Confused yet successful: Theoretical computation of distinguishers for monobit leakages in terms of confusion coefficient and SNR

Eloi De Cherisey, Sylvain Guilley, Olivier Rioul

Confused yet Successful:
Theoretical Comparison of Distinguishers for Monobit Leakages in Terms of Confusion Coefficient and SNR

Eloi de Chérisey[1(✉)], Sylvain Guilley[1,2], and Olivier Rioul[1]

[1] Télécom ParisTech, Paris, France
{eloi.decherisey,sylvain.guilley,olivier.rioul}@telecom-paristech.fr
[2] Secure-IC S.A.S., Rennes, France

Abstract. Many side-channel distinguishers (such as DPA/DoM, CPA, Euclidean Distance, KSA, MIA, etc.) have been devised and studied to extract keys from cryptographic devices. Each has pros and cons and find applications in various contexts. These distinguishers have been described theoretically in order to determine which distinguisher is best for a given context, enabling an unambiguous characterization in terms of success rate or number of traces required to extract the secret key.

In this paper, we show that in the case of monobit leakages, the theoretical expression of all distinguishers depend only on two parameters: the confusion coefficient and the signal-to-noise ratio. We provide closed-form expressions and leverage them to compare the distinguishers in terms of convergence speed for distinguishing between key candidates. This study contrasts with previous works where only the asymptotic behavior was determined—when the number of traces tends to infinity, or when the signal-to-noise ratio tends to zero.

Keywords: Side-channel distinguisher · Differential Power Analysis (DPA) · Difference of Means (DoM) · Correlation Power Analysis (CPA) · Mutual Information Analysis (MIA) · Kolmogorov-Smirnov Analysis (KSA) · Confusion coefficient · Signal-to-noise ratio · Success rate · Success exponent

1 Introduction

Today's ciphering algorithms such as AES are considered resistant to cryptanalysis. This means that the best possible way to extract a 128-bit key is about as complex as an exhaustive search over the 2^{128} possibilities. With our current computational power, this is not achievable within a reasonable amount of time.

However, it is possible to use plaintexts, ciphertexts, along with additional side information in order to recover the secret key of a device. Indeed, the secret key may leak via *side-channels*, such as the time to compute the algorithm, the power consumption of the device during the computation of the algorithm, or the electro-magnetic radiations of the chip.

In order to secure chips from side-channel attacks, designers have to understand how these work and what could be the future security breaches in the cryptographic algorithm as well as in the hardware implementation. A preliminary step is to identify how the secret keys leak and deduce leakage models. Then, mathematical functions—called *distinguishers*—take the leakage as argument and return an estimation of the secret key. Such distinguishers come in many flavours[1] and have different figures of merit in different contexts. A given context not only involves the cryptographic algorithm and the device through the leakage model, but also the side-channel acquisition setup through the measurement characterized by its signal-to-noise ratio (SNR). This is illustrated in Fig. 1 borrowed from Heuser *et al.* [12] (with our annotations in red).

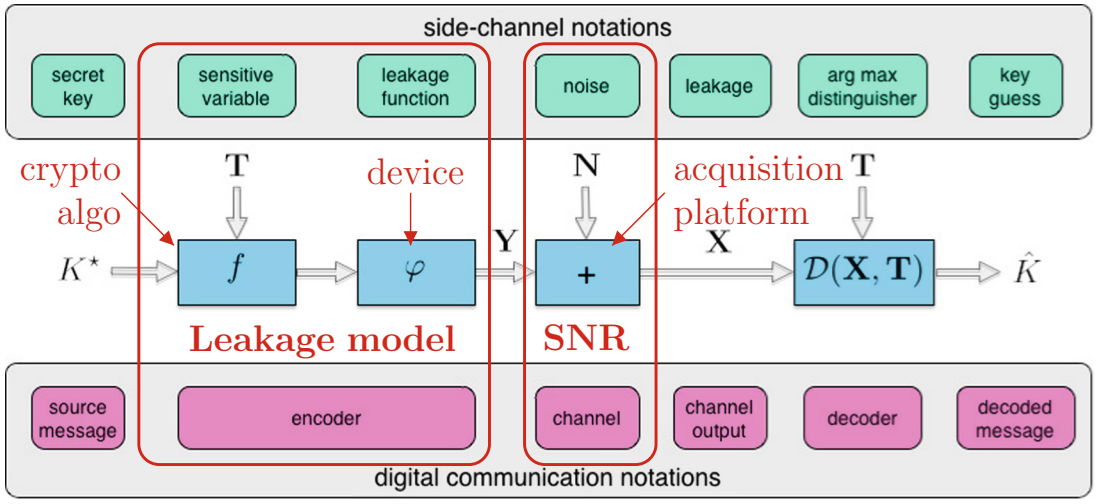

Fig. 1. Illustration of the two parts of the side-channel analysis context (in red). (Color figure online)

In practice one may encounter *monobit* leakages. This means that the output of the leakage model can only take two values. In this case, as we shall see, the mathematical computations turn to be simpler and information theoretic tools can be used to precisely describe the link between the leakage model and the real-world leaking traces. From another perspective, considering monobit leakages can also be seen as an "abstraction" trick meant to intentionally ignore the complex effect of the way the device leaks, thereby keeping only the contribution from the cryptographic algorithm in the leakage model.

A related question is how the choice of the substitution box in the cryptographic algorithm may "help" the attacker. The standard AES substitution box was designed to be very secure against linear and differential cryptanalysis [6]. On the contrary, under side-channel analysis, the substitution box may be helpful for the attacker, especially for monobit leakages as shown below.

[1] We cover in this paper the following distinguishers: Difference of Means (DoM) [13], Correlation Power Analysis (CPA) [3], Euclidean distance [12, §3], Kolmogorov-Smirnov Analysis (KSA) [22], and Mutual Information Analysis (MIA) [9].

Related Work. Distinguishers were often studied empirically, yet such an approach does not allow for generalizations to other contexts and measurement campaigns. A theoretical approach consists in analyzing the formal expressions of the distinguishers as mathematical functions. Fei et al. [8] have shown that distinguishers such as DoM and CPA can be expressed in terms of a *confusion coefficient*. They gave the impetus to extend this formal analysis to other types of distinguishers. In 2014, Heuser *et al.* [11] relate KSA to the confusion coefficient, and also noticed that the confusion coefficient can be related to the resistance of a substitution box against differential cryptanalysis.

Whitnall and Oswald [21] have proposed the *relative distinguishing margin* metric to compare distinguishers. However, it has been shown [18] that this metric may not be relevant in all contexts. Another way to compare distinguishers is to contrast how their success rate (SR) in key recovery depends on the number q of side-channel traces. Works such as [8,14] provide mathematically models for the SR. But the comparison between different distinguishers has never been actually carried out based on such frameworks. Instead, we shall leverage on the so-called *success exponent* (SE) [10] which allows to compare the SR of various distinguishers based on only one exponent parameter.

Our Contributions. In this paper, we consolidate the knowledge about side-channel attacks exploiting monobit leakages. We provide a rigorous proof that any distinguisher acting on monobit leakages depends only on two parameters: the *confusion coefficient* and the *noise variance*. Some distinguishers, namely DoM, CPA and KSA, have already been expressed as a function of those two parameters [8,11]. In this article, we derive this expression for MIA and we obtain a simple analytic function when the non zero values of the confusion coefficient are near $1/2$, which is the case of leakages occurring at cryptographically strong substitution boxes [4].

We derive the success exponent of these distinguishers in terms of the confusion coefficient and the standard deviation of the noise. Success exponents allow to characterize the efficiency (in terms of number of traces) of distinguishers to recover the key. Our closed-form expressions of the success exponent enable the comparison of distinguishers based only on these two parameters. The flow chart of Fig. 2 situates our contributions in relation to the current state of the art.

Organization. The remainder of this paper is organized as follows. In Sect. 2, we recall the main definitions. In Sect. 3, we consider all distinguishers in one mathematical framework and we show that they are only functions of two parameters. In Sect. 4, we compare the distinguishers in terms of the success exponent. Section 5 concludes. Appendices provide proofs for technical lemmas.

Notations. Throughout this paper, we use calligraphic letters to denote sets and lower-case letters for elements in this set (e.g. $x \in \mathcal{X}$). Capital letters denote random variables. For example, X is a random variable taking values in $\mathcal{X}$ and $x \in \mathcal{X}$ is a realization of X. The probability that X is x is noted $\mathbb{P}(X = x)$ or simply $\mathbb{P}(x)$ when there is no ambiguity. The expectation of a random variable is

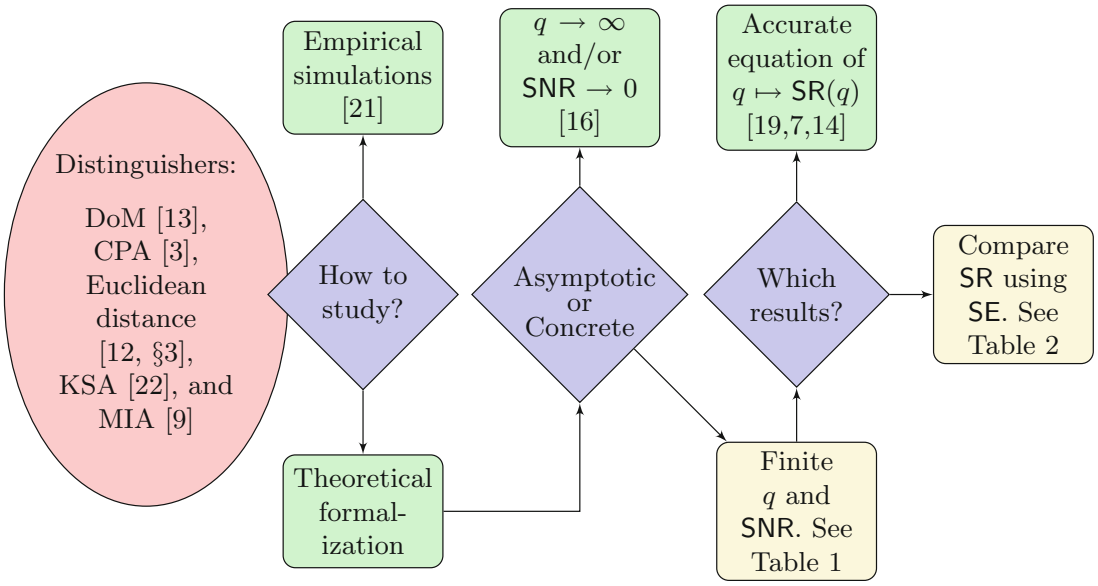

Fig. 2. The state of the art in relation to our contributions (in yellow boxes—see also Tables 1 and 2 below). (Color figure online)

noted $\mathbb{E}[X]$ and its variance $\mathrm{Var}(X)$. The differential entropy $h(X)$ of a random variable X following distribution $p(x)$ is defined as

$$h(X) = -\int_{\mathbb{R}} p(x) \log_2 p(x) \, \mathrm{d}x. \tag{1}$$

The mutual information between two random variables X and Y is defined as

$$I(X;Y) = h(X) - h(X|Y) = \mathbb{E}\left[\log_2 \frac{\mathbb{P}(X,Y)}{\mathbb{P}(X)\mathbb{P}(Y)}\right]. \tag{2}$$

2 Modelization and Definitions

2.1 The Leakage Model

In order to compare the different distinguishers for monobit leakages, we need a leakage model upon which our computations will be based. A plaintext t meets the secret key k^* through a leakage function $f(t, k^*)$. The resulting variable $y(k^*)$ is called the *sensitive* variable. The dependence in the plaintext t will be omitted to make equations easier to read when there is no ambiguity.

The attacker measures a noisy version of $y(k^*)$ called *trace* and denoted by x. When the key is unknown, the attacker computes a sensitive variable with a key hypothesis k, that is, $y(k) = f(t, k)$. Thus our model takes the form

$$\begin{cases} y(k) = f(t,k) \\ \quad x = y(k^*) + n \end{cases} \tag{3}$$

where n is an independent measurement noise.

As we consider monobit leakages, we suppose that $y(k)$ can take only two values. In practice, t (resp. k) are subsets of the full plaintext (resp. key). Typically, in the case of AES where attacks can be conducted using a divide-and-conquer approach on a per substitution box basis, t and k are 8-bit works (i.e., bytes).

The above leakage model can also be written using random variables. Let T the random variable for the plaintext, $Y(k)$ for the sensitive variable, X for the measurement, and N for the Gaussian noise. We have:

$$\begin{cases} Y(k) = f(T, k) \\ \quad X = Y(k^*) + N. \end{cases} \tag{4}$$

In a view to simplify further mathematical computations, we suppose that the leakage random variable is reduced, that is, centered ($\mathbb{E}[Y(k)] = 0$ for all k) and of unit variance ($\mathbb{E}[Y(k)^2] = 1$ for all k). The noise is also assumed Gaussian of zero mean and its standard deviation is noted $\sigma > 0$. Moreover, we assume that for any key hypothesis the sensitive variable is *balanced*, that is, $\mathbb{P}(y(k)) = \frac{1}{2}$. Since $Y(k)$ is a binary random variable, we necessarily have that $Y(k) \in \{\pm 1\}$ in our model, and consequently the signal-to-noise ratio equals $\mathsf{SNR} = 1/\sigma^2$.

Last, we suppose that the attacker has at his disposal a number of q traces $x_1, \ldots, x_q$ obtained from leaking sensitive variables $y_1(k^*), \ldots, y_q(k^*)$ under additive noise $n_1, \ldots, n_q$.

2.2 The Confusion Coefficient

In the side-channel context, the confusion coefficient was defined by Fei et al. as the probability that two sensitive variables arising from two different key hypotheses are different [8, Section 3.1]. Mathematically, the confusion coefficient is written as

$$\kappa(k, k^*) = \mathbb{P}(Y(k) \neq Y(k^*)). \tag{5}$$

As the secret key k^* is constant and understood from the context, we can write $\kappa(k, k^*) = \kappa(k)$. Notice that in practical situations, the EIS (Equal Images under different Subkeys [20, Def. 2]) assumption holds, therefore κ is actually a function of the key bitwise XOR difference $k \oplus k^*$.

Figure 3 illustrates the confusion coefficient for a monobit leakage $Y(k) = \mathsf{SubBytes}(T \oplus k) \bmod 2$, where $\mathsf{SubBytes}$ is the AES substitution box (application $\mathbb{F}_2^8 \to \mathbb{F}_2^8$) and $\oplus$ is the bitwise exclusive or. We notice that except for $k = k^*$ (here taken $= 178 = \texttt{0xb2}$), the confusion coefficient for the AES $\mathsf{SubBytes}$ is close to $1/2$. This results from the fact the AES $\mathsf{SubBytes}$ has been designed to be resistant against differential cryptanalysis. Specifically, Heuser et al. [11, Proposition 6] noticed that a "good" substitution box leads to confusion coefficients near $1/2$.

The original definition of the confusion coefficient [8] considers only monobit leakages. An extension for any type of leakage was proposed in [10] where $\kappa(k)$ is defined by

$$\kappa(k) = \mathbb{E}\left[\left(\frac{Y(k^*) - Y(k)}{2}\right)^2\right]. \tag{6}$$

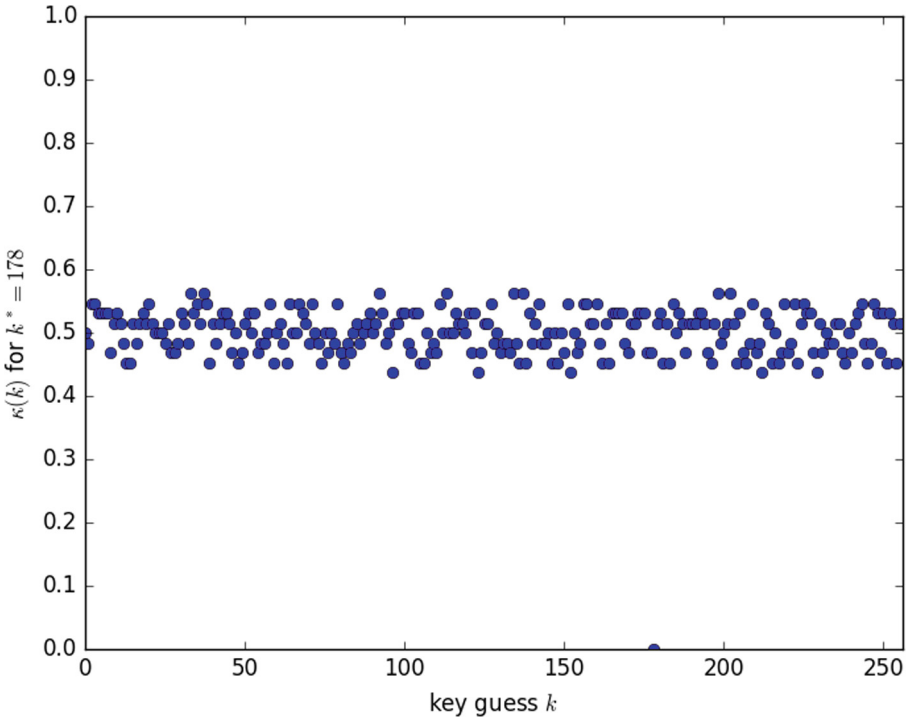

Fig. 3. Confusion coefficient for the AES SubBytes Least Significant Bit (LSB)

Equation (5) can be easily recovered from this more general expression by noting that when $Y(k)$ and $Y(k^*) \in \{\pm 1\}$, $\left(\frac{Y(k^*)-Y(k)}{2}\right)^2$ is 0 or 1 according to whether $Y(k) = Y(k^*)$ or $Y(k) \neq Y(k^*)$.

2.3 Distinguishers

Distinguishers aim at recovering the secret key k^* from the traces and the model. For every key k, the attacker computes the associated distinguisher. The key hypothesis that gives the highest value of the distinguisher is the estimated key. The attack is successful if the estimated key is equal to the secret key.

For every key hypothesis k, a distinguisher is noted $\widehat{\mathcal{D}}(k)$ and the estimated key is $\widehat{k} = \arg\max_k \widehat{\mathcal{D}}(k)$. Five classical distinguishers are:

– Difference of Means (DoM) [8], also known as the Differential Power Analysis (DPA) [13] where the attacker computes

$$\widehat{\mathcal{D}}(k) = \frac{\sum_{i|y_i(k)=+1} x_i}{\sum_{i|y_i(k)=+1}} - \frac{\sum_{i|y_i(k)=-1} x_i}{\sum_{i|y_i(k)=-1}}. \tag{7}$$

– Correlation Power Analysis (CPA) [3] where the attacker computes the absolute value of the Pearson coefficient

$$\widehat{\mathcal{D}}(k) = \left| \frac{\frac{1}{q}\sum_{i=1}^{q} x_i y_i(k) - \frac{1}{q}\sum_{i=1}^{q} x_i \cdot \frac{1}{q}\sum_{i=1}^{q} y_i(k)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y_i(k))}} \right|. \tag{8}$$

Notice that $\mathrm{Var}(Y_i(k))$ do not depend on the index i, since repeated measurements are i.i.d.

– Euclidean distance, which corresponds to the Maximum Likelihood (ML) attack under the Gaussian noise hypothesis, where the attacker actually computes the negative Euclidean distance between the model and the trace

$$\widehat{\mathcal{D}}(k) = -\frac{1}{q}\sum_{i=1}^{q}(x_i - y_i(k))^2. \tag{9}$$

Maximizing the value of the distinguisher amounts to minimizing the Euclidean distance. According to [12], as the noise is Gaussian and additive, the Euclidean distance is the optimal distinguishing rule (ML rule) that maximizes the success probability.

– Kolmogorov-Smirnov Analysis (KSA) [22] where the traces are used to build an estimation of the cumulative density function $\widehat{F}(x)$, and the distinguisher is

$$\widehat{\mathcal{D}}(k) = -\mathbb{E}_{Y(k)}\left[\|\widehat{F}(x|Y(k)) - \widehat{F}(x)\|_\infty\right] \tag{10}$$

where the infinite norm is defined as $\|\widehat{F}(x)\|_\infty = \sup_x |\widehat{F}(x)|$. Maximizing the value of the distinguisher amounts to minimizing the expected infinite norm.

– Mutual Information Analysis (MIA) [9] where the attacker computes the mutual information between the traces and each model. The traces are used to build an estimation of the joint distribution of X and $Y(k)$, denoted by $\widehat{p}(X, Y(k))$, and with this estimation, we calculate the mutual information

$$\widehat{\mathcal{D}}(k) = \sum_{x,y(k)} \widehat{p}(x, y(k)) \log_2 \frac{\widehat{p}(x, y(k))}{\widehat{p}(x) \cdot \widehat{p}(y(k))}. \tag{11}$$

Given the available data, the attacker computes the distinguisher as a function of $x_1, \ldots, x_q$ and $y_1(k), \ldots, y_q(k)$. To emphasize the dependence on the data, we may write $\widehat{\mathcal{D}}(k) = \widehat{\mathcal{D}}(X_1, \ldots, X_q, Y_1(k), \ldots, Y_q(k))$. As these traces are realizations of random variables, we may also consider $\widehat{\mathcal{D}}(k)$ as a random variable which is a function of $X_1, \ldots, X_q$ and $Y_1(k), \ldots, Y_q(k)$, with expectation $\mathbb{E}[\widehat{\mathcal{D}}(k)]$ and a variance $\mathrm{Var}(\widehat{\mathcal{D}}(k))$.

When the number of queries q tends to infinity, we assume that the distinguisher converges in the mean-squared sense:

Definition 1 (Theoretical Distinguisher [10]). *The theoretical value of the distinguisher is defined as the limit in the mean square sense when $q \to \infty$ of the distinguisher. The notation for the theoretical distinguisher is $\mathcal{D}(k)$, which is therefore implicitly defined as:*

$$\mathbb{E}[(\widehat{\mathcal{D}}(k) - \mathcal{D}(k))^2] \longrightarrow 0 \text{ as } q \to \infty. \tag{12}$$

Put differently, $\widehat{\mathcal{D}}(k)$ can be seen as an estimator of $\mathcal{D}(k)$. It is easily seen that as $q \to +\infty$ the distinguishers presented previously have the following theoretical distinguishers:

- For DoM, the theoretical distinguisher is

$$\mathcal{D}(k) = \mathbb{E}[XY(k)]. \tag{13}$$

- For CPA, the theoretical distinguisher is

$$\mathcal{D}(k) = \frac{\left|\mathbb{E}[XY(k)] - \mathbb{E}[X]\mathbb{E}[Y(k)]\right|}{1+\sigma^2}. \tag{14}$$

- For Euclidean distance (ML) distinguisher, we have:

$$\mathcal{D}(k) = -\mathbb{E}\big[(X - Y(k))^2\big]. \tag{15}$$

- For KSA, we have:

$$\mathcal{D}(k) = \mathbb{E}_{Y(k)}\big[\|F(x|Y(k)) - F(x)\|_\infty\big]. \tag{16}$$

- For MIA, it is the mutual information

$$\mathcal{D}(k) = I(X; Y(k)). \tag{17}$$

3 Theoretical Expressions for Distinguishers

In this section, we show that all distinguishers for monobit leakages are functions of only two parameters: the confusion coefficient $\kappa(k)$ and the $\mathsf{SNR} = 1/\sigma^2$. This is confirmed by the closed-form expressions for classical distinguishers. In particular we derive the one corresponding to MIA.

3.1 A Communication Channel Between $Y(k)$ and $Y(k^*)$

To understand the link between any sensitive variable $Y(k)$ and the leaking sensitive variable $Y(k^*)$, consider the following information-theoretic communication channel between these two variables described in Fig. 4. This communication channel is simply a theoretical construction that helps explain the link between $Y(k)$ and $Y(k^*)$, which are both binary and equiprobable random variables taking their values in $\{\pm 1\}$. The parameters p and p' are the transition probabilities defined as $p = \mathbb{P}(Y(k^*) = +1|Y(k) = -1)$ and $p' = \mathbb{P}(Y(k^*) = -1|Y(k) = +1)$.

Lemma 1. *The communication channel defined in Fig. 4 is a binary symmetric channel (BSC) with transition probability equal to the confusion coefficient $\kappa(k)$.*

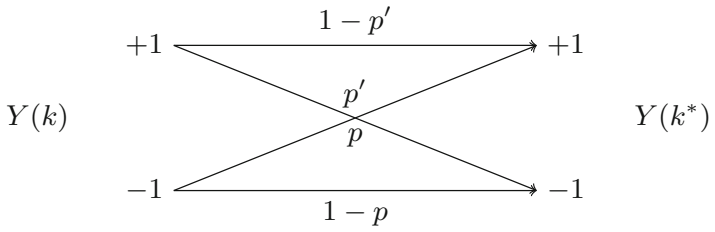

Fig. 4. Abstract communication channel between $Y(k)$ and $Y(k^*)$

Proof. To prove that the channel is symmetric, we show that both transition probabilities coincide: $p = p'$. In fact, from Fig. 4, $\frac{1}{2} = \mathbb{P}(Y(k^*) = 1) = p\mathbb{P}(Y(k) = -1) + (1-p')\mathbb{P}(Y(k) = 1) = \frac{1}{2}(p + 1 - p')$ hence $p = p'$. Now the confusion coefficient $\kappa(k) = \mathbb{P}(Y(k) \neq Y(k^*))$ can be expanded as

$$\begin{aligned}\kappa(k) &= \tfrac{1}{2}\big(\mathbb{P}(Y(k) \neq Y(k^*)|Y(k) = 1) + \mathbb{P}(Y(k) \neq Y(k^*)|Y(k) = -1)\big) && (18)\\ &= \tfrac{1}{2}\big(\mathbb{P}(Y(k^*) = -1|Y(k) = 1) + \mathbb{P}(Y(k^*) = 1|Y(k) = -1)\big) && (19)\\ &= \tfrac{1}{2}\big(p + p'\big) = p = p'. && (20)\end{aligned}$$

This proves that the BSC has transition probability equal to $\kappa(k)$. □

According to a well-known information theoretic result [5, p. 187], the Shannon's *capacity* in bits per bit of this channel is

$$\mathcal{C} = 1 - H_2(\kappa(k)), \tag{21}$$

where $H_2(x)$ is the binary entropy function defined by

$$H_2(x) = x\log_2\Big(\frac{1}{x}\Big) + (1-x)\log_2\Big(\frac{1}{1-x}\Big). \tag{22}$$

This is represented in Fig. 5 as a function of $\kappa(k)$. Interestingly, the value $\kappa(k) = 1/2$ corresponds to null capacity while the capacity is evidently 1 bit per bit for $\kappa(k^*) = 0$, since in this case the above communication channel reduces to the identity.

3.2 A General Result

We can now explain why all distinguishers for monobit leakages depend only on the two parameters $\kappa(k)$ and $\mathsf{SNR} = \sigma^{-2}$.

Theorem 1. *Any theoretical distinguisher $\mathcal{D}(k)$ for a binary leakage y can be expressed as a function of $\kappa(k)$ and σ.*

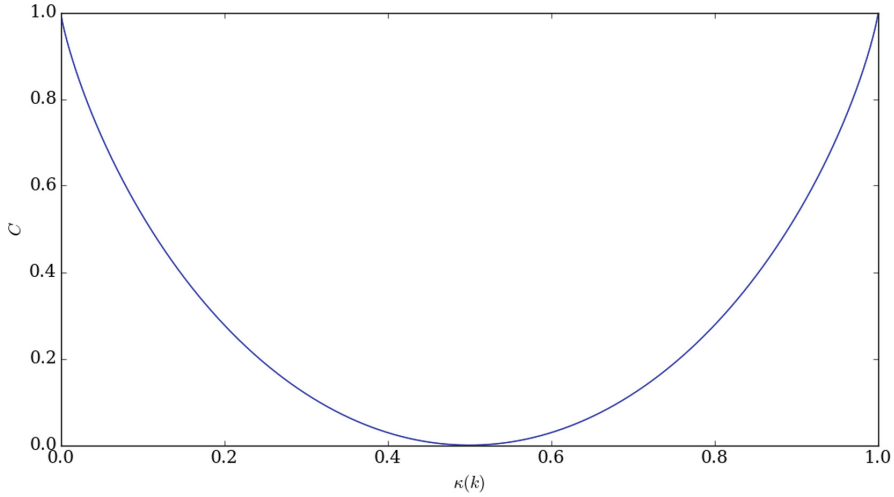

Fig. 5. Representation of the channel capacity according to $\kappa(k)$

Proof. Any theoretical distinguisher is defined in terms of the joint probability distribution of X and $Y(k)$, noted $p(x, y(k))$. Now for any $x \in \mathbb{R}$ and $y(k) = \pm 1$,

$$p(x, y(k)) = \mathbb{P}(y(k))\ p(x \mid y(k)) \tag{23}$$

$$= \frac{1}{2} p(y(k^*) + n \mid y(k)) \tag{24}$$

$$= \frac{1}{2} \sum_{y(k^*)} p(y(k^*) + n \mid y(k), y(k^*))\ \mathbb{P}(y(k^*) \mid y(k)) \tag{25}$$

where $\mathbb{P}(y(k^*) \mid y(k))$ is the transition probability of the channel defined in Fig. 4. There are two possibilities. Either $y(k) = y(k^*)$, and in this case $\mathbb{P}(y(k^*)|y(k)) = 1 - \kappa(k)$, or $y(k) \neq y(k^*)$ and in this case $\mathbb{P}(y(k^*)|y(k)) = \kappa(k)$. The sum over $y(k^*)$ has two terms and both cases are represented. Moreover, the Gaussian noise is independent from every other random variable. Therefore, we have two possibilities for the joint probability:

$$p(x, y(k)) = \begin{cases} \frac{1}{2}\left(\phi(\frac{1+n}{\sigma})\kappa(k) + \phi(\frac{-1+n}{\sigma})(1-\kappa(k))\right) \\ \frac{1}{2}\left(\phi(\frac{-1+n}{\sigma})\kappa(k) + \phi(\frac{1+n}{\sigma})(1-\kappa(k))\right) \end{cases} \tag{26}$$

where $\phi(x)$ is the probability density function of a standard normal random variable. As the noise is centered and Gaussian, the only parameter that characterizes ϕ is its standard deviation σ. Therefore, a joint distribution of a monobit leakage is fully characterized by σ and $\kappa(k)$. □

This proves that the knowledge of the confusion coefficient and the noise power are essential to predict the performances of the side-channel attacks for monobit leakages.

3.3 Classical Distinguishers as Functions of $\kappa(k)$ and σ^2

To highlight the result of Sect. 3.2, we compute the classical distinguishers according to the confusion coefficient and the noise power. As we mentioned in the introduction, some of them have already been expressed according to these variables: we recall these results in Table 1 with references to the articles where the expression of the distinguisher in terms of $\kappa(k)$ is proven.

Table 1. Summary of classical distinguishers. Among all the classical theoretical distinguishers, we notice that the expression of the theoretical value of DoM with $\kappa(k)$ does not depend on σ.

Distinguisher	Original paper	Theoretical expression with $\kappa(k)$	Reference
DoM	[13]	$\mathcal{D}(k) = 2(1/2 - \kappa(k))$	[15]
CPA	[3]	$\mathcal{D}(k) = 2\frac{\lvert 1/2 - \kappa(k)\rvert}{\sqrt{1+\sigma^2}}$	[15]
Euclidean distance	[12, §3]	Lemma 2	This paper
KSA	[22]	$\mathcal{D}(k) = \mathrm{erf}\left(\frac{1}{2\sigma^2}\right) \lvert 1/2 - \kappa(k)\rvert$	[11]
MIA	[9]	Lemma 3	This paper

The new results are given by the following lemmas.

Lemma 2. *For monobit leakages, the Euclidean distance distinguisher can be expressed as:*

$$\mathcal{D}(k) = 4(1/2 - \kappa(k)) - (\sigma^2 + 2). \tag{27}$$

Proof. We have $\mathcal{D}(k) = -\mathbb{E}\big[(X - Y(k))^2\big] = -\mathbb{E}\big[(Y(k^*) - Y(k) + N)^2\big] = -\mathbb{E}\big[(Y(k^*) - Y(k))^2\big] - \sigma^2$ since the noise is independent from $Y(k^*) - Y(k)$. Then by (6), $\mathcal{D}(k) = -4\kappa(k) - \sigma^2 = 4(1/2 - \kappa(k)) - 2 - \sigma^2$ where we have stressed the dependence in $1/2 - \kappa(k)$ as in Table 1. □

Lemma 3. *For monobit leakages, when $\kappa(k) \approx 1/2$ for $k \neq k^*$, the MIA distinguisher can be expressed at first order as:*

$$\mathcal{D}(k) = 2\log_2(e)(\kappa(k) - 1/2)^2 g(\sigma) \tag{28}$$

where

$$g(\sigma) = \frac{1}{2}\mathbb{E}\left[\tanh^2\left(\frac{Z}{\sigma} + \frac{1}{\sigma^2}\right) + \tanh^2\left(\frac{Z}{\sigma} - \frac{1}{\sigma^2}\right)\right] \tag{29}$$

and $Z \sim \mathcal{N}(0,1)$. The function g satisfies

$$\lim_{\sigma \to 0} g(\sigma) = 1 \qquad \textit{and} \qquad \lim_{\sigma \to \infty} \sigma^2 \times g(\sigma) = 1. \tag{30}$$

Proof. See Appendix A. □

Figure 6 plots the shape of $g(\sigma)$ which tends to 1 when $\sigma \to 0$ and is equivalent to $\frac{1}{\sigma^2}$ when $\sigma \to \infty$.

When $k = k^*$ the MIA distinguisher also has a simple expression since it reduces to the known expression of the channel capacity for channels with binary input and additive Gaussian noise [2, p. 274]:

$$\mathcal{D}(k^*) = \frac{1}{\sigma^2} - \int_{\mathbb{R}} \frac{e^{-\frac{1}{2}y^2}}{2\pi} \log_2 \cosh(\frac{1}{\sigma^2} - \frac{y}{\sigma^2}) \mathrm{d}y. \tag{31}$$

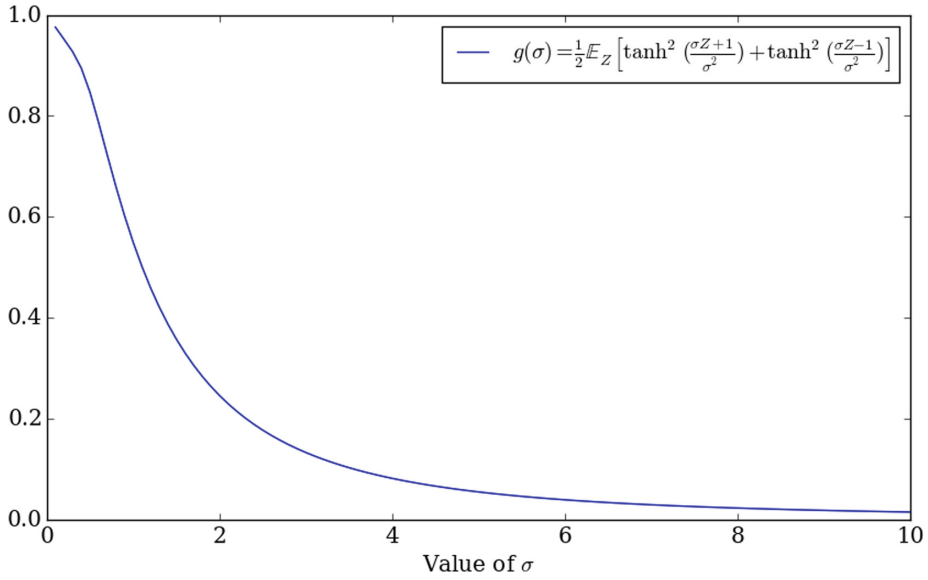

Fig. 6. Representation of $g(\sigma)$

Remark 1. With respect to their theoretical distinguishers, DoM is in bijection with the Euclidean distance, and CPA is in bijection with KSA. Indeed, the Euclidean distance is $\mathcal{D}(k) = 4(1/2 - \kappa(k)) - 2 - \sigma^2$ and σ is independent from the choice of the key. Therefore, there is a bijection between $4(1/2 - \kappa(k)) - 2 - \sigma^2$ and $2(1/2 - \kappa(k))$ which is the theoretical value of DoM. Regarding CPA and KSA, both distinguishers are functions of $|1/2 - \kappa(k)|$.

We also notice that MIA is in bijection with CPA (and therefore KSA). Indeed, according to the value of MIA with $\kappa(k)$, the distinguisher is a function of $(1/2 - \kappa(k))^2$ which is in bijection with $|1/2 - \kappa(k)| = \sqrt{(1/2 - \kappa(k))^2}$. This means that for monobit leakages, any attack that works with one of these distinguishers will also work with another, and *vice versa*.

4 Comparing Distinguishers with the Success Exponent

In the previous section, we have computed the theoretical values of the classical distinguishers in terms of $\kappa(k)$ and σ. Now, we wish to compare their success rate. As we mentioned Sect. 2.3, the attacker computes the estimated distinguisher

$\widehat{\mathcal{D}}(k)$ to recover the secret key. This is the main reason why all distinguishers do not perform equally in key recovery; indeed, they do not converge at the same speed towards their theoretical value.

In order to compare them, we have computed their *success exponent*, a metric proposed by Guilley *et al.* in [10] that evaluates how fast the success rate of a distinguisher converges to 100%. With a Gaussian assumption, they prove that the success rate can be modeled as

$$\mathsf{SR} = 1 - \exp(-q \times \mathsf{SE}), \tag{32}$$

where q is the number of traces and $\mathsf{SE} \in \mathbb{R}^+$ is the so-called success exponent. Therefore, the greater the success exponent is, the faster the convergence of the success rate.

Table 2. Success exponents for the classical distinguishers. The numerical values of SE are obtained for AES SubBytes least significant bit leakage model and noise of standard deviation $\sigma = 4$. Notice that in the monobit case, Euclidean distance and DoM have strictly the same success rate because $-(X - Y(k))^2 = -X^2 + 2XY(k) - 1$, and X^2 is independent of the choice of the key.

Distinguisher	Closed form SE with $\kappa(k)$ and σ	Reference	Numerical value for AES SubBytes
DoM	$\frac{1}{2}\min_{k\neq k^*} \frac{\kappa(k)}{1+\sigma^2-\kappa(k)}$	[10, Proposition 4]	3.39×10^{-3}
CPA	Lemma 4	This paper	3.39×10^{-3}
Euclidean distance	$\frac{1}{2}\min_{k\neq k^*} \frac{\kappa(k)}{1+\sigma^2-\kappa(k)}$	[10, Proposition 5]	3.39×10^{-3}
KSA	Lemma 5	This paper	1.08×10^{-3}
MIA	Lemma 6	[10, Proposition 6]	8.52×10^{-5}

We present the theoretical values of the success exponent for the different distinguishers in Table 2. As a direct consequence of Theorem 1, all of these success exponents are function of $\kappa(k)$ and σ. Therefore, if the attacker only knows the type of substitution box that is used and the SNR of the leakage, he can predict how fast he recovers the secret key.

Lemma 4 (Success exponent of CPA). *The success exponent of CPA*[2] *is:*

$$\mathsf{SE} = \frac{1}{2}\min_{k\neq k^*} \frac{1 - 2|1/2 - \kappa(k)|}{1 + 2\sigma^2 + 2|1/2 - \kappa(k)|}. \tag{33}$$

Proof. See Appendix B. □

[2] In [10], CPA is treated as a distinguisher, but without the absolute values. Those remove false positives which occur in monobit leakages when there are anti-correlations. Our value of the success exponent is, therefore, different from theirs.

Lemma 5 (Success exponent of KSA). *Assuming that the distributions are estimated with the kernel method using Heaviside step function, the success exponent of KSA is*

$$\mathsf{SE} = \frac{1}{2} \min_{k \neq k^*} \frac{\operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}\right)^2 (1/2 - |1/2 - \kappa(k)|)}{2 - \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}\right)^2 (1/2 - |1/2 - \kappa(k)|)}. \tag{34}$$

Proof. See Appendix C. □

Lemma 6 (Success exponent of MIA). *When $\sigma \gg 1$, the success exponent for an MIA computed with histograms is*

$$\mathsf{SE} = \frac{4 \log_2(e)^2}{\sigma^4} \min_{k \neq k^*} \kappa(k)^2 (1 - \kappa(k))^2. \tag{35}$$

Proof. See Appendix D. □

In order to validate our theoretical results, we have simulated attacks within the monobit model presented in Sect. 2. The success rates of these attacks are presented in Fig. 7. In this figure, we notice that, as expected, the Euclidean distance (ML) is the best distinguisher, closely followed by CPA. Both have similar same success rate. The small difference is due to the use the the absolute values in the distinguishing function of CPA (see discussion in Remark 9 of [12]). The KSA is requiring a bit less than the double of traces, compared to Euclidean distance, DoM and CPA. The MIA performs really bad compared to the other distinguishers. Error bars represent the inaccuracy while estimating the SR (here, we ran 100 simulations).

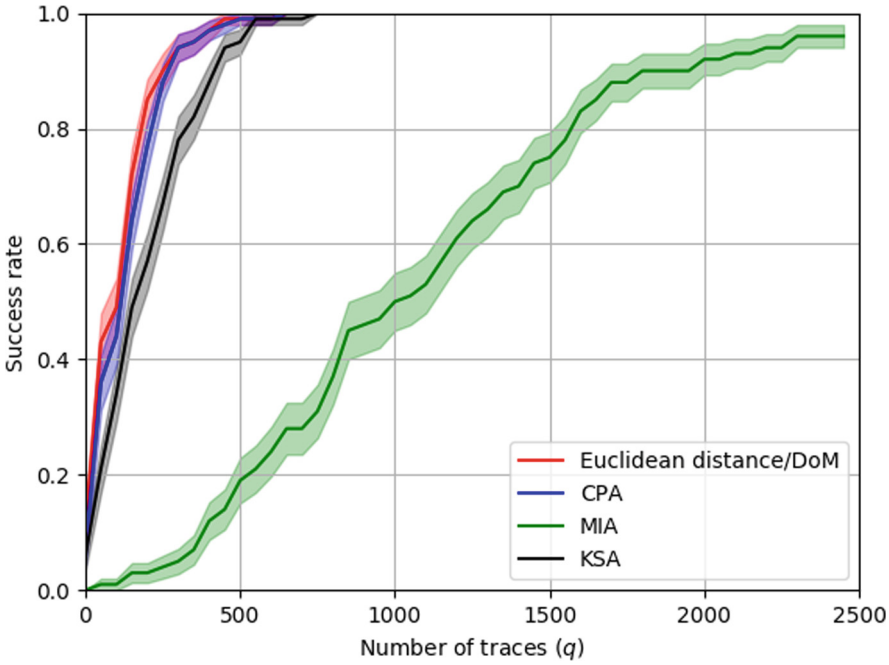

Fig. 7. Success rate for classical distinguishers ($\sigma = 4$)

These simulations are therefore in complete coherence with the theoretical results of Table 2. Indeed, the order of the distinguishers is the same w.r.t. the success rate and w.r.t. the success exponent. In addition, according to the definition of the success exponent SE in (32), the number of traces q to reach a given success rate (e.g., $\mathsf{SR} = 80\%$) is proportional to the inverse of SE. This quantitative law is satisfied in the simulation of Fig. 7.

5 Conclusion

In this paper, we have mathematically proven that only two parameters, the confusion coefficient and the SNR, determine the side-channel distinguishing efficiency for monobit leakages. Both of them are easy to compute because the confusion coefficient can be calculated with the knowledge of the operating substitution box and the SNR can be measured offline.

Our work is useful to predict how fast a distinguisher will succeed to recover the secret key. Long and painful simulations can be advantageously replaced by the computation of the success exponent using closed-form expressions.

This paper also consolidates the state of the art about the classical distinguishers, especially for MIA and KSA. We have derived the success exponent for these two distinguishers as a function of the confusion coefficient and the standard deviation of the noise.

A Proof of Lemma 3

The MIA distinguisher is expressed as

$$\mathcal{D}(k) = I(Y(k^*) + N; Y(k)) = h(Y(k^*) + N) - h(Y(k^*) + N \mid Y(k)). \tag{36}$$

From Sect. 3.1, $Y(k^*)$ knowing $Y(k)$ is a binary random variable with probability $\kappa(k)$. As N is Gaussian independent from $Y(k)$, the pdf of $Y(k^*) + N$ knowing $Y(k)$ is a Gaussian mixture that can take two forms:

$$p_{\kappa(k)}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}[\kappa(k)e^{-\frac{(x-1)^2}{2\sigma^2}} + (1-\kappa(k))e^{-\frac{(x+1)^2}{2\sigma^2}}] \\ \frac{1}{\sqrt{2\pi}\sigma}[\kappa(k)e^{-\frac{(x+1)^2}{2\sigma^2}} + (1-\kappa(k))e^{-\frac{(x-1)^2}{2\sigma^2}}] \end{cases}, \tag{37}$$

By symmetry, their entropy $h(Y(k^*) + N \mid Y(k))$ will be the same and we can take any of these pdfs. Letting ϕ be the standard normal density, we can write

$$p_{\kappa(k)}(x) = p_{1/2}(x) - 2(1/2 - \kappa(k))\phi(x)e^{-\frac{1}{\sigma^2}} \sinh(\frac{x}{\sigma^2}) \tag{38}$$

$$= p_{1/2}(x)(1 - 2(1/2 - \kappa(k)) \tanh(\frac{x}{2\sigma^2}). \tag{39}$$

where

$$p_{1/2}(x) = \frac{1}{2\sqrt{2\pi}\sigma}[e^{-\frac{(x-1)^2}{2\sigma^2}} + e^{-\frac{(x+1)^2}{2\sigma^2}}] = \frac{1}{\sigma}e^{-\frac{1}{2\sigma^2}}\phi(\frac{x}{\sigma})\cosh(\frac{x}{\sigma^2}). \tag{40}$$

For notational convenience define $\epsilon = 2(1/2 - \kappa(k))$, $p = p_{1/2}(x)$, and $t = \tanh(x)$. Then

$$I(X;Y(k)) = h(Y(k^*) + N) - h(Y(k^*) + N \mid Y(k)) \tag{41}$$

$$= -\int p \log_2 p + \int (p(1-\epsilon t)) \log_2(p(1-\epsilon t)) \tag{42}$$

$$= -\int \epsilon p t \log_2 p + \int p \log_2(1-\epsilon t) - \int p\epsilon t \log_2(1-\epsilon t). \tag{43}$$

The first term vanishes since p is even and t odd. We apply a Taylor expansion:

$$I(X;Y(k)) = \int p[-\epsilon t - \frac{\epsilon^2 t^2}{2} - \frac{\epsilon^3 t^3}{3} + O(\epsilon^4)] - \int \epsilon p t[-\epsilon t - \frac{\epsilon^2 t^2}{2} - \frac{\epsilon^3 t^3}{3} + O(\epsilon^4)]. \tag{44}$$

The odd terms of the expansion are null as t is odd and p even. We therefore obtain:

$$I(X;Y(k)) = \int p[-\frac{\epsilon^2 t^2}{2} + O(\epsilon^4)] - \int [-\epsilon^2 p t^2 + O(\epsilon^4)] = \int \frac{\epsilon^2 p t^2}{2} + O(\epsilon^4). \tag{45}$$

Thus, finally,

$$\mathcal{D}(k) = 2\log_2(e)(1/2 - \kappa(k))^2 g(\sigma), \tag{46}$$

where

$$g(\sigma) = \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}} \int_{\mathbb{R}} \phi(\frac{x}{\sigma}) \cosh(\frac{x}{\sigma^2}) \tanh^2(\frac{x}{\sigma^2}) \mathrm{d}x. \tag{47}$$

There are several ways to express $g(\sigma)$. For example, we have:

$$g(\sigma) = e^{-\frac{1}{2\sigma^2}} \int_{\mathbb{R}} \phi(x) \cosh(\frac{x}{\sigma}) \tanh^2(\frac{x}{\sigma}) \mathrm{d}x. \tag{48}$$

This expression can be reduced to:

$$g(\sigma) = \frac{1}{2} \mathbb{E}_X \left[\tanh^2(\frac{X}{\sigma} + \frac{1}{\sigma^2}) + \tanh^2(\frac{X}{\sigma} - \frac{1}{\sigma^2}) \right], \tag{49}$$

where $X \sim \mathcal{N}(0,1)$. By the dominated convergence theorem ($\tanh^2(\frac{X}{\sigma} + \frac{1}{\sigma^2})$ is always smaller than 1) when $\sigma \to 0$, we obtain $g(0) = 1$ and when $\sigma \to \infty$ we obtain the equivalent $\frac{1}{\sigma^2}$.

B Proof of Lemma 4

The success exponent is defined by

$$\mathsf{SE} = \frac{\mathbb{E}[\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)]^2}{2\mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k))}. \tag{50}$$

where in our case

$$\widehat{\mathcal{D}}(k) = \frac{1}{q\sqrt{1+\sigma^2}}\Big|\sum_{i=1}^{q} X_i Y_i(k)\Big|. \tag{51}$$

First for large q we can consider that $\mathbb{E}[|\sum_i X_i Y_i(k)|] = |\mathbb{E}[\sum_i X_i Y_i(k)]|$.

$$\mathbb{E}[\widehat{\mathcal{D}}(k)] = |\mathbb{E}[XY(k)]| = \frac{2 \times |1/2 - \kappa(k)|}{\sqrt{1+\sigma^2}} \tag{52}$$

hence

$$\mathbb{E}[\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)] = \frac{1 - 2 \times |1/2 - \kappa(k)|}{\sqrt{1+\sigma^2}}. \tag{53}$$

Secondly we have

$$\mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)) = \frac{1}{q^2(1+\sigma^2)}\mathrm{Var}\Big(\Big|\sum_{i=1}^{q} X_i Y_i(k^*)\Big| - \Big|\sum_{i=1}^{q} X_i Y_i(k)\Big|\Big). \tag{54}$$

To remove the absolute values, we distinguish two cases whether the sum is positive or negative. We consider that q is large enough to have strictly positive or negative values.

$$\begin{aligned}
\mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)) &= \frac{1}{q^2(1+\sigma^2)}\mathrm{Var}\Big(\sum_{i=1}^{q} X_i Y_i(k^*) \mp \sum_{i=1}^{q} X_i Y_i(k)\Big) && (55)\\
&= \frac{1}{q^2(1+\sigma^2)}\mathrm{Var}\Big(\sum_{i=1}^{q} X_i\big(Y_i(k^*) \mp Y_i(k)\big)\Big) && (56)\\
&= \frac{1}{q(1+\sigma^2)}\mathrm{Var}\big(X\big(Y(k^*) \mp Y(k)\big)\big) && (57)\\
&= \frac{1}{q(1+\sigma^2)}\mathrm{Var}\big((Y(k^*) + N\big)\big(Y(k^*) \mp Y(k)\big)\big) && (58)\\
&= \frac{1}{q(1+\sigma^2)}\mathrm{Var}\big(\mp Y(k^*)Y(k) + N(Y(k^*) \mp Y(k))\big). && (59)
\end{aligned}$$

The variance term is the difference of the two following quantities

$$\begin{aligned}
\mathbb{E}\Big[(\mp Y(k^*)Y(k) + N(Y(k^*) \mp Y(k)))^2\Big] &= 1 + 2\sigma^2(1 - 2|1/2 - \kappa(k)|) && (60)\\
\mathbb{E}\Big[\mp Y(k^*)Y(k) + N(Y(k^*) \mp Y(k))\Big]^2 &= \Big(2(1/2 - \kappa(k))\Big)^2. && (61)
\end{aligned}$$

Combining all the above expressions we obtain (33).

C Proof of Lemma 5

To prove the success rate of KSA, we first need an estimator for the cumulative density function. We take as kernel a function Φ as simple as possible i.e. the Heaviside function $\Phi(x) = 0$ if $x < 0$ and $\Phi(x) = 1$ if $x \geq 0$.

With this function and for $x \in \mathbb{R}$, we can estimate $F(x|Y(k)=1) - F(x)$ by the following estimator:

$$\tilde{F}(x|Y(k)=1) - \tilde{F}(x) = \frac{\sum_{i|Y_i(k)=1} \Phi(x - X_i)}{\sum_{i|Y_i(k)=1} 1} - \frac{\sum_i \Phi(x - X_i)}{q}. \tag{62}$$

We suppose that q is large enough to consider that $\sum_{i|Y_i(k)=1} 1 = \frac{q}{2}$ (by the law of large numbers). Therefore we have:

$$\tilde{F}(x|Y(k)=1) - \tilde{F}(x) = \frac{\sum_{i|Y_i(k)=1} \Phi(x - X_i)}{q} - 2\frac{\sum_i \Phi(x - X_i)}{q}. \tag{63}$$

We notice that $\sum_{i|Y_i(k)=1} \Phi(x - X_i) = \frac{1}{2}\sum_i (Y_i(k)+1)\Phi(x - X_i)$. Therefore

$$\tilde{F}(x|Y(k)=1) - \tilde{F}(x) = \frac{1}{q}\sum_{i=1}^{q} Y_i(k)\Phi(x - X_i). \tag{64}$$

This estimator is a sum of i.i.d. random variables. We can therefore apply the central limit theorem.

$$\begin{aligned}
\mathbb{E}[\tilde{F}(x|Y(k)=1) - \tilde{F}(x)] &= \mathbb{E}[Y(k)\Phi(x - X_i)] &(65)\\
&= \mathbb{E}[Y(k)\Phi(x - Y(k^*) - N)] &(66)\\
&= \frac{1}{2}(\kappa(k) - 0.5)\Big(\mathrm{erf}\Big(\frac{1-x}{\sigma\sqrt{2}}\Big) + \mathrm{erf}\Big(\frac{1+x}{\sigma\sqrt{2}}\Big)\Big). &(67)
\end{aligned}$$

The maximum of the absolute value is for $x = 0$ and we obtain:

$$\|\mathbb{E}[\tilde{F}(x|Y(k)=1) - \tilde{F}(x)]\|_\infty = |0.5 - \kappa(k)|\mathrm{erf}\Big(\frac{1}{\sigma\sqrt{2}}\Big). \tag{68}$$

We notice that $\|\mathbb{E}[\tilde{F}(x|Y(k)=1) - \tilde{F}(x)]\|_\infty = \|\mathbb{E}[\tilde{F}(x|Y(k)=-1) - \tilde{F}(x)]\|_\infty$. To calculate the variance, we consider that $x = 0$ as it is the value that maximizes the expectation of the distinguisher.

$$\mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)) = \mathrm{Var}\Big(\frac{1}{q}\Big(\sum_{i=1}^{q} \Phi(x - X_i)(Y_i(k^*) - Y_i(k))\Big)\Big) \tag{69}$$

The computation of this variance gives:

$$\mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k)) = 2(0.5 - |0.5 - \kappa(k)|) - \mathrm{erf}\Big(\frac{1}{\sigma\sqrt{2}}\Big)^2 (0.5 - |0.5 - \kappa(k)|)^2. \tag{70}$$

Overall, the success exponent is:

$$\mathsf{SE} = \frac{1}{2}\min_{k \neq k^*} \frac{\mathrm{erf}\big(\frac{1}{\sqrt{2}\sigma}\big)^2 (1/2 - |1/2 - \kappa(k)|)}{2 - \mathrm{erf}\big(\frac{1}{\sqrt{2}\sigma}\big)^2 (1/2 - |1/2 - \kappa(k)|)}. \tag{71}$$

D Proof of Lemma 6

For MIA, we refer to [10, Section 5.3] for the theoretical justifications. In order to obtain a simple closed-form expression of the success exponent, we suppose that $\sigma \gg 1$ and that the probability density functions are all Gaussian. This means that $X|Y(k)$ is a Gaussian random variable of standard deviation $\sqrt{4\kappa(k)(1-\kappa(k))+\sigma^2}$. Moreover, we will keep only the first order approximation in $\mathsf{SNR} = \sigma^{-2}$ of the SE.

$$h(X|Y(k)) - h(X|Y(k^*)) = \frac{1}{2}\log_2(2\pi e \cdot (4\kappa(k)(1-\kappa(k))+\sigma^2) - \frac{1}{2}\log_2(2\pi e \cdot \sigma^2) \tag{72}$$

$$= \frac{1}{2}\log_2 \frac{4\kappa(k)(1-\kappa(k))+\sigma^2}{\sigma^2} \tag{73}$$

$$\approx \frac{\log_2(e) 4\kappa(k)(1-\kappa(k))}{2\sigma^2} \tag{74}$$

The Fisher information of a Gaussian random variable of standard deviation ζ is equal to $\frac{1}{\zeta^2}$. Therefore the Fisher information of X knowing $Y = y(k)$ is:

$$F(X|Y(k) = y(k)) = \frac{1}{4\kappa(k)(1-\kappa(k))+\sigma^2}. \tag{75}$$

As this value does not depend on the value of $Y(k)$, we have:

$$F(X|Y(k)) = \frac{1}{4\kappa(k)(1-\kappa(k))+\sigma^2} \tag{76}$$

$$J(X|Y(k)) - J(X|Y(k^*)) = \frac{1}{4\kappa(k)(1-\kappa(k))+\sigma^2} - \frac{1}{\sigma^2} \tag{77}$$

$$\approx -\frac{\kappa(k)(1-\kappa(k))}{\sigma^4}. \tag{78}$$

Last, we have to calculate $\mathrm{Var}(-\log_2 p(X|Y(k) = y(k)))$. Let $\zeta^2 = \sigma^2 + 4\kappa(k)(1-\kappa(k))$ and C the normalization constant. We have:

$$\mathrm{Var}(-\log_2 p(X|Y(k) = y(k))) = \mathrm{Var}\Big(-\log_2\Big(C\exp\Big(-1/2\frac{(X-\mu)^2}{\zeta^2}\Big)\Big)\Big) \tag{79}$$

$$= \mathrm{Var}\Big(-\log_2(C) + 1/2\frac{(X-\mu)^2}{\zeta^2}\Big) \tag{80}$$

$$= \frac{1}{4}\mathrm{Var}\Big(\frac{(X-\mu)^2}{\zeta^2}\Big) = \frac{1}{4\zeta^4}\mathrm{Var}(X^2) \tag{81}$$

$$= \frac{1}{4(\sigma^2+4\kappa(k)(1-\kappa(k)))^2} 2(1+\sigma^2)^2 \approx \frac{1}{2}. \tag{82}$$

Overall, the success exponent defined in [10, Proposition 6] can be simplified in the case of monobit leakage as:

$$\mathsf{SE} \approx \min_{k \neq k^*} 4\frac{\log_2(e)^2 \kappa(k)^2 (1-\kappa(k))^2}{\sigma^4}. \tag{83}$$

References

1. Batina, L., Robshaw, M. (eds.): CHES 2014. LNCS, vol. 8731. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44709-3
2. Blahut, R.E.: Principles and Practice of Information Theory. Addison-Wesley Longman Publishing Co. Inc., Boston (1987)
3. Brier, É., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28632-5_2
4. Carlet, C., Heuser, A., Picek, S.: Trade-offs for S-boxes: cryptographic properties and side-channel resilience. In: Gollmann, D., Miyaji, A., Kikuchi, H. (eds.) ACNS 2017. LNCS, vol. 10355, pp. 393–414. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61204-1_20
5. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley-Interscience, New York (2006). ISBN-10: 0471241954, ISBN-13: 978-0471241959
6. Daemen, J., Rijmen, V.: Rijndael for AES. In: AES Candidate Conference, pp. 343–348 (2000)
7. Fei, Y., Ding, A.A., Lao, J., Zhang, L.: A statistics-based success rate model for DPA and CPA. J. Cryptographic Eng. **5**(4), 227–243 (2015). https://doi.org/10.1007/s13389-015-0107-0
8. Fei, Y., Luo, Q., Ding, A.A.: A statistical model for DPA with novel algorithmic confusion analysis. In: Prouff, E., Schaumont, P. (eds.) CHES 2012. LNCS, vol. 7428, pp. 233–250. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33027-8_14
9. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual information analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85053-3_27
10. Guilley, S., Heuser, A., Rioul, O.: A key to success. In: Biryukov, A., Goyal, V. (eds.) INDOCRYPT 2015. LNCS, vol. 9462, pp. 270–290. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26617-6_15
11. Heuser, A., Rioul, O., Guilley, S.: A theoretical study of Kolmogorov-Smirnov distinguishers – side-channel analysis vs. differential cryptanalysis. In: Prouff [17], pp. 9–28. https://doi.org/10.1007/978-3-319-10175-0_2
12. Heuser, A., Rioul, O., Guilley, S.: Good is not good enough - deriving optimal distinguishers from communication theory. In: Batina and Robshaw [1], pp. 55–74. https://doi.org/10.1007/978-3-662-44709-3_4
13. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48405-1_25
14. Lomné, V., Prouff, E., Rivain, M., Roche, T., Thillard, A.: How to estimate the success rate of higher-order side-channel attacks. In: Batina and Robshaw [1], pp. 35–54. https://doi.org/10.1007/978-3-662-44709-3_3
15. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks. Revealing the Secrets of Smart Cards. Springer, Boston (2007). https://doi.org/10.1007/978-0-387-38162-6
16. Mangard, S., Oswald, E., Standaert, F.: One for all - all for one: unifying standard differential power analysis attacks. IET Inf. Secur. **5**(2), 100–110 (2011). https://doi.org/10.1049/iet-ifs.2010.0096
17. Prouff, E. (ed.): COSADE 2014. LNCS, vol. 8622. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10175-0

18. Reparaz, O., Gierlichs, B., Verbauwhede, I.: A note on the use of margins to compare distinguishers. In: Prouff [17], pp. 1–8. https://doi.org/10.1007/978-3-319-10175-0_1
19. Rivain, M.: On the exact success rate of side channel analysis in the Gaussian model. In: Avanzi, R.M., Keliher, L., Sica, F. (eds.) SAC 2008. LNCS, vol. 5381, pp. 165–183. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04159-4_11
20. Schindler, W., Lemke, K., Paar, C.: A stochastic model for differential side channel cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005). https://doi.org/10.1007/11545262_3
21. Whitnall, C., Oswald, E.: A fair evaluation framework for comparing side-channel distinguishers. J. Cryptographic Eng. **1**(2), 145–160 (2011)
22. Whitnall, C., Oswald, E., Mather, L.: An exploration of the Kolmogorov-Smirnov test as a competitor to mutual information analysis. In: Prouff, E. (ed.) CARDIS 2011. LNCS, vol. 7079, pp. 234–251. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-27257-8_15