Statistical Analysis of Second Order Differential Power Analysis

Emmanuel Prouff, Matthieu Rivain, and Régis Bévan

Abstract—Second order Differential Power Analysis (2O-DPA) is a powerful side-channel attack that allows an attacker to bypass the widely used masking countermeasure. To thwart 2O-DPA, higher order masking may be employed but it implies a nonnegligible overhead. In this context, there is a need to know how efficient a 2O-DPA can be, in order to evaluate the resistance of an implementation that uses first order masking and, possibly, some hardware countermeasures. Different methods of mounting a practical 2O-DPA attack have been proposed in the literature. However, it is not yet clear which of these methods is the most efficient. In this paper, we give a formal description of the higher order DPA that are mounted against software implementations. We then introduce a framework in which the attack efficiencies may be compared. The attacks we focus on involve the combining of several leakage signals and the computation of correlation coefficients to discriminate the wrong key hypotheses. In the second part of this paper, we pay particular attention to 2O-DPA that involves the product combining or the absolute difference combining. We study them under the assumption that the device leaks the Hamming weight of the processed data together with an independent gaussian noise. After showing a way to improve the product combining, we argue that in this model, the product combining is more efficient not only than absolute difference combining, but also than all the other combining techniques proposed in the literature.

Index Terms—Embedded systems security, cryptographic implementations, side-channel analysis, higher order differential power analysis.

1 Introduction

Side-channel analysis (SCA) exploits information that leaks from physical implementations of cryptographic algorithms. This leakage (e.g., the power consumption or the electromagnetic emanations) may indeed reveal information on the secret data manipulated by the implementation. Among the SCA attacks, two classes may be distinguished. The set of so-called *Profiling SCA* corresponds to a powerful adversary who has a copy of the attacked device under control and who uses it to evaluate the distribution of the leakage according to the processed values. Once such an evaluation is obtained, a maximum likelihood approach is followed to recover the secret data manipulated by the attacked device. The second set of attacks is the set of so-called *Differential Power Analysis* (DPA) [1]. It corresponds to a more realistic (and much weaker) adversary than the one considered in Profiling SCA, as the focused adversary is only able to observe the device behavior and has no a priori knowledge of the implementation details. This paper only deals with the set of DPA as it includes a great majority of the attacks encountered, e.g., by the Smart Card Industry. For further information about Profiling SCA, the different studies conducted, for instance, in [2], [3], [4] may be read.

A DPA is a statistical attack that correlates a physical leakage with a prediction on the values taken by one or several intermediate variable(s) of the implementation that depend on both the plaintext and the secret key (such variables are called here *sensitive variables*). To avoid information leakage, the manipulation of sensitive variables must be protected by adding countermeasures to the algorithm.

A very common countermeasure to protect block ciphers implementations is to randomize their sensitive variables by masking techniques [5], [6]. All of these are essentially based on the same principle which can be stated as follows: every sensitive variable Z is randomly split into d shares $M_1, \ldots, M_d$ in such a way that the relation $M_1 \star \ldots \star M_d = Z$ is satisfied for a group operation $\star$ (e.g., the X-OR or the modular addition). Usually, the $d-1$ shares $M_1, \ldots, M_{d-1}$ (called *the masks*) are randomly picked up and the last one M_d (called *the masked variable*) is processed such that it satisfies $M_1 \star \ldots \star M_d = Z$. This technique is usually called a $(d-1)$*th-order masking*. When it is applied to protect the software implementation of an algorithm, the elements $M_1, \ldots, M_d$ are manipulated at different times $t_1, \ldots, t_d$ and an attacker needs to get information on all of them if he wants to get information on Z. The class of *Higher Order DPA* (HO-DPA) attacks have been introduced to defeat this kind of countermeasures.

When a $(d-1)$th-order masking is used, a dth-order DPA can be performed by combining the leakage signals $L(t_1), \ldots, L(t_d)$ resulting from the manipulation of the d shares $M_1, \ldots, M_d$. This enables the construction of a signal that is correlated to the targeted sensitive variable Z. Such an attack can theoretically bypass any $(d-1)$th-order masking. However, the noise effects imply that the difficulty of carrying out an HO-DPA in practice increases exponentially

- *E. Prouff is with the Society Oberthur Technologies, Oberthur Technologies, 71-73 rue des Hautes Pâtures, 92726 Nanterre Cedex, France. E-mail: e.prouff@oberthurcs.com.*
- *M. Rivain is with the University of Luxembourg, Faculty of Sciences, Technology and Communication 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, and with the Society Oberthur Technologies, Oberthur Technologies, 71-73 rue des Hautes Pâtures, 92726 Nanterre Cedex, France. E-mail: m.rivain@oberthurcs.com.*
- *R. Bévan is with NAGRA France. E-mail: regis.bevan@nagra.com.*

Manuscript received 4 Jun. 2008; revised 9 Oct. 2008; accepted 4 Dec. 2008; published online 8 Jan. 2009.
Recommended for acceptance by S. Olariu.
For information on obtaining reprints of this article, please send e-mail to: tc@computer.org, and reference IEEECS Log Number TC-2008-06-0273
Digital Object Identifier no. 10.1109/TC.2009.15.

with its order [5], [7]. On the other hand, the design of a higher order masking scheme that is efficient and secure against dth-order DPA for $d \geq 2$ is still an issue [8]. Therefore, first order masking (together with hardware countermeasures) is widely used to protect block ciphers implementations against DPA [6], [9], [10].

In this context, second order DPA have been widely investigated [5], [7], [11], [12], [13], [14], [15], [16]. Mainly, two combining functions have been proposed to mount sound second order DPA attacks against masked implementations. The first one, proposed by Chari et al. in [5], simply consists of processing the product of the two leakages $L(t_1)$ and $L(t_2)$ (in the sequel we call it the *product combining*). The second one, proposed by Messerges in [11], consists of computing the absolute value of the difference between $L(t_1)$ and $L(t_2)$ (we call it the *absolute difference combining*). Recently, a formal analysis of these combining functions has been initiated. In [13], Joye et al. analyzed the *single-bit* second order DPA (that is, the DPA targeting a single bit of the sensitive data) based on the absolute difference combining and they proposed a way to improve it. In [7], Schramm and Paar analyzed the *multibit* second order DPA based on the product combining. Although these separate analyses allow to better understand the drawbacks and the assets of each of these combining functions, they do not allow to clearly establish which approach is the most suitable. In [16], Oswald et al. compared the two combining functions by evaluating some correlation coefficients in a noise-free model. Based on their results, they argued that the absolute difference combining is more efficient than the product combining. However, the limitation of the leakage model used in [16] does not allow to draw definitive conclusions.

In this paper, we conduct an in-depth analysis of an HO-DPA attack that involves a combining function and target software implementations of cryptographic algorithms. We define a theoretical framework in which the efficiency of such an HO-DPA can be measured and optimized once the combining function has been chosen. Then, we analyze both the product combining and the absolute difference combining according to a realistic leakage model (namely the Hamming Weight model with noise) and we show how the efficiency of the product combining can be improved by preprocessing the leakage measurements. Our analysis states that this improved product combining leads to the best efficiency. We also argue that this function is the best published function to perform a second order DPA when devices leak the Hamming weight of the processed data and when the noise is nonnegligible.

2 Preliminaries

2.1 Notations and Useful Definitions

We use the calligraphic letters, like $\mathcal{X}$, to denote finite sets (e.g., $\mathbb{F}_2^n$). The corresponding large letter X is used to denote a random variable over $\mathcal{X}$, while the lowercase letter x a particular element from $\mathcal{X}$. The probability of the event $(X = x)$ is denoted by $\mathrm{P}[X = x]$ or $p_X(x)$. The uniform probability distribution over a set $\mathcal{X}$ is denoted by $\mathcal{U}(\mathcal{X})$ and the gaussian probability distribution with expectation μ and standard deviation σ is denoted as $\mathcal{N}(\mu, \sigma)$. The expectation of X is denoted by $\mathrm{E}[X]$, its variance by $\mathrm{Var}[X]$, and its standard deviation by $\sigma[X]$. The correlation coefficient between X and Y is denoted by $\rho[X, Y]$. It measures the linear interdependence between X and Y and is defined by

$$\rho[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sigma[X]\sigma[Y]}, \tag{1}$$

where $\mathrm{Cov}[X, Y]$, called *covariance of X and Y*, equals $\mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$ or $\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$ equivalently.

The empirical version of the correlation coefficient is the *Pearson coefficient*:

$$\widehat{\rho}(< x_1, \ldots, x_N >, < y_1, \ldots, y_N >) = \frac{\sum_{j=1}^{N}(x_j - \overline{x})(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{N}(x_j - \overline{x})^2}\sqrt{\sum_{j=1}^{N}(y_j - \overline{y})^2}}, \tag{2}$$

where $< x_1, \ldots, x_N >$ (resp. $< y_1, \ldots, y_N >$) denotes a sample of N values taken by X (resp. Y) over $\mathcal{X}$ (resp. $\mathcal{Y}$) and where $\overline{x}$ (resp. $\overline{y}$) denotes the mean $\frac{1}{N}\sum_{j=1}^{N} x_j$ (resp. $\frac{1}{N}\sum_{j=1}^{N} y_j$).

We recall hereafter a well-known property of the (Pearson) correlation coefficient.

Property 1. *The correlation coefficient (resp. the Pearson correlation coefficient) stays unchanged when an increasing affine transformation is applied to one of its input random variables (resp. input samples).*

In this paper, we often use the notion of Hamming weight. For every vector $x \in \mathbb{F}_2^n$, we denote by $\mathrm{H}(x)$ the Hamming weight of x. It equals $\sum_{i=1}^{n} x[i]$, where $x[i]$ denotes the ith bit-coordinate of x. The Hamming weight function has the following property, which will be often used in Section 4.

Property 2. *For every $z, m \in \mathbb{F}_2^n$, the Hamming weight of $z \oplus m$ satisfies*

$$\mathrm{H}(z \oplus m) = \mathrm{H}(z) + \mathrm{H}(m) - 2\mathrm{H}(z \wedge m), \tag{3}$$

where $\oplus$ denotes the bitwise addition and $\wedge$ denotes the bitwise multiplication.

2.2 Context of DPA Attacks

DPA attacks exploit the leakage that results from the manipulation of some sensitive variables. In the following definition, we formalize the notion of sensitive variable.

Definition 1 (Sensitive variable). *A variable Z is sensitive if it depends on both a public variable X (derived from the plaintext) and a secret variable K (derived from the secret key).*

In the rest of the paper, Z, X, and K are modeled as uniformly distributed random variables satisfying

$$Z = g(X, K), \tag{4}$$

where g corresponds to an intermediate calculus (e.g., an SBox function or a simple logic operation such as the bitwise addition) during the processing of the algorithm.[1] Moreover, we shall only consider variables K and Z defined over small sets (e.g., isomorphic to $\mathbb{F}_2^n$ with $n \leq 8$). Indeed,

1. The fact that Z is uniformly distributed holds if and only if g is balanced, which is very usual for a block cipher intermediate calculus.

(HO)-DPA requires to carry out statistical tests for almost all the possible values of K. Hence, the complexity (e.g., in terms of leakage measurements) of the attack increases exponentially with the dimension of $\mathcal{K}$ and only sensitive data of small length n can be targeted.

Since g and X are public, information leakage on Z implies information leakage on K. As a consequence, the manipulation of Z has to be protected against DPA and the most common algorithmic protection consists of using masking techniques [5], [6]. As recalled in Section 1, when $(d-1)$th-order masking is involved, every sensitive variable Z appearing in the algorithm is represented by d shares $M_1, \ldots, M_d$ such that

$$M_1 \star \cdots \star M_d = Z, \tag{5}$$

where $\star$ denotes a group law. The shares $M_1, \ldots, M_{d-1}$ are mutually independent random variables uniformly distributed over $\mathcal{Z}$ and the share M_d is the random variable satisfying (5).

To ensure the security, the variables M_is are manipulated at different times t_is. Thus, the leakage signal $L(t_i)$ generated by the algorithm execution, at each time t_i, can be modeled as a noisy function of M_i. More generally, we will denote by $L(t)$ the leakage generated at any time t.

As every tuple of $d-1$ shares is independent of Z, an attacker has to consider the d leakages $L(t_i)$s simultaneously in order to recover information on Z. This is the core principle of the HO-DPA attacks we formally describe in the next section.

3 Higher Order Differential Power Analysis

3.1 Adversary Model

In this paper, we assume that the attacker can query the targeted cryptographic primitive with an arbitrary number of plaintexts and obtain the corresponding physical observations. It is also assumed that the attacker cannot profile the leakage distribution according to the values of the manipulated data (Template and Profiling Attacks are thus impossible). In fact, we shall assume in the following that a correlation distinguisher is used to isolate the expected sensitive data. The attacker who is modeled in such a way is weaker than the one considered in Template Attacks. However, he corresponds quite well to the kind of adversary encountered in a large variety of applications such as the banking and GSM ones. This adversary model, which is very classical in SCA, has been considered in many other studies (e.g., [11], [13], [16], [17]).

Additionally, we assume that the attacker is able to precisely determine the manipulation time of every intermediate variable (e.g., masks, masked variables, etc.) that appears in the algorithm whose implementation is under attack. This assumption simplifies the study of the attacks. It may, however, be noted that the attacker is usually weaker than the one we consider. The manipulation times of the focused intermediate variables are indeed a priori unknown by the attacker who usually needs to consider numerous possible times within a given interval (see, for instance, [12], [16]).

3.2 Attack Description

HO-DPA aims at recovering information on $Z = g(X, K)$ (and thus on K) by simultaneously considering the leakage signals at the d times $t_1, \ldots, t_d$ that correspond to the manipulations of the d shares.

The attack starts by combining the d signals $L(t_1), \ldots, L(t_d)$ with a *combining function* $\mathcal{C}$ and by defining a *prediction function* f according to some assumptions on the device leakage model. Then, for every guess k on the value of the secret K, the attacker computes the so-called *prediction* $f \circ g(X, k)$ and checks its validity by estimating the following correlation coefficient:

$$\rho_k = \rho[\mathcal{C}(L(t_1), \ldots, L(t_d)), f \circ g(X, k)]. \tag{6}$$

Remark 2. Due to (4), the coefficient ρ_K (that corresponds to the correct guess) can be rewritten:

$$\rho_K = \rho[\mathcal{C}(L(t_1), \ldots, L(t_d)), f(Z)]. \tag{7}$$

The attack rests on the following fact: if the functions $\mathcal{C}$ and f are well chosen, then $f \circ g(X, K)$ (i.e., $f(Z)$) is highly correlated to $\mathcal{C}(L(t_1), \ldots, L(t_d))$, and thus, the coefficient ρ_K corresponding to the correct guess must be greater than every coefficient ρ_k such that $k \neq K$.

To estimate the different correlation coefficients ρ_ks, the attacker processes N leakage measurements $L_1(t), \ldots, L_N(t)$ (where the $L_j(t)$s can be modeled as N mutually independent random variables sharing the same distribution as $L(t)$). For every k, the estimation of ρ_k is obtained by computing the Pearson coefficient $\widehat{\rho}_k(N)$ between the samples $< f \circ g(X_1, k), \ldots, f \circ g(X_N, k) >$ and $< \mathcal{C}(L_1(t_1), \ldots, L_1(t_d)), \ldots, \mathcal{C}(L_N(t_1), \ldots, L_N(t_d)) >$, where X_j denotes the public variable corresponding to the jth measurement $L_j(t)$. As $\widehat{\rho}_k(N)$ tends toward ρ_k when N increases, for N large enough, the secret K must be the one that maximizes $\widehat{\rho}_k(N)$. Hence, the attacker selects the guess k that maximizes $\widehat{\rho}_k(N)$.

An HO-DPA such as described above successfully makes it possible to recover the secret K iff $\widehat{\rho}_K(N) > \widehat{\rho}_k(N)$ holds for every $k \neq K$. When the pair $(\mathcal{C}, f)$ is s.t. $\rho_K = \max_k \rho_k$, the quality of the estimations $\widehat{\rho}_k(N)$s increases with the number of measurements N and the success of the attack essentially depends on N. Then a natural definition for the efficiency of an HO-DPA involving a pair of functions $(\mathcal{C}, f)$ can be deduced.

Definition 3 (Efficiency of HO-DPA). *The efficiency of an HO-DPA given a success rate β is the smallest value N such that*

$$\mathrm{P}\left[\widehat{\rho}_K(N) > \max_{k \neq K} \widehat{\rho}_k(N)\right] \geq \beta. \tag{8}$$

The definition above allows us to evaluate the efficiency of an HO-DPA in a formal way. However, since the probability in (8) relies on the structure of the function g, it cannot be straightforwardly used to decide on the efficiency of an HO-DPA in the general case (i.e., whatever the targeted variable $Z = g(X, K)$). To render such a decision possible, one usually assumes a very low correlation between correct and incorrect guesses.[2] Under this assumption, which implies that the correlation coefficients ρ_k are almost null for every

2. This assumption which depends on the structure of g is fairly realistic if g is highly nonlinear (e.g., the AES SBox).

$k \neq K$, the efficiency of an HO-DPA mainly relies on the correlation coefficient ρ_K. This fact has been argued in [18], [19], [20], where it is shown that the number of leakage measurements N for a successful attack is around α/ρ_K^2, where α is a value that depends on the required success rate β and on the number of key guesses $|\mathcal{K}|$. In this paper, we will therefore compare attack efficiencies by means of the correlation values ρ_Ks. For a given HO-DPA attack, we will refer to ρ_K as the *correlation of the attack*: the higher the correlation of an HO-DPA, the more efficient the HO-DPA.

Remark 4. In Section 3.6, some experimental results are provided which confirm that the correlation is effectively a good efficiency indicator for HO-DPA.

At this point, a natural issue arises that is the search for pairs $(\mathcal{C}, f)$ which maximize the correlation ρ_K. As a first step, we show in the next section how to deduce the prediction function f maximizing ρ_K from a given combining function $\mathcal{C}$.

3.3 Optimal Prediction Function

Let us begin our discussion with the following important result which will be intensively used in the rest of the paper. In the following proposition as well as in the rest of the paper, we shall consider the conditional expectation $\mathrm{E}[\mathcal{C}|Z]$ as a function $\mathrm{E}[\mathcal{C}|\cdot]$ applied to Z.

Proposition 5. *Let $\mathcal{C}$ and Z be two random variables. Then, for every function f defined over $\mathcal{Z}$, we have*

$$\rho[f(Z), \mathcal{C}] = \rho[f(Z), \mathrm{E}[\mathcal{C}|Z]] \times \rho[\mathrm{E}[\mathcal{C}|Z], \mathcal{C}]. \tag{9}$$

Before proving Proposition 5, let us introduce the following useful lemma.

Lemma 6. *Let $\mathcal{C}$ and Z be two random variables. Then, for every function f defined over $\mathcal{Z}$, we have*

$$\mathrm{E}[f(Z)\mathcal{C}] = \mathrm{E}[f(Z)\mathrm{E}[\mathcal{C}|Z]]. \tag{10}$$

Proof. We assume that $\mathcal{C}$ and Z are discrete (the continuous case holds straightforwardly from the discrete one). We have

$$\mathrm{E}[f(Z)\mathcal{C}] = \sum_{z,c} \mathrm{P}[Z = z, \mathcal{C} = c] f(z) c. \tag{11}$$

Since $\mathrm{P}[Z = z, \mathcal{C} = c]$ equals $\mathrm{P}[Z = z]\mathrm{P}[\mathcal{C} = c|\mathrm{Z} = z]$, we get

$$\begin{aligned}\mathrm{E}[f(Z)\mathcal{C}] &= \sum_z \mathrm{P}[Z = z] f(z) \sum_c \mathrm{P}[\mathcal{C} = c|Z = z]c \\ &= \sum_z \mathrm{P}[Z = z] f(z) \mathrm{E}[\mathcal{C}|Z = z],\end{aligned}$$

which leads to (10). □

Remark 7. Lemma 6 implies $\mathrm{E}[\mathcal{C}] = \mathrm{E}[\mathrm{E}[\mathcal{C}|Z]]$ (for $f : z \mapsto 1$), which is known as the law of total expectation, and it implies $\mathrm{E}[\mathrm{E}[\mathcal{C}|Z]\mathcal{C}] = \mathrm{E}[\mathrm{E}[\mathcal{C}|Z]^2]$ (for $f : z \mapsto \mathrm{E}[\mathcal{C}|Z = z]$).

Based on Lemma 6, we give hereafter the proof of Proposition 5.

Proof (Proposition 5). According to Remark 7, the covariance between $f(Z)$ and $\mathcal{C}$ satisfies

$$\begin{aligned}\mathrm{Cov}[f(Z), \mathcal{C}] &= \mathrm{E}[f(Z)\mathrm{E}[\mathcal{C}|Z]] - \mathrm{E}[f(Z)]\mathrm{E}[\mathrm{E}[\mathcal{C}|Z]] \\ &= \mathrm{Cov}[f(Z), \mathrm{E}[\mathcal{C}|Z]].\end{aligned}$$

Hence, the correlation $\rho[f(Z), \mathcal{C}]$ satisfies

$$\rho[f(Z), \mathcal{C}] = \rho[f(Z), \mathrm{E}[\mathcal{C}|Z]] \times \frac{\sigma[\mathrm{E}[\mathcal{C}|Z]]}{\sigma[\mathcal{C}]}. \tag{12}$$

On the other hand, we have

$$\rho[\mathrm{E}[\mathcal{C}|Z], \mathcal{C}] = \frac{\mathrm{Cov}[\mathrm{E}[\mathcal{C}|Z], \mathcal{C}]}{\sigma[\mathrm{E}[\mathcal{C}|Z]]\sigma[\mathcal{C}]}. \tag{13}$$

Due to Lemma 6, the covariance $\mathrm{Cov}[\mathrm{E}[\mathcal{C}|Z], \mathcal{C}]$ equals $\mathrm{Cov}[\mathrm{E}[\mathcal{C}|Z], \mathrm{E}[\mathcal{C}|Z]]$, namely it equals the variance $\mathrm{Var}[\mathrm{E}[\mathcal{C}|Z]]$. Hence, (12) and (13) together imply (9). □

As a direct consequence of Proposition 5, we have the next corollary.

Corollary 8. *Let d be an integer and let $\mathcal{C}$ denote a combined leakage $\mathcal{C}(L(t_1), \ldots, L(t_d))$. The prediction function f that maximizes the correlation $\rho[f(Z), \mathcal{C}]$ is defined by*

$$f_{opt}(z) = \mathrm{E}[\mathcal{C}|Z = z]. \tag{14}$$

Let ρ_{opt} be the correlation $\rho[f_{opt}(Z), \mathcal{C}]$. If f_{opt} is not constant, then ρ_{opt} satisfies

$$\rho_{opt} = \frac{\sigma[\mathrm{E}[\mathcal{C}|Z]]}{\sigma[\mathcal{C}]}. \tag{15}$$

Proof. Let f be a function defined over $\mathcal{Z}$ and let $\rho_K{}'$ denote the correlation $\rho[f(Z), \mathcal{C}]$. Then, due to Proposition 5, we have $\rho_K = \rho[f(Z), \mathrm{E}[\mathcal{C}|Z]] \times \rho_{opt}$. As $\rho[f(Z), \mathrm{E}[\mathcal{C}|Z]]$ is always smaller than or equal to 1 and since ρ_{opt} is greater than or equal to 0, we deduce $\rho_K \leq \rho_{opt}$. This implies that the function $f = f_{opt} : z \mapsto \mathrm{E}[\mathcal{C}|Z = z]$ maximizes ρ_K. Finally, (15) holds by definition of f_{opt} and by Lemma 6. □

Corollary 8 exhibits the optimal prediction function f_{opt} and the optimal correlation of an HO-DPA according to a given combining function and the leakage distribution. Moreover, Proposition 5 gives us a mean to quantify the effectiveness loss occurring when a suboptimal function f is involved. Indeed, in this case, (9) implies that making a suboptimal prediction f decreases the optimal correlation ρ_{opt} by a factor $\rho[f, f_{opt}]$.

In practice, the kind of adversary considered in this paper is not able to compute the optimal prediction function exhibited in Corollary 8. Indeed, such a computation requires to determine the exact relationship between the leakages $L(t_i)$s and the shares M_is. In the next section, we will estimate this relationship by modeling the leakage, and then we will study the optimal prediction function and the optimal correlation for two widely used second order combining functions. We will show that some prediction functions proposed in the literature are, in fact, suboptimal, and we will compute how much they decrease the correlation ρ_{opt} (and thus the attack efficiency) from the optimal one defined in (15).

4 Analysis of the Existing Second Order DPA

The different 2O-DPA that are studied in this section are assumed to target an implementation that processes a masked sensitive variable $Z \oplus M$ at a time t_1 and the corresponding mask M at a time t_2. Variables Z and M are assumed to be mutually independent and uniformly distributed over $\mathbb{F}_2^n$.

As argued in Section 3, studying a 2O-DPA essentially amounts to studying the combining function it involves. Hereafter, we pay particular attention to the product combining [5] and the absolute difference combining [11] which are the most widely used functions in the literature. For both combining functions, we exhibit the optimal prediction f_{opt} and we calculate the optimal correlation ρ_{opt} by applying (15). We also compare f_{opt} with the Hamming weight prediction function (which is often involved in the published HO-DPA) and we study their impact on the attack efficiency. Eventually, we analyze the obtained results and address other combining functions that have been proposed in the literature.

Before presenting our analysis (and to allow us to exhibit explicit formulas), we need to make the following assumption which we claim is very usual and realistic.

Assumption 1 (Leakage Model). *The leakages $L(t_1)$ and $L(t_2)$ satisfy*

$$L(t_1) = \delta_1 + \mathrm{H}(Z \oplus M) + B_1\,, \tag{16}$$

$$L(t_2) = \delta_2 + \mathrm{H}(M) + B_2, \tag{17}$$

where δ_1 and δ_2 denote the constant parts of the leakages and $\mathrm{H}(\cdot)$ is the Hamming weight function. B_1 and B_2 are two gaussian random variables centered in zero with a standard deviation σ and Z, M, B_1, and B_2 are mutually independent.[3]

The model defined by Assumption 1 allows us to have a quite good formal representation of the device leakage. It will be referred to as the *Hamming Weight Model* in the rest of the paper.

Remark 9. In some cases, it may be sound to assume that the device does not leak the Hamming weight of the processed data but the Hamming distance between these data and an initial state (see, for instance, [17]). Extending our analysis to this so-called *Hamming distance model* is straightforward. Let $L(t_1)$ equal $\delta_1 + \mathrm{H}(IS_1 \oplus Z \oplus M) + B_1$ and $L(t_2)$ equal $\delta_2 + \mathrm{H}(IS_2 \oplus M) + B_2$, where IS_1 and IS_2 are two initial states independent of Z and M. After denoting by Z' the summation $IS_1 \oplus IS_2 \oplus Z$ and by M' the summation $M \oplus IS_2$, it can be checked that $L(t_1)$ and $L(t_2)$, respectively, equal $\delta_1 + \mathrm{H}(Z' \oplus M') + B_1$ and $\delta_2 + \mathrm{H}(M') + B_2$. As Z' and M' are uniformly distributed and mutually independent, this model is equivalent to the one defined in Assumption 1.

3. For the sake of simplicity, we assume that both noises B_1 and B_2 have the same standard deviation. The analysis can be straightforwardly generalized for $\sigma[B_1] \neq \sigma[B_2]$.

When the noises B_1 and B_2 are both null, we shall say that the model is *idealized*. The analysis of 2O-DPA in this model is of interest. First, because some devices leak quite perfect nonnoisy information. Second, because it is generic (it does not take the component noise into account) and theoretical analyses conducted in this model are usually simple. In such an idealized model, exhibiting pertinent properties and/or characteristics for new combining and prediction functions $(\mathcal{C}, f)$ is often much more simple than in a model with noise. However, this primary study is not sufficient alone and, once defined in the idealized model, a pair of functions $(\mathcal{C}, f)$ must also be analyzed in the noisy model. Indeed, the combining of leakage points always results in an amplification of the noise (e.g., the noises B_1 and B_2 are added or multiplied), and it is therefore important to study the relationship between the efficiency of a combining function and the noise variations. For this reason, in the following, we conduct our analysis in the context of both the idealized and the nonidealized model.

4.1 Product Combining Second Order DPA

In this section, we investigate the product combining function:

$$\mathcal{C}_{prod}(L(t_1), L(t_2)) = L(t_1) \times L(t_2). \tag{18}$$

This function has already been studied by Schramm and Paar in [7]. Our main contribution compared to their work is that we consider a leakage model where the offsets δ_i are not null. This makes our analysis more practical since the leakage often has a nonzero offset due to the contribution of the device activity apart from the variable manipulation. During our study, we show in particular that the efficiency of the product combining is related to the values of these offsets and we show how to significantly improve it by applying a preprocessing to the leakage signals before combining them.

Let us start our analysis by computing the optimal prediction function corresponding to $\mathcal{C}_{prod}$. According to Corollary 8, it is the function $f_{opt} = z \mapsto \mathrm{E}[L(t_1) \times L(t_2) | Z = z]$. In the next proposition, we give an explicit formula for it.

Proposition 10. *Let $L(t_1)$ and $L(t_2)$ satisfy (16) and (17). Then, for every $z \in \mathbb{F}_2^n$, we have*

$$\begin{aligned}\mathrm{E}[L(t_1) \times L(t_2) | Z = z] = -\frac{1}{2}\mathrm{H}(z) + \frac{n^2+n}{4} \\ + \frac{n}{2}(\delta_1 + \delta_2) + \delta_1\delta_2.\end{aligned} \tag{19}$$

Proof. Since B_1 and B_2 are independent from M and satisfy $\mathrm{E}[B_1] = \mathrm{E}[B_2] = 0$, the expectation $\mathrm{E}[L(t_1) \times L(t_2) | Z = z]$ is equal to $\mathrm{E}[\mathrm{H}(z \oplus M)\mathrm{H}(M)] + \delta_1 \mathrm{E}[\mathrm{H}(M)] + \delta_2 \mathrm{E}[\mathrm{H}(z \oplus M)] + \delta_1\delta_2$. Moreover, since M is uniformly distributed over $\mathbb{F}_2^n$, we have $\mathrm{E}[\mathrm{H}(z \oplus M)] = \mathrm{E}[\mathrm{H}(M)] = \frac{n}{2}$ and, from Lemma 21 (see Appendix 1), we have $\mathrm{E}[\mathrm{H}(z \oplus M)\mathrm{H}(M)] = -\frac{1}{2}\mathrm{H}(z) + \frac{n^2+n}{4}$. Hence, we get (19). □

Proposition 10 together with Corollary 8 implies that the function $z \mapsto \mathrm{H}(z)$, or any decreasing affine function of it, may be used as an optimal prediction function for a 2O-DPA involving the product combining.

Corollary 11. *In the Hamming weight model, the optimal prediction function f_{opt} corresponding to $\mathcal{C}_{prod}$ is of the form*

$$f_{opt} : z \mapsto A \circ \mathrm{H}(z), \tag{20}$$

where A is an affine decreasing function defined over $\mathrm{H}(\mathcal{Z})$.

Proof. This a straightforward consequence of Corollary 8 and Proposition 10. □

It must be noted that the Hamming weight function has already been used as prediction function in previous works [7], [16]. Corollary 11 shows that this choice maximizes the amplitude of the correlation coefficient (in the Hamming weight model) and that it results in a negative correlation (as observed in [16], for instance).

To compute the optimal correlation corresponding to one of the functions satisfying (20), we exhibit in the following a formula for the variance of $L(t_1) \times L(t_2)$.

Proposition 12. *Let $L(t_1)$ and $L(t_2)$ satisfy (16) and (17). Then, the variance of $L(t_1) \times L(t_2)$ satisfies*

$$\mathrm{Var}[L(t_1) \times L(t_2)] = \frac{2n^3 + n^2}{16} + \frac{n}{4}\left(n\delta_1 + \delta_1^2 + n\delta_2 + \delta_2^2\right) + \frac{n^2+n}{2}\sigma^2 + \left(n\delta_1 + \delta_1^2 + n\delta_2 + \delta_2^2\right)\sigma^2 + \sigma^4. \tag{21}$$

Proof. As Z and M are mutually independent and uniformly distributed, one can check that M and $Z \oplus M$ are mutually independent. This implies that $L(t_1)$ and $L(t_2)$ are also mutually independent and we get

$$\mathrm{Var}[L(t_1) \times L(t_2)] = \mathrm{E}[L(t_1)^2]\mathrm{E}[L(t_2)^2] - \mathrm{E}[L(t_1)]^2\mathrm{E}[L(t_2)]^2. \tag{22}$$

Since Z and M are uniformly distributed over $\mathbb{F}_2^n$ and mutually independent, Lemma 20 (see Appendix 1) implies $\mathrm{E}[\mathrm{H}(M)^2] = \mathrm{E}[\mathrm{H}(Z \oplus M)^2] = \frac{n^2+n}{4}$. Then, since we have $B_i \sim \mathcal{N}(0, \sigma)$, one deduces that $\mathrm{E}[L(t_i)]$ and $\mathrm{E}[L(t_i)^2]$, respectively, equal $\frac{n}{2} + \delta_i$ and $\frac{n^2+n}{4} + n\delta_i + \delta_i^2 + \sigma^2$ for $i = 1, 2$. Finally, simplifying (22) leads to (21). □

It can be noted in (21) that $\mathrm{Var}[L(t_1) \times L(t_2)]$ is an increasing function of $n\delta_1 + \delta_1^2 + n\delta_2 + \delta_2^2$. Hence, the offsets values that minimize the variance are $\delta_1 = \delta_2 = -n/2$. Actually, this is not surprising: with such offsets, the leakages are centered in zero (i.e., $\mathrm{E}[L(t_1)] = \mathrm{E}[L(t_2)] = 0$) which alleviates the noise amplification caused by the product combining. As a direct consequence, minimizing the variance of $L(t_1) \times L(t_2)$ (and thus maximizing the correlation) can be done by centering the leakage signals $L(t_1)$ and $L(t_2)$ in zero (namely by substituting $L(t) - \mathrm{E}[L(t)]$ for $L(t)$). This can be simply achieved by averaging the leakage for a large number of measurements, then subtracting the average to each measurements. In the sequel, this preprocessing is called *normalization step*.

In the Hamming weight model, if the data V_t manipulated at time t is uniformly distributed over $\mathbb{F}_2^n$, then the leakage after the preprocessing step equals $L(t) - \mathrm{E}[L(t)]$ and satisfies

$$L(t) - \mathrm{E}[L(t)] = -\frac{n}{2} + \mathrm{H}(V_t) + B_t.$$

After assuming that the preprocessing step is part of the combining computation, we get the improved product combining function:

$$\mathcal{C}_{prod^\star}(L(t_1), L(t_2)) = (L(t_1) - \mathrm{E}[L(t_1)]) \times (L(t_2) - \mathrm{E}[L(t_2)]).$$

Then, we have the following proposition.

Proposition 13. *For every $z \in \mathbb{F}_2^n$, we have*

$$\mathrm{E}\left[\mathcal{C}_{prod^\star}(L(t_1), L(t_2)) | Z = z\right] = -\frac{1}{2}\mathrm{H}(z) + \frac{n}{4}$$

and

$$\mathrm{Var}\left[\mathcal{C}_{prod^\star}(L(t_1), L(t_2))\right] = \frac{n^2}{16} + \frac{n}{2}\sigma^2 + \sigma^4.$$

Proof. Proposition 13 straightforwardly results from Propositions 10 and 12 by setting $\delta_1 = \delta_2 = -n/2$. □

As a consequence of the proposition above, in the Hamming weight model, an optimal prediction function f_{opt} corresponding to $\mathcal{C}_{prod^\star}$ is of the form

$$f_{opt} : z \mapsto A \circ \mathrm{H}(z),$$

where A is an affine decreasing function defined over $\mathrm{H}(\mathcal{Z})$.

Due to Proposition 13 and Corollary 8, we can propose an explicit formula for the optimal correlation $\rho_{opt}^{prod^\star}$ corresponding to the improved product combining $\mathcal{C}_{prod^\star}$ and f_{opt}. In the Hamming weight, the correlation satisfies

$$\rho_{opt}^{prod^\star} = \frac{\sqrt{n}}{\sqrt{n^2 + 8n\sigma^2 + 16\sigma^4}}. \tag{23}$$

In particular, in the idealized model ($\sigma = 0$), it satisfies $\rho_{opt}^{prod^\star} = 1/\sqrt{n}$, and in the *very* noisy model ($\sigma \gg n$), it satisfies $\rho_{opt}^{prod^\star} \approx \sqrt{n}/4\sigma^2$. As an illustration to (23), Table 1 gives some values of the correlation for $n \in \{0, \ldots, 8\}$ and $\sigma \in \{0, 1, 5, 10\}$.

TABLE 1
(Optimal) Correlation for the Improved Product Combining

$\sigma \backslash n$	1	2	3	4	5	6	7	8
0	1.00	0.707	0.577	0.500	0.447	0.408	0.378	0.354
1	0.200	0.236	0.247	0.250	0.248	0.245	0.241	0.236
5	0.010	0.014	0.017	0.019	0.021	0.023	0.025	0.026
10	0.002	0.004	0.004	0.005	0.006	0.006	0.007	0.007

To illustrate the gain of efficiency resulting from the normalization step we propose in this paper, let us now consider the correlation ρ_{opt}^{prod-0} for the classical product combining function (18) in the Hamming weight model without offsets (such as computed in [7]). It satisfies

$$\rho_{opt}^{prod-0} = \frac{\sqrt{n}}{\sqrt{2n^3 + n^2 + 8(n^2+n)\sigma^2 + 16\sigma^4}}.$$

It can be checked that ρ_{opt}^{prod-0} is strictly lower than the correlation $\rho_{opt}^{prod^\star}$ we obtained for the product combining with preprocessing $\mathcal{C}_{prod^\star}$. Figs. 1 and 2 show how the value of the offsets (assuming $\delta_1 = \delta_2 = \delta$) affects the correlation ρ_{opt}^{prod} for

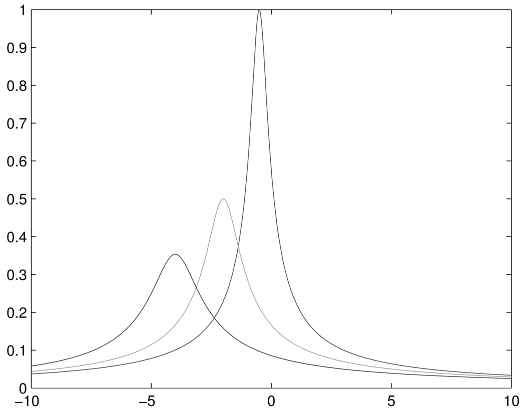

Fig. 1. Correlation ρ_{opt}^{prod} for $n = 8$ (on the left), $n = 4$ (in the middle), and $n = 1$ (on the right) in the idealized model, according to the offset δ.

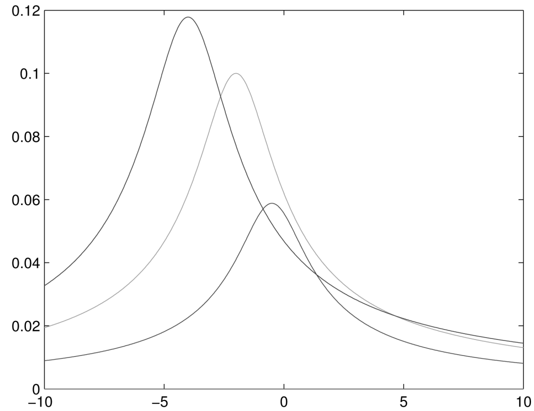

Fig. 2. Correlation ρ_{opt}^{prod} for $n = 8$ (on the left), $n = 4$ (in the middle), and $n = 1$ (on the right) in a noisy model ($\sigma = 2$), according to the offset δ.

$n \in \{1, 4, 8\}$ in the idealized model and in a noisy model ($\sigma = 2$). The maximum of this correlation is always reached for $\delta = -n/2$. Moreover, we observe that the correlation quickly decreases when the offset deviates from $-n/2$, which demonstrates the effectiveness of our improvement.

4.2 Absolute Difference Combining Second Order DPA

In this section, we investigate the absolute difference combining function, i.e., we take interest in the variable

$$\mathcal{C}_{diff}(L(t_1), L(t_2)) = |L(t_1) - L(t_2)|.$$

The absolute difference combining has already been studied by Joye et al. in [13]. In their paper, the authors consider the idealized model (i.e., without noise) and analyze a single-bit 2O-DPA (i.e., with a binary prediction function: $f(Z) \in \{0, 1\}$).

In the present paper, we extend this analysis to the multibit case (i.e., where f is not a binary function but the optimal prediction function) not only in the idealized but also in the noisy model. In the Hamming weight model, $\mathcal{C}_{diff}(L(t_1), L(t_2))$ equals $|\delta_1 - \delta_2 + \mathrm{H}(Z \oplus M) - \mathrm{H}(M) + B_1 - B_2|$. For this combining to work correctly, it is important that δ_1 be equal to δ_2. Indeed, if there is a great difference between these values, then the effect of the absolute value is reduced (or even canceled) by the constant term $\delta_1 - \delta_2$. For instance (neglecting the noise), if we have $|\delta_1 - \delta_2| > n$, then $\delta_1 - \delta_2 + \mathrm{H}(Z \oplus M) - \mathrm{H}(M)$ is either strictly positive or strictly negative and, as noticed by Messerges in [11], difference without absolute value is not a sound combining function (i.e., the difference between the two leakages is not correlated to the sensitive variable). Consequently, as for the product combining, we point out that the leakages must be normalized in order to have identical offsets in both leakage signals. Thus, as in Section 3.4, we will consider in this section that the leakages are normalized before being combined in order to ensure that they have similar offsets (i.e., we define the combining function $\mathcal{C}_{diff^*}$ such that $\mathcal{C}_{diff^*}(L(t_1), L(t_2)) = |L(t_1) - \mathrm{E}[L(t_1)] - L(t_2) + \mathrm{E}[L(t_2)]|$). In that case, the combined leakage after preprocessing satisfies

$$\mathcal{C}_{diff^*}(L(t_1), L(t_2)) = |\mathrm{H}(Z \oplus M) - \mathrm{H}(M) + B|, \qquad (24)$$

where B denotes $B_1 - B_2$ and satisfies $B \sim \mathcal{N}(0, \sqrt{2}\sigma)$.

For the absolute difference combining, it is not possible to exhibit a simple formula for the expectation that would be pertinent in the general case. Hence, we structure our study of the combining function in two steps: the first one is performed in the idealized model and the second one in the noisy model.

4.2.1 Study in the Idealized Model

If B is null, then (24) becomes

$$\mathcal{C}_{diff^*}(L(t_1), L(t_2)) = |\mathrm{H}(Z \oplus M) - \mathrm{H}(M)|.$$

In the following proposition, we exhibit an explicit formula for the expectation of $|\mathrm{H}(Z \oplus M) - \mathrm{H}(M)|$.

Proposition 14. *Let z be an element of $\mathbb{F}_2^n$. Then we have*

$$\mathrm{E}[|\mathrm{H}(M) - \mathrm{H}(z \oplus M)|] = 2^{1-\mathrm{H}(z)}\mathrm{H}(z)\binom{\mathrm{H}(z) - 1}{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor}. \qquad (25)$$

Proof. The proof of Proposition 14 is given in Appendix 2.2. □

As a consequence of Proposition 14, the optimal prediction for the absolute difference combining in the idealized Hamming weight model is not the Hamming weight of Z but a nonaffine function of it.

Corollary 15. *In the Hamming weight model, the optimal prediction function f_{opt} corresponding to $\mathcal{C}_{diff*}$ is of the form*

$$f_{opt} : z \mapsto [A \circ f](z),$$

where f is the function $z \mapsto 2^{1-\mathrm{H}(z)}\mathrm{H}(z)\binom{\mathrm{H}(z)-1}{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor}$ and A is either the identity function or an affine increasing function defined over $f(\mathcal{Z})$.

Proof. This a straightforward consequence of Corollary 8 and Proposition 14. □

Our main interest in Corollary 15 is that it tells us that even when the leakage satisfies the Hamming weight model, the Hamming weight of the targeted variable is not necessarily the optimal prediction for an HO-DPA. It actually depends on the combining function.

TABLE 2
Correlations for the Absolute Difference Combining in the Idealized Model

n	1	2	3	4	5	6	7	8
H	1.00	0.53	0.41	0.35	0.31	0.28	0.26	0.24
f_{opt}	1.00	0.65	0.50	0.41	0.35	0.31	0.28	0.26

The variance of $|\mathrm{H}(Z \oplus M) - \mathrm{H}(M)|$ has already been computed by Joye et al. [13]. The authors prove that it satisfies

$$\mathrm{Var}[|\mathrm{H}(Z \oplus M) - \mathrm{H}(M)|] = \frac{n}{2} - \left(2^{-2n} n \binom{2n}{n}\right)^2. \quad (26)$$

By Corollary 8 and in view of formulas (25) and (26), we deduce the optimal correlation related to $\mathcal{C}_{diff^\star}$:

$$\rho_{opt}^{diff^\star} = \frac{2^n \sum_{i=0}^n 2^{-2i} i^2 \binom{n}{i} \binom{i-1}{\lfloor \frac{i}{2} \rfloor}^2 - \left(\sum_{i=0}^n 2^{-i} i \binom{n}{i} \binom{i-1}{\lfloor \frac{i}{2} \rfloor}\right)^2}{2^{2n-2}\left(\frac{n}{2} - \left(2^{-2n} n \binom{2n}{n}\right)^2\right)}.$$

We have computed in Table 2 the optimal correlation $\rho_{opt}^{diff^\star}$ for some values of n. For comparison, we have also computed the correlation ρ_{HW} that corresponds to the Hamming weight prediction function (i.e., $f : z \mapsto \mathrm{H}(z)$). As expected, choosing our new prediction function makes it possible to slightly increase the correlation value (especially for low values of n). Furthermore, it can be checked that, as stated in Proposition 5, the efficiency gain is $\rho(f, f_{opt})$.

When the leakage is noisy, the previous analysis is no longer valid and cannot be extended to take the noise into account. Therefore, in the next section, we conduct a complementary analysis which addresses the noisy model.

4.2.2 Study in the Noisy Model

In the analysis that follows, we shall use the notation erf to denote the *error function* defined for every $x \in \mathbb{R}$ by $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$. We recall that the probability distribution function Φ of the standard gaussian distribution $\mathcal{N}(0,1)$ and the error function satisfy $\Phi(x) = \frac{1}{2}(1 + \mathrm{erf}(x/\sqrt{2}))$. The following proposition shall be useful to study $\mathcal{C}_{diff^\star}$ when the leakage is noisy.

Proposition 16. *Let s be a real number and B be a Gaussian random variable centered in zero with a standard deviation σ_0. The expectation of the variable $|s + B|$ satisfies*

$$\mathrm{E}[|s + B|] = s\,\mathrm{erf}\left(\frac{s}{\sqrt{2}\sigma_0}\right) + \frac{\sqrt{2}\sigma_0}{\sqrt{\pi}} \exp\left(-\frac{s^2}{2\sigma_0^2}\right). \quad (27)$$

Proof. The proof of Proposition 16 is given in Appendix 1. □

As a straightforward consequence of Proposition 16, we have the following corollary.

Corollary 17. *Let $L(t_1)$ and $L(t_2)$ satisfy (16) and (17). For every $z \in \mathbb{F}_2^n$, we have*

$$\begin{aligned} &\mathrm{E}\left[\mathcal{C}_{diff^\star}(L(t_1), L(t_2)) | Z = z\right] \\ &= \mathrm{E}\left[(\mathrm{H}(z \oplus M) - \mathrm{H}(M)) \mathrm{erf}\left(\frac{\mathrm{H}(z \oplus M) - \mathrm{H}(M)}{2\sigma}\right)\right] \\ &+ \frac{2\sigma}{\sqrt{\pi}} \mathrm{E}\left[\exp\left(-\frac{(\mathrm{H}(z \oplus M) - \mathrm{H}(M))^2}{4\sigma^2}\right)\right] \end{aligned}, \quad (28)$$

and

$$\begin{aligned} &\mathrm{Var}\left[\mathcal{C}_{diff^\star}(L(t_1), L(t_2))\right] \\ &= 2\sigma^2 + \frac{n}{2} - \mathrm{E}\left[\mathcal{C}_{diff^\star}(L(t_1), L(t_2))\right]^2 \end{aligned}. \quad (29)$$

Proof. After denoting $S = \mathrm{H}(z \oplus M) - \mathrm{H}(M)$, we get $\mathrm{E}[|L(t_1) - L(t_2)| | Z = z] = \mathrm{E}[|S + B|]$ and Proposition 16 directly leads to (28). Since we have $\mathrm{E}[B] = 0$, then $\mathrm{E}[|S + B|^2]$ equals $\mathrm{E}[S^2] + \mathrm{E}[B^2]$. Due to the linearity of the expectation, $\mathrm{E}[S^2]$ equals $\mathrm{E}[\mathrm{H}(M)^2] + \mathrm{E}[\mathrm{H}(z \oplus M)^2] - 2\mathrm{E}[\mathrm{H}(M)\mathrm{H}(z \oplus M)]$. Then, from Lemmas 20 and 21 (see Appendix 1), we deduce $\mathrm{E}[S^2] = \mathrm{H}(z)$. On the other hand, we have $\mathrm{E}[B^2] = 2\sigma^2$; hence, we deduce $\mathrm{E}[|S + B|^2] = 2\sigma^2 + \mathrm{H}(z)$, which finally gives (31) by definition of the variance. □

Corollary 17 does not allow to exhibit explicit formulas for f_{opt} and ρ_{opt} in the noisy model. However, (30) and (31) may be involved to efficiently compute the optimal prediction function and the optimal correlation corresponding to $\mathcal{C}_{diff^\star}$ in the noisy model for every pair (n, σ). As an illustration, we give in Table 3 the exact optimal correlation $\rho_{opt}^{diff^\star}$ for $n \in \{1, \ldots, 8\}$ and $\sigma \in \{0, 1, 5, 10\}$.

TABLE 3
Optimal Correlation for the Absolute Difference Combining

$\sigma \backslash n$	1	2	3	4	5	6	7	8
0	1.00	0.655	0.495	0.405	0.348	0.308	0.280	0.258
1	0.143	0.166	0.173	0.173	0.171	0.168	0.164	0.161
5	0.007	0.009	0.011	0.013	0.014	0.015	0.016	0.017
10	0.002	0.002	0.003	0.003	0.004	0.004	0.004	0.005

In order to determine the efficiency loss resulting from the use of the Hamming weight as prediction function instead of the one defined in (28), we computed the correlation $\rho(\mathrm{H}(Z), f_{opt}(Z))$ (as suggested in Proposition 5) for different values of n and σ. Table 4 lists some of our results.

Table 4 suggests that whatever the dimension n, the correlation $\rho(\mathrm{H}(Z), f_{opt}(Z))$ tends toward 1 when σ increases. This suggests that in the noisy model, the Hamming weight of Z (or an affine function of it) is a good prediction for the absolute difference combined leakage and that it becomes optimal as the noise increases. The following corollary brings an explanation to this phenomenon.

TABLE 4
Correlation between the Optimal Prediction Function and the Hamming Weight

$\sigma \backslash n$	1	2	3	4	5	6	7	8
0	1	0.816	0.832	0.861	0.886	0.905	0.919	0.930
1	1	0.996	0.996	0.996	0.996	0.996	0.996	0.996
5	1	0.998	0.997	0.999	0.999	0.999	0.999	0.999
10	1	1	1	1	0.999	0.999	0.999	0.999

Corollary 18. *Let $L(t_1)$ and $L(t_2)$ satisfy (16) and (17). Then, for every integer n and for every $z \in \mathbb{F}_2^n$, we have*

$$\mathrm{E}\big[\mathcal{C}_{diff^\star}(L(t_1), L(t_2))|Z = z\big] = \frac{2\sigma}{\sqrt{\pi}} + \frac{\mathrm{H}(z)}{2\sqrt{\pi}\sigma} + \varepsilon\left(\frac{1}{\sigma^3}\right),$$

and

$$\mathrm{Var}\big[\mathcal{C}_{diff^\star}(L(t_1), L(t_2))\big] = \frac{2\pi - 4}{\pi}\sigma^2 + \frac{\pi - 2}{2\pi}n + \varepsilon\left(\frac{1}{\sigma^2}\right).$$

Proof. Let us focus on (28) asymptotically. For every a, we have $\mathrm{erf}(a) = \frac{2}{\sqrt{\pi}}a + \varepsilon(a^3)$ and $\exp(a) = 1 + a + \varepsilon(a^2)$. Since we also have $\mathrm{H}(z \oplus M) - \mathrm{H}(M) = \varepsilon(1)$ (as n is a constant), we can rewrite (28) in the following form:

$$\begin{aligned}&\mathrm{E}\big[\mathcal{C}_{diff^\star}(L(t_1), L(t_2))|Z = z\big]\\ &\quad= \frac{1}{\sqrt{\pi}\sigma}\mathrm{E}\Big[(\mathrm{H}(z \oplus M) - \mathrm{H}(M))^2\Big] + \varepsilon\left(\frac{1}{\sigma^3}\right)\\ &\quad+ \frac{2\sigma}{\sqrt{\pi}}\left(1 - \frac{1}{4\sigma^2}\mathrm{E}\Big[(\mathrm{H}(z \oplus M) - \mathrm{H}(M))^2\Big] + \varepsilon\left(\frac{1}{\sigma^4}\right)\right)\end{aligned}. \quad (30)$$

Then, $\mathrm{E}[(\mathrm{H}(z \oplus M) - \mathrm{H}(M))^2]$ equals $\mathrm{E}[\mathrm{H}(M)^2] + \mathrm{E}[\mathrm{H}(z \oplus M)^2] - 2\mathrm{E}[\mathrm{H}(M)\mathrm{H}(z \oplus M)]$. From Lemmas 20 and 21 (see Appendix 1), one verifies that this expression equals $\mathrm{H}(z)$, which together with (30) and (29) implies Corollary 18. □

Corollary 18 confirms the empirical study presented in Table 4: in the noisy model, the Hamming weight is a good prediction for the absolute difference combined leakage. Indeed, the function $z \mapsto \mathrm{E}[|L(t_1) - L(t_2)||Z = z]$ (which corresponds to the optimal prediction function) tends toward an affine function of $\mathrm{H}(z)$ when the noise increases. Moreover, we can deduce from Corollaries 8 and 18 an approximation of the correlation $\rho_{opt}^{diff^\star}$ when n is negligible compared to σ:

$$\rho_{opt}^{diff^\star} \approx \frac{\sqrt{n}}{4\sqrt{2\pi - 4}\sigma^2}.$$

4.3 Product versus Absolute Difference

In the two previous sections, we have investigated the correlation of 2O-DPA involving either the product or the absolute difference as combining function. Tables 1 and 3 give the correlations for $n \in \{0, \ldots, 8\}$ and $\sigma \in \{0, 1, 5, 10\}$ and show that, for all these parameters, the correlation for the product combining is greater than the correlation for the absolute difference combining.

In a *very* noisy model ($\sigma \gg n$), we have shown that the correlations satisfy

$$\rho_{opt}^{prod^\star} \approx \frac{\sqrt{n}}{4\sigma^2} = 0.25\frac{\sqrt{n}}{\sigma^2}$$

and

$$\rho_{opt}^{diff^\star} \approx \frac{\sqrt{n}}{4\sqrt{2\pi - 4}\sigma^2} \approx 0.165\frac{\sqrt{n}}{\sigma^2}.$$

We observe a linear relationship between the two approximations of the correlations in the *very* noisy model: $\rho_{opt}^{prod^\star} \approx 1.5\rho_{opt}^{diff^\star}$. As a straightforward consequence of this relation, the correlation $\rho_{opt}^{prod^\star}$ is always greater than $\rho_{opt}^{diff^\star}$ when the noise is high and the two correlations are asymptotically equivalent when the noise increases.

TABLE 5
Number of Required Measurement for the Product Combining

n	1	2	3	4
$\sigma = 0,\ SR = 90.0\%$	20	30	40	60
$\sigma = 0,\ SR = 99.9\%$	30	50	80	130
$\sigma = 1,\ SR = 90.0\%$	430	310	280	280
$\sigma = 1,\ SR = 99.9\%$	940	690	600	600
$\sigma = 5,\ SR = 90.0\%$	190000	100000	65000	55000
$\sigma = 5,\ SR = 99.9\%$	410000	205000	135000	120000
n	5	6	7	8
$\sigma = 0,\ SR = 90.0\%$	80	100	110	130
$\sigma = 0,\ SR = 99.9\%$	150	190	230	280
$\sigma = 1,\ SR = 90.0\%$	280	290	300	310
$\sigma = 1,\ SR = 99.9\%$	570	600	650	700
$\sigma = 5,\ SR = 90.0\%$	40000	35000	30000	25000
$\sigma = 5,\ SR = 99.9\%$	85000	75000	65000	55000

TABLE 6
Number of Required Measurement for the Absolute Difference Combining

n	1	2	3	4
$\sigma = 0,\ SR = 90.0\%$	20	40	60	90
$\sigma = 0,\ SR = 99.9\%$	30	60	130	190
$\sigma = 1,\ SR = 90.0\%$	900	800	800	700
$\sigma = 1,\ SR = 99.9\%$	1800	1700	1650	1600
$\sigma = 5,\ SR = 90.0\%$	420000	300000	225000	155000
$\sigma = 5,\ SR = 99.9\%$	800000	770000	410000	380000
n	5	6	7	8
$\sigma = 0,\ SR = 90.0\%$	130	170	210	250
$\sigma = 0,\ SR = 99.9\%$	270	340	440	550
$\sigma = 1,\ SR = 90.0\%$	700	750	750	800
$\sigma = 1,\ SR = 99.9\%$	1550	1500	1600	1750
$\sigma = 5,\ SR = 90.0\%$	115000	90000	80000	70000
$\sigma = 5,\ SR = 99.9\%$	200000	200000	170000	160000

4.3.1 Empirical Verification

In order to empirically verify the analysis carried out in the previous sections, we ran some 2O-DPA attack simulations according to the defined Hamming weight model. The targeted sensitive variable Z was a vector of $n \leq 8$ bits chosen among the output bits of the AES S-Box (taking $X \oplus K$ as input). The different values of X were randomly picked up to model a known (but not chosen) plaintext attack. Tables 5 and 6 give the number of measurements required to reach a success rate of either 90 percent or 99.9 percent for the product and the absolute difference according to the values of $n \in \{0, \cdots, 8\}$ and $\sigma \in \{0, 1, 5\}$ (10,000—resp. 1,000—simulations were performed for $\sigma \in \{0, 1\}$—resp. $\sigma = 5$).

Remark 19. We can observe that the results printed in Tables 5 and 6 match very well the correlation values given in Tables 1 and 3. Indeed, there is a kind of one-to-one correspondance between the correlation values and the number of measurements required to reach a given success rate. These results confirm that the correlation is a good indicator of the efficiency of an HO-DPA.

The number of measurements required by an HO-DPA quickly increases as the noise increases. Consequently, we were not able to derive some precise success rates for $\sigma \geq 10$. However, we have done several simulations with different noise deviations that all led to the same results: the number of measurements required to retrieve the targeted secret was almost all the time smaller for the product combining than for the absolute difference combining.

From our observations, we conclude that the product combining is more efficient than the absolute difference combining not only in the idealized but also in the noisy model (under the assumption that the leakage is normalized before being combined, as explained in Section 3.4).

4.4 Further Combining Functions

Other combining functions have been proposed in the literature [13], [16], [21]. In this section, we discuss these different proposals.

4.4.1 Raising to the Power

In [13], Joye et al. suggest to improve the efficiency of the absolute difference combining by raising it to a power α. They analyze the new combining functions $\mathcal{C}^{\alpha}_{diff^*}$ in the idealized model (corresponding to our model with $\sigma = 0$) for a single-bit 2O-DPA (i.e., with a binary combining function $f : z \mapsto z[i]$). Oswald et al. carry on with this approach in [16]: for a prediction function equal to the Hamming weight (i.e., $f : z \mapsto \mathrm{H}(z)$), they evaluate the correlation coefficients for $\mathcal{C}^{\alpha}_{diff^*}$ and $\mathcal{C}^{\alpha}_{prod^*}$ according to different α in the idealized model without offset (corresponding to our model with $\delta_1 = \delta_2 = 0$).

For several values n and α, we have computed in the idealized model the optimal correlations for both $\mathcal{C}^{\alpha}_{prod^*}$ and $\mathcal{C}^{\alpha}_{diff^*}$.[4] Table 7 lists the obtained values.

For both combining functions and every n, the maximum of the optimal correlations is reached for $\alpha = 1$. Thus, our analysis shows that raising the combined leakage to a power is not a sound approach to increase the efficiency of a 2O-DPA when the noise is null. This seems to contradict the analyses presented in [13], [16], where the authors report that raising to some values α improves the efficiency of the combining. The difference between our conclusions and the ones in [13], [16] is a consequence of the following fact: our study compares 2O-DPA that have been optimized by involving the optimal prediction function (introduced in Section 3.3) and by normalizing the leakage signals (as shown in Section 3.4). Besides, for every α we have tested, our correlation values are greater than the ones reported by Oswald et al. in [16].

In fact, we observed that raising to the power also decreases the efficiency of 2O-DPA in the noisy model. To summarize, our analysis suggests that raising the combining function to a power α decreases the efficiency of the second order DPA, the noise being null or not.

4. When n equals 1 and α is even, the product of the leakages does not depend on Z (and the expectation is constant with Z) which results in an undefined correlation.

TABLE 7
Optimal Correlation for $\mathcal{C}^{\alpha}_{prod^*}$ and $\mathcal{C}^{\alpha}_{diff^*}$

$\alpha \backslash n$	1	2	3	4	5	6	7	8
Product								
1	1.00	0.71	0.58	0.50	0.45	0.41	0.38	0.35
2	und.	0.58	0.37	0.27	0.21	0.17	0.15	0.13
3	1.00	0.71	0.50	0.39	0.33	0.29	0.26	0.24
4	und.	0.58	0.44	0.32	0.24	0.19	0.16	0.14
5	1.00	0.71	0.50	0.36	0.26	0.21	0.17	0.15
6	und.	0.58	0.45	0.33	0.24	0.18	0.14	0.11
Absolute difference								
1	1.00	0.65	0.50	0.41	0.35	0.31	0.28	0.26
2	1.00	0.58	0.45	0.38	0.33	0.30	0.28	0.26
3	1.00	0.60	0.45	0.37	0.33	0.29	0.27	0.25
4	1.00	0.62	0.45	0.36	0.31	0.28	0.25	0.24
5	1.00	0.64	0.45	0.35	0.30	0.26	0.24	0.22
6	1.00	0.65	0.45	0.35	0.29	0.25	0.22	0.20

4.4.2 Sine-Based Combining Function

In [21], Oswald and Mangard propose a combining function based on the sine function. It takes as parameters the exact Hamming weights of the mask and of the masked variable[5]:

$$\mathcal{C}_{sin}(\mathrm{H}(Z \oplus M), \mathrm{H}(M)) = \sin((\mathrm{H}(Z \oplus M) - \mathrm{H}(M))^2). \quad (31)$$

They also suggest to use the above combining function together with the following prediction function:

$$f_{sin}(Z) = -89.95 \sin(\mathrm{H}(Z))^3 - 7.82 \sin(\mathrm{H}(Z))^2 + 67.66 \sin(\mathrm{H}(Z)). \quad (32)$$

In the idealized model and for $n = 8$, the use of the couple $(\mathcal{C}_{sin}, f_{sin})$ allows an attacker to reach a correlation of 0.83, which is quite high. However, f_{sin} is not optimal. Indeed, Corollary 8 states that the optimal prediction function for $\mathcal{C}_{sin}$ is the function f_{opt} defined by

$$f_{opt}(Z) = \mathrm{E}_M[\mathcal{C}_{sin}(\mathrm{H}(Z \oplus M), \mathrm{H}(M))]. \quad (33)$$

Actually, for such a function, we have $\rho(f_{sin}, f_{opt}) = 0,97$, which implies that the use of f_{sin} instead of f_{opt} results in an efficiency loss of 3 percent.

Without the above improvement, it is difficult to compare the efficiencies of $\mathcal{C}_{sin}$ and $\mathcal{C}_{prod^*}$. Indeed, the attack scenario presented in [21] does not correspond to the kind of attacker we focus in this paper (see Section 3.1). In [21], the authors consider a very strong adversarial model where the attacker is able to recover the exact Hamming weights of the mask and the masked variable based on preprocessed templates (see [2] for further details on *Template Attacks*). However, in such a scenario, combining the obtained Hamming weights is a suboptimal attack strategy and, as explained in [21], a better strategy is to use a Bayesian classification (or maximum likelihood test). Moreover, the recovering of the exact Hamming weight values is only possible in an almost noise-free model.

As argued at the beginning of this section, in a classical HO-DPA scenario, the evaluation process of a combining function must include an analysis in a noisy environment.

5. The formulas given in [21] are erroneous, and (31) and (32) are their corrected versions.

TABLE 8
Correlations for $\mathcal{C}_{sin}$ and $\mathcal{C}_{prod^\star}$ According to σ

$(\mathcal{C}, f) \setminus \sigma$	0	0.1	0.3	0.4	0.5	0.7	1	5
$(\mathcal{C}_{sin}, f_{opt})$	0.87	0.74	0.38	0.21	0.11	0.05	0.037	0
$(\mathcal{C}_{sin}, f_{sin})$	0.83	0.70	0.35	0.19	0.08	0.01	0	0
$(\mathcal{C}_{prod^\star}, \mathrm{H})$	0.36	0.36	0.34	0.33	0.32	0.29	0.24	0.03

Therefore, we analyzed the efficiency of the sine-based combining in the presence of noise. Namely, we added Gaussian noises $N_1, N_2 \sim \mathcal{N}(0, \sigma)$ to the Hamming weights in (33) and (35). We list in Table 8 the values of the correlation according to an increasing noise (with n equal to 8).

It can be observed that the correlation for $\mathcal{C}_{sin}$ quickly decreases as σ increases. For a noise deviation σ greater than or equal to 0.4 (which is quite low), the product combining offers a greater correlation. This suggests that in an HO-DPA scenario (where the leakage is noisy), the sine-based combining function is not suitable.

4.4.3 Final Comparison

To conclude this section, Fig. 3 plots the correlations ρ_K with respect to the noise deviation $\sigma \in [0, 2]$ for the combining functions $\mathcal{C}_{sin}$, $\mathcal{C}^{\alpha}_{prod^\star}$, and $\mathcal{C}^{\alpha}_{diff^\star}$, $\alpha \in \{1, 2, 3\}$. This plot underlines the previous conclusion: among the known combining functions, the improved product combining offers the best efficiency in a general leakage model.

5 Conclusion

In this paper, we have investigated higher order DPA attacks that combine several leakage signals to defeat masking countermeasures. We have first defined a theoretical framework allowing us to evaluate the efficiency of such an HO-DPA and have shown how to optimize it according to the combining technique and the leakage model. This enabled us to study the existing combining techniques for second order DPA in the Hamming weight model with noise, paying particular attention to product combining and absolute difference combining. Our analysis allowed us to exhibit a way of significantly improving the product combining in this model and we showed that this improved product combining is more efficient than all the other techniques previously proposed in the literature.

Our work introduces the basis for a practically oriented analysis of HO-DPA attacks that may be used for future research. In particular, the framework proposed in this paper makes it possible to analyze the efficiency of new combining techniques in a general model. Moreover, our approach could be extended to the investigation HO-DPA of orders greater than 2.

Appendix 1
Useful Lemmas

Lemma 20. *Let n be a positive integer and let M be a random variable uniformly distributed over $\mathbb{F}_2^n$. Then, we have*

$$\mathrm{E}[\mathrm{H}(M)^2] = \frac{n^2+n}{4}. \quad (34)$$

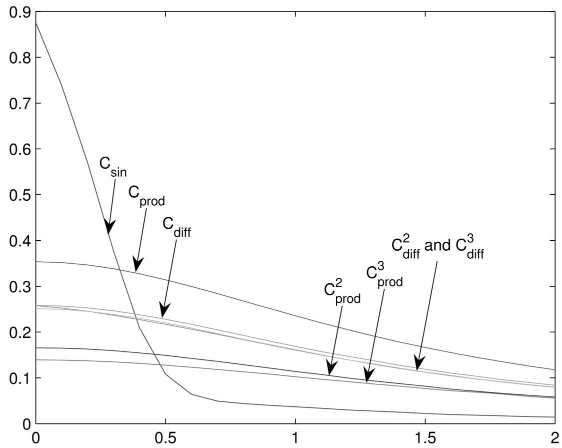

Fig. 3. Correlation ρ_K for different combining functions according to the noise deviation σ.

Proof. Since M is uniformly distributed over $\mathbb{F}_2^n$, we have

$$\mathrm{E}[\mathrm{H}(M)^2] = \mathrm{E}\left[\sum_{i,j=1}^{n} M[i]M[j]\right],$$

that is

$$\mathrm{E}[\mathrm{H}(M)^2] = \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathrm{E}[M[i]M[j]] + \sum_{i=1}^{n} \mathrm{E}[M[i]].$$

For every $i \neq j$, we have $\mathrm{E}[M[i]M[j]] = \frac{1}{4}$ and $\mathrm{E}[M[i]] = \frac{1}{2}$. Hence, we deduce that $\mathrm{E}[\mathrm{H}(M)^2] = n(n-1) \times \frac{1}{4} + n \times \frac{1}{2} = \frac{n^2+n}{4}$.

Lemma 21. *Let n be a positive integer and M be a random variable uniformly distributed over $\mathbb{F}_2^n$. Then, for every $z \in \mathbb{F}_2^n$, we have*

$$\mathrm{E}[\mathrm{H}(z \oplus M)\mathrm{H}(M)] = -\frac{1}{2}\mathrm{H}(z) + \frac{n^2+n}{4}. \quad (35)$$

Proof. From Property 2, we have

$$\begin{aligned}\mathrm{E}[\mathrm{H}(z \oplus M)\mathrm{H}(M)] = \mathrm{H}(z)\mathrm{E}[\mathrm{H}(M)] + \mathrm{E}[\mathrm{H}(M)^2] \\ - 2\mathrm{E}[\mathrm{H}(z \wedge M)\mathrm{H}(M)]\end{aligned}. \quad (36)$$

Since M is uniformly distributed, we have $\mathrm{E}[\mathrm{H}(M)] = \frac{n}{2}$ and $\mathrm{E}[\mathrm{H}(M)^2] = \frac{n^2+n}{4}$ (from Lemma 20). On the other hand, $\mathrm{E}[\mathrm{H}(z \wedge M)\mathrm{H}(M)]$ satisfies

$$\mathrm{E}[\mathrm{H}(z \wedge M)\mathrm{H}(M)] = \sum_{i=0}^{n} z[i]\mathrm{E}[M[i]\mathrm{H}(M)]. \quad (37)$$

Since M is uniformly distributed over $\mathbb{F}_2^n$, $\mathrm{E}[M[i]\mathrm{H}(M)]$ is equal to $\frac{1}{n}\mathrm{E}[\mathrm{H}(M)^2]$, i.e., to $\frac{n+1}{4}$ (from Lemma 20). Hence, simplifying (36) leads to (35). □

Lemma 22. *Let m and n be two integers and r be a positive integer:*

$$\sum_k \binom{r}{m+k}\binom{s}{n+k} = \binom{r+s}{r-m+n}. \quad (38)$$

Proof. Lemma 22 is a well-known result whose proof can be found in [22]. □

Appendix 2
Proofs of Propositions 14 and 16

2.1 Proof of Proposition 14

Proof. For every pair $(z, m) \in \mathbb{F}_2^n$, Property 2 implies $|\mathrm{H}(z \oplus m) - \mathrm{H}(m)| = |\mathrm{H}(z) - 2\mathrm{H}(z \wedge m)|$ from which we deduce

$$\mathrm{E}[|\mathrm{H}(z \oplus M) - \mathrm{H}(M)|] = \sum_{i=0}^{\mathrm{H}(z)} |\mathrm{H}(z) - 2i| \mathrm{P}[\mathrm{H}(z \wedge M) = i]. \tag{39}$$

Since M is uniformly distributed, $\mathrm{P}[\mathrm{H}(z \wedge M) = i]$ equals $2^{-\mathrm{H}(z)}\binom{h}{i}$. Hence, we deduce

$$\mathrm{E}[|\mathrm{H}(z \oplus M) - \mathrm{H}(M)|] = 2^{-\mathrm{H}(z)} \sum_{i=0}^{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor} \binom{\mathrm{H}(z)}{i} (\mathrm{H}(z) - 2i). \tag{40}$$

By symmetry, we have $\sum_{i=0}^{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor} \binom{\mathrm{H}(z)}{i}$ equal to $\frac{1}{2}(\sum_{i=0}^{\mathrm{H}(z)} \binom{\mathrm{H}(z)}{i} + \binom{\mathrm{H}(z)}{\frac{\mathrm{H}(z)}{2}} (\mathrm{H}(z) \bmod 2))$. Then, $\sum_i \binom{\mathrm{H}(z)}{i} = 2^{\mathrm{H}(z)}$ implies

$$\sum_{i=0}^{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor} \binom{\mathrm{H}(z)}{i} = 2^{\mathrm{H}(z)-1} + \frac{1}{2} \binom{\mathrm{H}(z)}{\frac{\mathrm{H}(z)}{2}} (\mathrm{H}(z) + 1 \bmod 2). \tag{41}$$

On the other hand, $\binom{\mathrm{H}(z)}{i} i$ equals $\mathrm{H}(z)\binom{\mathrm{H}(z)-1}{i-1}$, which in a similar way gives

$$\sum_{i=0}^{\lfloor \frac{\mathrm{H}(z)}{2} \rfloor} \binom{\mathrm{H}(z)}{i} i = \frac{\mathrm{H}(z)}{2} 2^{\mathrm{H}(z)-1} - \frac{\mathrm{H}(z)}{2} \binom{\mathrm{H}(z)-1}{\frac{\mathrm{H}(z)-1}{2}} \times (\mathrm{H}(z) \bmod 2). \tag{42}$$

Finally, (40), (41), and (42) lead to (25). □

2.2 Proof of Proposition 16

Proof. Let ϕ_B and Φ_B, respectively, denote the probability density function and the probability distribution function of B (that is, $\Phi_B(y) = \mathrm{P}[\mathrm{B} \leq y] = \int_{-\infty}^{\mathrm{y}} \phi_{\mathrm{B}}(x) \mathrm{d}x$). As B has a gaussian distribution $\mathcal{N}(0, \sigma_0)$, we have $\phi_B(x) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp(-x^2/2\sigma_0^2)$. Then, we have

$$\begin{aligned} \mathrm{E}[|s + B|] &= \int_{-\infty}^{+\infty} |s + x| \phi_B(x) dx \\ &= s \int_{-s}^{s} \phi_B(x) dx + \int_{-s}^{s} x\phi_B(x) dx \\ &\quad + 2 \int_{s}^{+\infty} x\phi_B(x) dx. \end{aligned}$$

Since the function $x \mapsto x\phi_B(x)$ is odd, the term $\int_{-s}^{s} x\phi_B(x)dx$ equals zero. Moreover, we have $\int_{-s}^{s} \phi_B(x)dx = 2(\Phi_B(s) - \frac{1}{2})$ and $\int_{s}^{+\infty} x\phi_B(x)dx = \frac{\sigma_0}{\sqrt{2\pi}} \exp(-s^2/2\sigma_0^2)$. Hence, we get

$$\mathrm{E}[|s + B|] = 2s\left(\Phi_B(s) - \frac{1}{2}\right) + \frac{\sqrt{2}\sigma_0}{\sqrt{\pi}} \exp(-s^2/2\sigma_0^2). \tag{43}$$

Finally, since B has a gaussian distribution $\mathcal{N}(0, \sigma_0)$, its probability distribution function Φ_B satisfies $\Phi_B(y) = \frac{1}{2}(1 + \mathrm{erf}(\frac{y}{\sqrt{2}\sigma_0}))$ for every $y \in \mathbb{R}$; hence, (43) directly implies (27). □

Acknowledgments

R. Bévan worked for Oberthur Technologies during the redaction of this paper.

References

[1] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," *Advances in Cryptology—Proc. Int'l Crytology Conf. (CRYPTO '99),* M. Wiener, ed., pp. 388-397, 1999.

[2] S. Chari, J. Rao, and P. Rohatgi, "Template Attacks," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2002),* B. Kaliski, Jr., Ç. Koç, and C. Paar, eds., pp. 13-29, 2002.

[3] W. Schindler, K. Lemke, and C. Paar, "A Stochastic Model for Differential Side Channel Cryptanalysis," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2005),* J. Rao and B. Sunar, eds., pp. 30-46, 2005.

[4] C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater, "Template Attacks in Principal Subspaces," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2006),* L. Goubin and M. Matsui, eds., pp. 1-14, 2006.

[5] S. Chari, C. Jutla, J. Rao, and P. Rohatgi, "Towards Sound Approaches to Counteract Power-Analysis Attacks," *Advances in Cryptology—Proc. Int'l Crytology Conf. (CRYPTO '99),* M. Wiener, ed., pp. 398-412, 1999.

[6] L. Goubin and J. Patarin, "DES and Differential Power Analysis—The Duplication Method," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES '99),* Ç. Koç and C. Paar, eds., pp. 158-172, 1999.

[7] K. Schramm and C. Paar, "Higher Order Masking of the AES," *Topics in Cryptology—Proc. Cryptographers' Track (CT)-RSA Conf. 2006,* pp. 208-225, 2006.

[8] J.-S. Coron, E. Prouff, and M. Rivain, "Side Channel Cryptanalysis of a Higher Order Masking," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2007),* P. Paillier and I. Verbauwhede, eds., pp. 28-44, 2007,

[9] T. Messerges, "Securing the AES Finalists Against Power Analysis Attacks," *Proc. Int'l Conf. Fast Software Encryption (FSE 2000),* B. Schneier, ed., pp. 150-164, 2000.

[10] M.-L. Akkar and C. Giraud, "An Implementation of DES and AES, Secure against Some Attacks," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2001),* Ç. Koç, D. Naccache, and C. Paar, eds., pp. 309-318, 2001.

[11] T. Messerges, "Using Second Order Power Analysis to Attack DPA Resistant Software," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2000),* Ç. Koç and C. Paar, eds., pp. 238-251, 2000.

[12] J. Waddle and D. Wagner, "Toward Efficient Second Order Power Analysis," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2004),* M. Joye and J.-J. Quisquater, eds., pp. 1-15, 2004.

[13] M. Joye, P. Paillier, and B. Schoenmakers, "On Second Order Differential Power Analysis," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2005),* J. Rao and B. Sunar, eds., pp. 293-308, 2005.

[14] E. Peeters, F.-X. Standaert, N. Donckers, and J.-J. Quisquater, "Improving Higher-Order Side Channel Attacks with FPGA Expermiments," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2005),* J. Rao and B. Sunar, eds., pp. 309-321, 2005.

[15] F.-X. Standaert, E. Peeters, and J.-J. Quisquater, "On the Masking Countermeasure and Higher-Order Power Analysis Attacks," *Proc. Int'l Conf. Information Technology: Coding and Computing (ITCC '05),* vol. I, pp. 562-567, 2005.

[16] E. Oswald, S. Mangard, C. Herbst, and S. Tillich, "Practical Second Order DPA Attacks for Masked Smart Card Implementations of Block Ciphers," *Topics in Cryptology—Proc. Cryptographers' Track (CT)-RSA Conf. 2006*, pp. 192-207, 2006.
[17] E. Brier, C. Clavier, and F. Olivier, "Correlation Power Analysis with a Leakage Model," *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2004),* M. Joye and J.-J. Quisquater, eds., pp. 16-29, 2004.
[18] S. Mangard, "Hardware Countermeasures against DPA—A Statistical Analysis of Their Effectiveness," *Topics in Cryptology—Proc. Cryptographers' Track (CT)-RSA Conf. 2004,* T. Okamoto, ed., pp. 222-235, 2004.
[19] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks—Revealing the Secrets of Smartcards.* Springer, 2007.
[20] F.-X. Standaert, E. Peeters, G. Rouvroy, and J.-J. Quisquater, "An Overview of Power Analysis Attacks against Field Programmable Gate Arrays," *Proc. IEEE,* vol. 94, no. 2, pp. 383-394, Feb. 2006.
[21] E. Oswald and S. Mangard, "Template Attacks on Masking-Resistance is Futile," *Topics in Cryptology—Proc. Cryptographers' Track (CT)-RSA 2007 Conf.*, M. Abe, ed., pp. 562-567, 2007.
[22] D. Knuth, *The Art of Computer Programming,* third ed., vol. 1. Addison Wesley, 1997,
[23] *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2004),* M. Joye and J.-J. Quisquater, eds., 2004.
[24] *Advances in Cryptology—Proc. Int'l Crytology Conf. (CRYPTO '99),* M. Wiener, ed., 1999.
[25] *Topics in Cryptology—Proc. Cryptographers' Track (CT)-RSA Conf. 2006,* 2006.
[26] *Proc. Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES 2005),* J. Rao and B. Sunar, eds., 2005.

Emmanuel Prouff received the PhD degree in computer science and applied mathematics from the University of Caen, France, in 2004. He has been a university lecturer in computer science at the University of Orsay (Paris XI) and at the French Engineering School of Bourges. He is currently working at Oberthur Technologies for payment and GSM smartcard security. His specific research interests include theoretical aspects of symmetric cryptology and the different ways of thwarting side channel attacks.

Matthieu Rivain received the BS degree from the University of Rennes 1 in 2003 and the MS degree from the Grande Ecole of Computer Science and Applied Mathematics of Grenoble (ENSIMAG) in 2006. Currently, he is a PhD student at the University of Luxembourg and Oberthur Technologies. His research interests include cryptography, cryptographic implementations, and physical attacks (side channel and fault analysis).

Régis Bévan received the MS degree in electronics and signal processing in 1999, and the PhD degree in computer science and signal processing in 2004. His PhD thesis was about side channel attacks against smartcards. He worked from 1999 to 2007 in the Cryptography and Security Group at Oberthur for payment and GSM smartcard security. Currently, he works at Nagravision for PayTV security. His main scientific interests are computer security and cryptography.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.