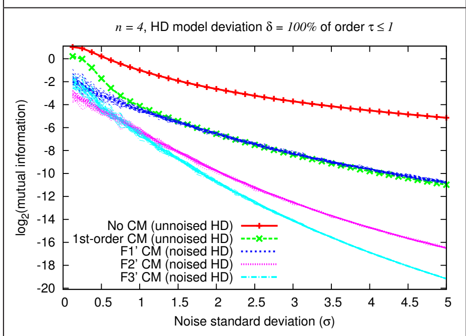
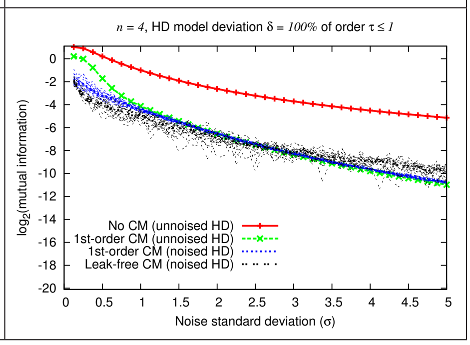
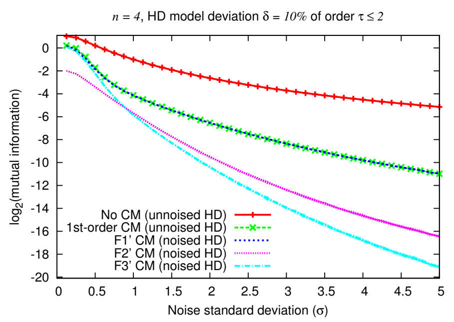
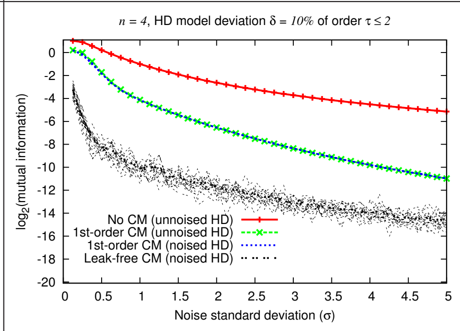
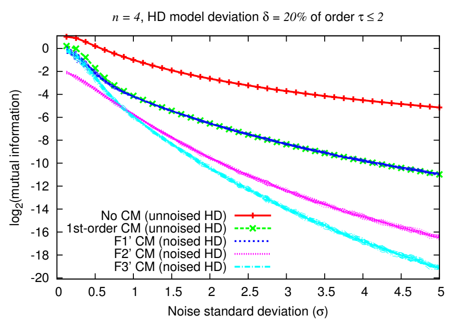
| Our leakage squeezing CM (see Fig. 2) | Leak-free CM (see reference [13]) |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |

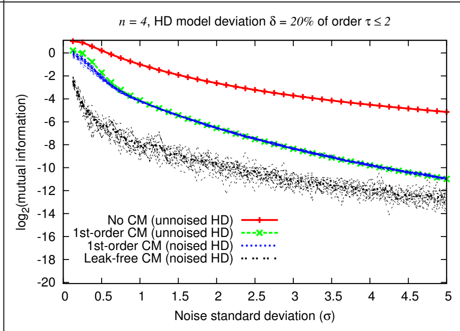
**Table 6.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect HD leakage model.
*Nota bene: The smaller the mutual information, the better the countermeasure.*

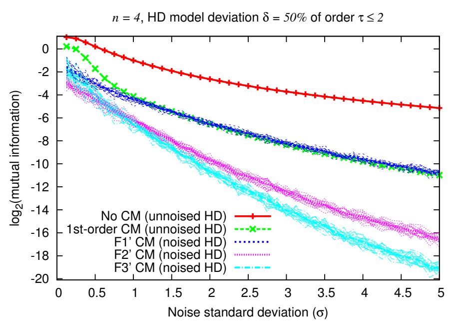
| Our leakage squeezing CM (see Fig. 2) | Leak-free CM (see reference [13]) |
|---|---|

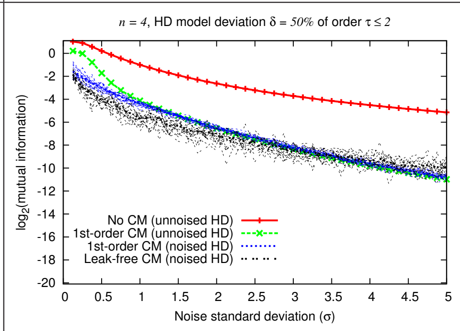


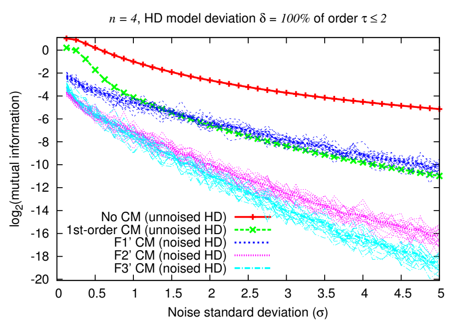



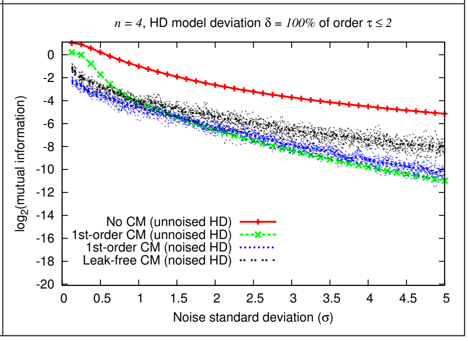



**Table 7.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect HD leakage model.
*Nota bene: The smaller the mutual information, the better the countermeasure.*

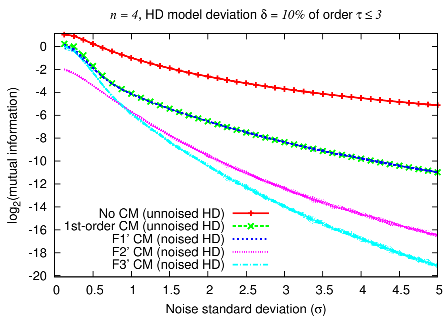
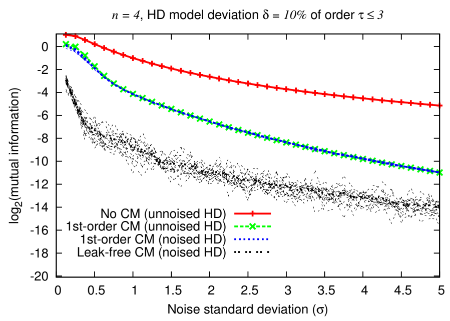
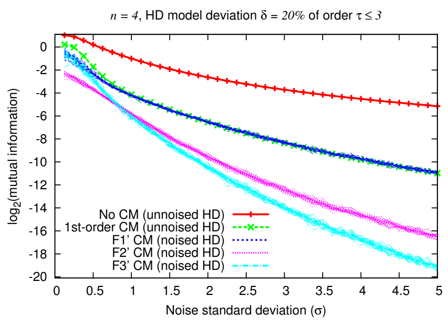
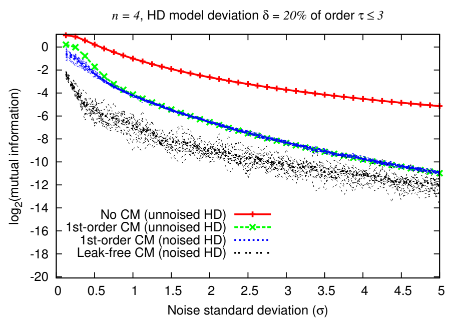
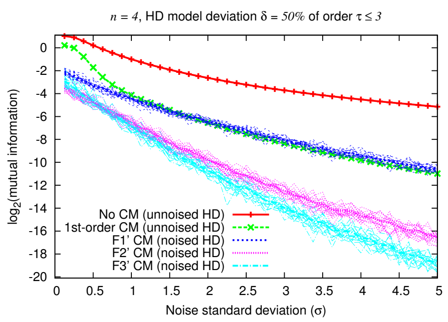
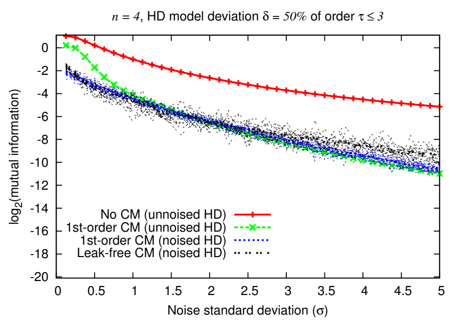
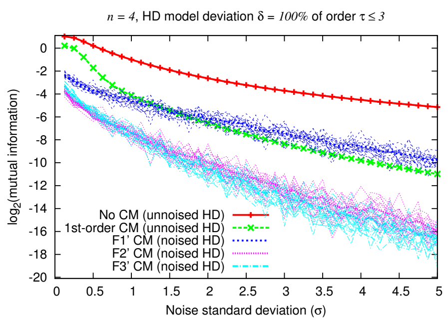
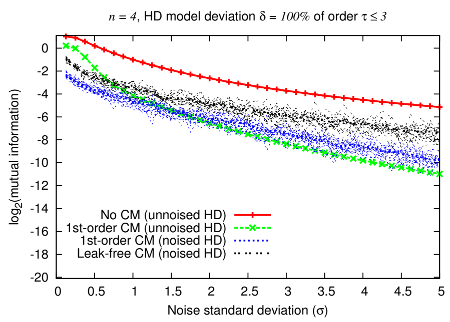
<table>
<tr><th>Our leakage squeezing CM (see Fig. 2)</th><th>Leak-free CM (see reference [13])</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

**Table 8.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect "Null" leakage model.
<u>*Nota bene*</u>*: The smaller the mutual information, the better the countermeasure.*

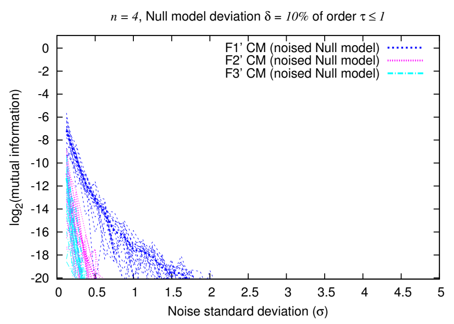
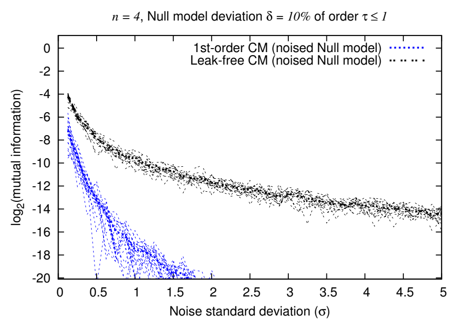
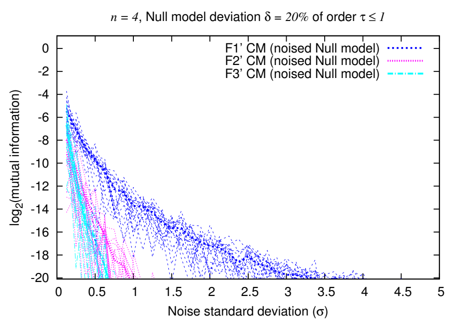
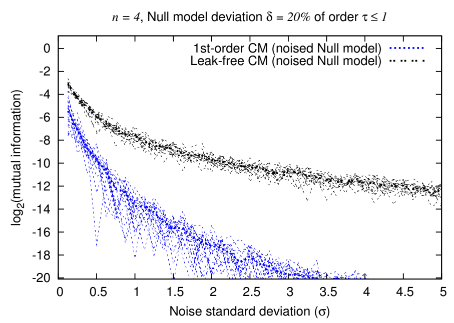
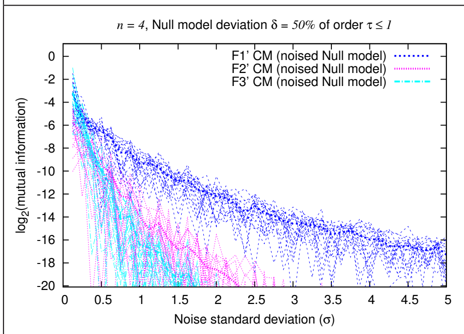
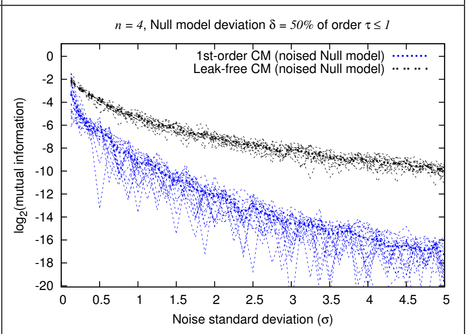
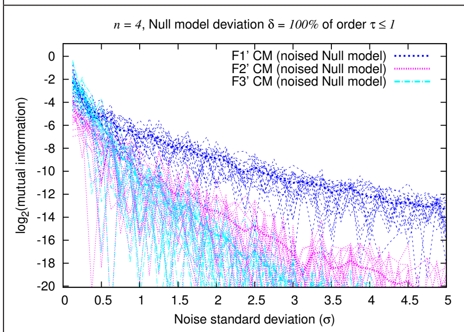
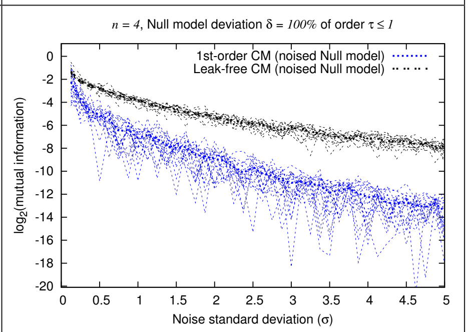
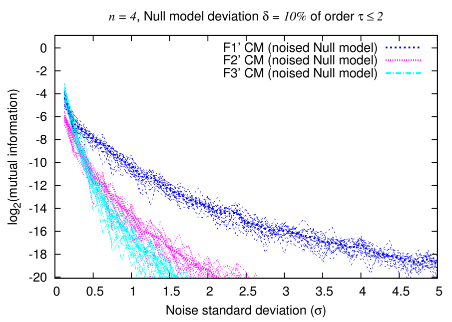
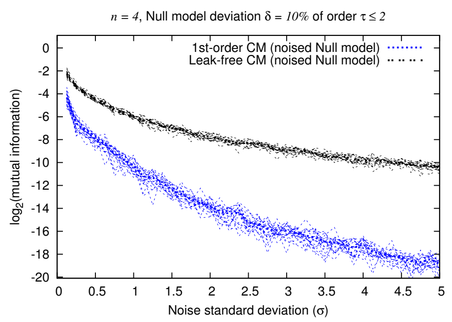
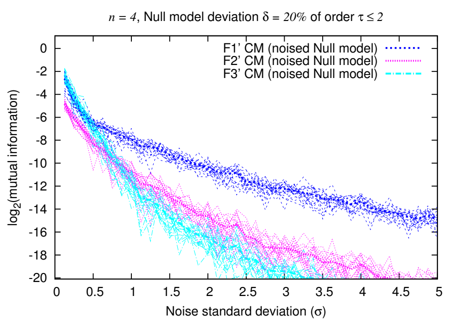
| Our leakage squeezing CM (see Fig. 2) | Leak-free CM (see reference [13]) |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

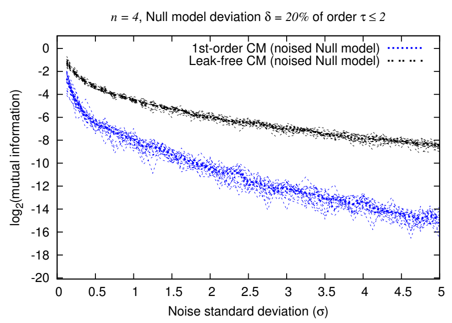
**Table 9.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect "Null" leakage model.
*Nota bene: The smaller the mutual information, the better the countermeasure.*

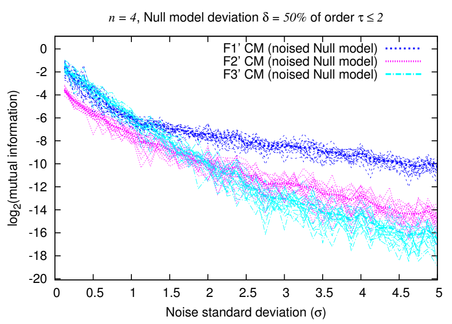
| Our leakage squeezing CM (see Fig. 2) | Leak-free CM (see reference [13]) |
|---|---|

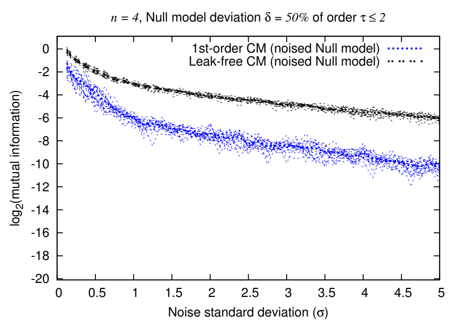


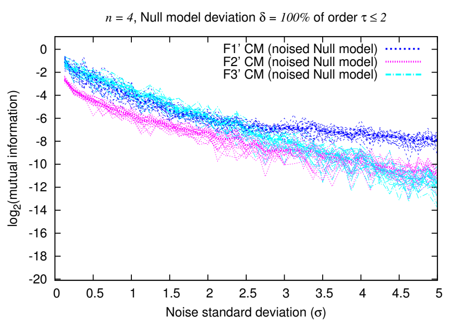



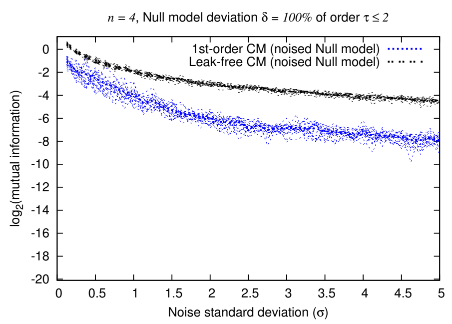
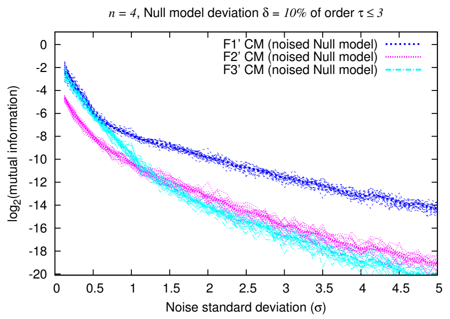
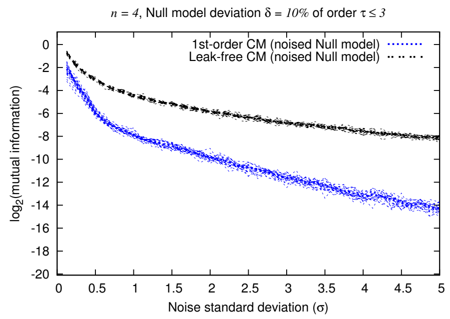
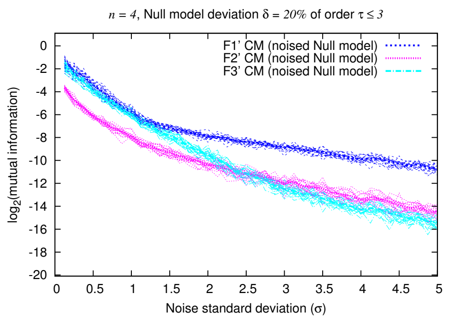
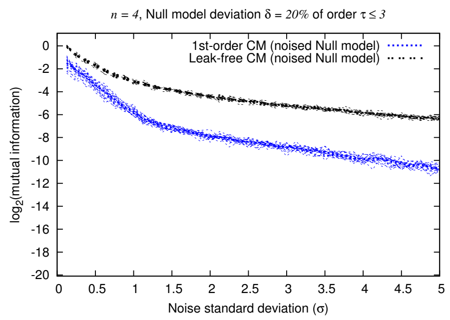
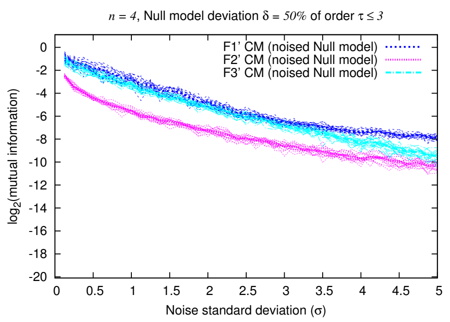
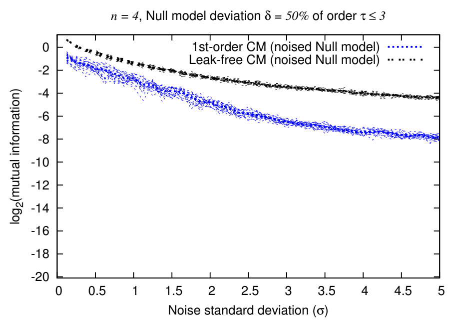
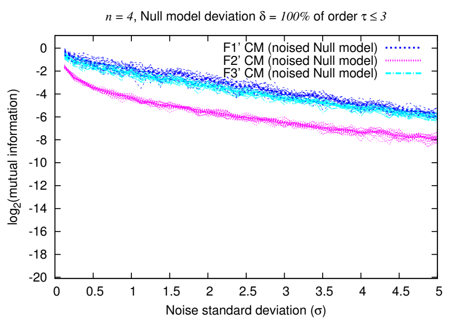
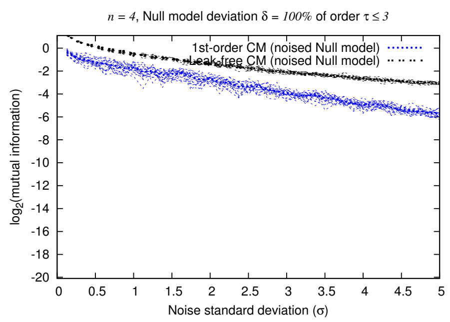



**Table 10.** Leakage comparison of the proposed CM (*left column*) and the leak-free CM [13] (*right column*) in the imperfect "NULL" leakage model.
*Nota bene: The smaller the mutual information, the better the countermeasure.*

<table>
<tr><th>Our leakage squeezing CM (see Fig. 2)</th><th>Leak-free CM (see reference [13])</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>