Noise reduction in side channel attack using fourth-order cumulants

Tanh Ha Le, Jessy Clédière, Christine Serviere, Jean-Louis Lacoume

Noise Reduction in Side Channel Attack Using Fourth-Order Cumulant

Thanh-Ha Le, Jessy Clédière, Christine Servière, and Jean-Louis Lacoume, *Senior Member, IEEE*

***Abstract*—Side channel attacks exploit physical information leaked during the operation of a cryptographic device (e.g., a smart card). The confidential data, which can be leaked from side channels, are timing of operations, power consumption, and electromagnetic emanation. In this paper, we propose a preprocessing method based on the fourth-order cumulant, which aims to improve the performance of side channel attacks. It takes advantages of the Gaussian and nonGaussian properties, that respectively characterize the noise and the signal, to remove the effects due to Gaussian noise coupled into side channel signals. The proposed method is then applied to analyze the electromagnetic signals of a synthesized application-specific integrated circuit during a data encryption standard operation. The theoretical and experimental results show that our method significantly reduces the number of side channel signals needed to detect the encryption key.**

***Index Terms*—Correlation power analysis (CPA), data encryption standard (DES), differential power analysis (DPA), fourth-order cumulant, Gaussian noise, higher order statistics, side channel attack.**

I. Introduction

SIDE channel analysis was first introduced in the form of timing attacks by Kocher in 1996 [1]. Some years later, Kocher *et al.* proposed another attack based on power consumption information, known as differential power analysis (DPA) [2]. Power consumption signals of complementary metal–oxide semiconductor (CMOS) chips were used to deduce the key of the DES algorithm [3] by the difference of mean curves selected on defined criteria. Later, electromagnetic emanation signals obtained by different kinds of sensors were successfully used to replace power consumption signals [4]–[6]. This kind of attack is known as differential electromagnetic analysis (DEMA). The effectiveness of DPA and DEMA has been verified in different types of devices [application-specific integtated circuit (ASIC), field-programmable gate array (FPGA)], implemented with different cryptographic algorithms (DES, AES, RC4, ECC, RSA). Several countermeasures have been proposed to secure them from first- and high-order differential attacks [7]–[10]. Numerous authors have extended Kocher's *et al.* point of view by introducing multibit DPA methods to improve the differential attack [11]–[14]. Recently, the new technique of correlation power analysis (CPA) has been investigated alot [28], [30], [31]. For the sake of simplicity, we will use the terms "DPA" and "CPA" in this paper for any differential or correlation analysis on power or electromagnetic signals.

Since the detection of an encryption key is mainly based on side channel signals, their signal-to-noise ratio (SNR) may significantly influence the key guess accuracy. If the undesirable noise level is extremely high, the secret key can be undetectable. Therefore, adding noise to side channel signals is one of the countermeasures against side channel analysis. The averaging operation can be used to reduce noise as in [2] and [15]. However, this method requires many power consumption signals. Messerges *et al.* introduced another method which consists of filtering noise and using the multibit DPA attack to improve the SNR of DPA signals [12]. Contrary to many approaches which try to eliminate noise, the template attack technique [15] is based on a precise noise model to collect the maximum information from a single signal. Template attacks were then developed in [16]–[18].

Our work is directed toward filling the gap between signal processing and what has been previously proposed. The idea of improving the detection of transient signals embedded in additive Gaussian noise using higher order statistics was investigated in [19] and [20]. Transient and impulsive signals have super-Gaussian probability densities and, thus, high values of kurtosis. As a result, by using the fourth-order cumulant of the observations, the effects due to Gaussian noise can be removed and the dynamics of the signal can be enhanced.

Our contribution focuses on exploiting the fourth-order cumulant properties as a preprocessing phase before the standard DPA/CPA methods. We calculate the probability of detection and the SNR which represent the capacity of the secret key detection. We show theoretically and experimentally that our method supports the reduction of the number of signals needed to detect the encryption key.

This paper is structured as follows. The background of side channels attacks and higher order statistics is presented in Section II and Section III. In Section IV, we provide a detailed explanation of the proposed method. Section V describes the theoretical analysis of our solution which is then experimentally validated in Section VI.

II. Side Channel Attacks

A. Information Leaked From Side Channel Signals

Today, CMOS technology is the most widely used in digital design applications, such as smart cards. Two main side channels which can be leaked in CMOS circuits are the power dissipation and the electromagnetic emanation.

T.-H. Le and J. Clédière are with the CEA Leti, Grenoble 38054, France (e-mail: thanhha.le@cea.fr; jessy.clediere@cea.fr).

C. Servière and J.-L. Lacoume are with the LIS, INPG, Saint Martin d'Heres 38402, France (e-mail: christine.serviere@inpg.fr; jean-louis.lacoume@inpg.fr).

1) Power Dissipation: The amount of power dissipated in a CMOS circuit is the sum of static and dynamic dissipation [21]. The static dissipation, which is, in general, very small, is due to leakage current or other currents drawn continuously from the power supply. The dynamic dissipation is due to the switching transient current, the charging, and the discharging of load capacitance. From a side channel attack point of view, the dynamic dissipation contains significant information which can be exploited by attackers.

2) Electromagnetic Emanations: A sudden current pulse in a CMOS circuit causes a sudden variation of the electromagnetic field surrounding the device, which can be captured by inductive sensors. The relation between the magnetic field and its source current is given by Biot–Savart's law $\overrightarrow{dB} = (\mu I \overrightarrow{dl} \overrightarrow{r}/4\pi r^3)$, where $\overrightarrow{dl}$ is an infinitesimal length of the conductor carrying electric current I, μ is the magnetic permeability, and $\overrightarrow{r}$ is the directional vector representing the distance between the current and the field point. According to Faraday's law, any change in the magnetic environment of a coil of wire will cause a voltage to be induced in the coil $V = (d\phi/dt)$, where the magnetic flux is $\phi = \int_S \overrightarrow{B}\,\overrightarrow{dS}$. Hence, if useful information is contained in I, it can also be detected by measuring V. The advantage of electromagnetic signals compared to power consumption signals is the possibility of measuring V without direct device access. Furthermore, for each message, several electromagnetic signals can be captured by placing sensors in different positions [5] to obtain more localized information.

B. Side Channel Noise

1) Gaussian Noise: When performing a power analysis of smart cards, the following kinds of noise should be taken into account.

- Intrinsic noise is due to physical fluctuations in circuits. Such noise can be distinguished into at least four different types: thermal noise, shot noise, $1/f$ noise, and generation–recombination noise [22].
- Added noise is due to voluntary physical fluctuations in circuits. It can be added by using a linear feedback shift register (LFSR) or random generators, which allow chip developers to partially block side channel attacks.
- Quantization noise is caused by analog-to-digital conversion and is assumed to be an uncorrelated stationary white noise source [23]. Numerical noise can also be generated during DPA/CPA computation.
- External noise is generated by external sources, such as measuring equipment or environment conditions.

In practice, all fluctuating currents and voltages generated in electrical devices have a probability density function of Gaussian form [22] since the fluctuating quantity is the sum of a large number of independent random variables. In such a case, the central limit theorem holds and, thus, the intrinsic noise is Gaussian. By the same way, we can consider the quantization noise and the external noise as Gaussian noise.

2) Temporal Misalignment: The temporal misalignment of signals provokes a great amount of noise into signals and destabilizes side channel attacks. The misalignment sources in power analysis can be divided into two groups. The first one consists of unintentional sources generated by the device or measurements [24]. The second one includes intentional sources added by device developers, for example, the random process interrupts (RPIs) [25]. Some solutions were proposed in [25]–[27] to solve the temporal misalignment.

C. Differential Power Analysis

DPA exploits the dependence between the handled data and the power consumption of the circuit. The original DPA attack proposed by Kocher *et al.* [2] is based on the fact that the power dissipation to manipulate one bit to 1 is different from the power dissipation to manipulate it to 0. To test different keys K_s, DPA uses N ciphertexts (or plaintexts) C_i $(i = 1 \dots N)$ and a selection function $D(C_i, b, K_s)$ which predicts the value of an examined bit b. DPA computes the differential trace $\Delta_D(b)$ as the difference between the average of the traces for which $D(C_i, b, K_s)$ is 1 and the average of the traces for which $D(C_i, b, K_s)$ is 0. If we denote W_{C_i} as the power consumption signal corresponding to the message C_i, the trace $\Delta_D(b)$ is computed as follows:

$$\Delta_D(b) = \frac{\sum_{i=1}^{N} D(C_i, b, K_s) W_{C_i}}{\sum_{i=1}^{N} D(C_i, b, K_s)} - \frac{\sum_{i=1}^{N} (1 - D(C_i, b, K_s)) W_{C_i}}{\sum_{i=1}^{N} (1 - D(C_i, b, K_s))}. \quad (1)$$

In theory, if the bits inside the algorithm are uniformly distributed and if the choice of b and text messages is suitable, then for the correct hypothesis K_s, the $\Delta_D(b) \neq 0$ at the instant τ when the bit b is handled. It is thus represented by a peak in the differential trace at the instant τ, which is called the DPA peak. For incorrect keys, $\Delta_D(b)$ tends to 0 and no significant peak appears. However, in practice, the bit distribution conditions are never perfect, and some output correlations can occur with incorrect key guess, so we observe other peaks which are not the DPA peak. We define a ghost peak as the one which appears at the instant τ and in a differential curve corresponding to an incorrect key hypothesis. The ghost peak problem was explained in [28] and [29]. We call also a secondary peak as the one which appears at an instant other than τ in a differential curve corresponding to any key hypothesis (wrong or correct). In our experiment, we detect the subkey K_s used in the first S-box of the first round of DES. The size of K_s is 6 b, so we have 64 key assumptions. The bit b is one bit of the S-box output.

D. Correlation Power Analysis

Correlation approaches are based on the relation between the actual power consumption of a circuit and a power consumption model (e.g., the Hamming weight model [30], [31]). In [28], the Hamming distance model was used. The relationship between the power consumption W and the Hamming distance is linear and the correct key is the one which maximizes their correlation factor. If we denote $H_{i,R} = H(R \oplus C_i)$ as the Hamming

distance between the actual state of message C_i and a reference state R, the correlation factor $\hat{\rho}_{WH}(R)$ is formulated as

$$\hat{\rho}_{WH}(R) = \frac{\frac{\sum_{i=1}^{N} W_{C_i} H_{i,R}}{N} - \frac{\sum_{i=1}^{N} W_{C_i}}{N} \frac{\sum_{i=1}^{N} H_{i,R}}{N}}{\sigma_W \sigma_H} \quad (2)$$

where σ_W and σ_H are the standard deviations of W_{C_i} and $H_{i,R}$.

In our evaluation, we examine four bits of an S-box output and $H_{i,R}$ is the Hamming distance between the S-box output and its reference state.

III. Higher Order Statistics

Moments and cumulants are statistical measures which characterize signal properties. The first-order moment (the mean) and the second-order cumulant (the variance) have been widely used to characterize the probability distribution of a signal. If a signal has a Gaussian probability density function, it is sufficient to use the first- and second-order measures to characterize it. However, many real-life signals are nonGaussian and higher order statistics (HOS, moments and cumulants of orders higher than 2) are needed to fully describe them. As for applications, HOS first play an important role in blind array signal processing [32], [33]. The idea of the Gaussian noise suppression using cumulants was investigated in [34]. Another application using cumulants is the retrieval harmonics in noise [35], [36]. Blind source separation also obtains much success using HOS [37], [38].

Consider a 1-D real random variable x which is associated with its first and second characteristic functions. The moments of x can be obtained by deriving the first characteristic function at point 0, whereas the cumulants can be obtained by deriving the second characteristic function at point 0 [39]. The rth-order cumulant is a function of the moments of orders up to r. If the variable x is centered (i.e., $\mu_1(x) = 0$), for the orders from 1 to 4, these relations are

$$\begin{aligned}
\kappa_1(x) &= 0 \\
\kappa_2(x) &= \mu_2(x) = E[x^2] \\
\kappa_3(x) &= \mu_3(x) = E[x^3] \\
\kappa_4(x) &= \mu_4(x) - 3\mu_2(x)^2 = E[x^4] - 3E[x^2]^2
\end{aligned}$$

where $\mu_r(x)$ and $\kappa_r(x)$ are the r^{th}-order moment and cumulant, respectively.

Many interesting properties of cumulants can be found in [40]. In our work, we are mainly interested in the following characteristic: cumulants of order higher than 2 can remove the Gaussian noise present in the signal. It means that if $(z_i)_{i=1}^n$ are Gaussian random variables independent of $(y_i)_{i=1}^n (n > 2)$, then we have $\text{cumulant}(y_1 + z_1, y_2 + z_2, \ldots, y_n + z_n) = \text{cumulant}(y_1, y_2, \ldots, y_n)$.

In general, we do not have the knowledge about the probability density of the signal, the moments and cumulants are calculated by estimators. Let x be a centered scalar random variable, x_n $(n = 1 \ldots N_c)$ be realizations of x. The unbiased estimator of the fourth-order cumulant is formulated as [39]

$$\widehat{k}_4(x) = \frac{N_c + 2}{N_c(N_c - 1)} \sum_{i=1}^{N_c} x_i^4 - \frac{3}{N_c(N_c - 1)} \sum_{i,j=1}^{N_c} x_i^2 x_j^2. \quad (3)$$

IV. Cumulant-Based Analysis

A. Gaussian Noise Suppression Using the Fourth-Order Cumulant

Consider the side channel signal $W_{C_i}(t)$ corresponding to the message C_i. This signal can be considered as the sum of a useful signal and Gaussian noise $W_{C_i}(t) = S_{C_i}(t) + B(t)$. As cumulants of an order higher than two of a Gaussian random variable are equal to zero, the cumulants of the signal plus Gaussian noise are equal to the cumulants of the useful signal

$$\kappa_4(W_{C_i}) = \kappa_4(S_{C_i}) + \kappa_4(B) = \kappa_4(S_{C_i}). \quad (4)$$

The fourth-order cumulant is generally used versus the third-order one since for any signal with a symmetric probability density, its third-order cumulant is equal to zero. Therefore, we use the fourth-order cumulant in our case.

We perform the cumulant computation by sliding a window of N_w samples with a step $p = 1$ sample as illustrated in Fig. 1. The fourth-order cumulant of the signal in each window is computed using (3). As $p = 1$, the influence of $W_{C_i}(\tau)$ can be observed on N_w consecutive values $\widehat{k}_4(W_{C_i}, l)$ of the corresponding cumulant signal ($l \in [(\tau - N_w + 1), \tau]$). The value of $\widehat{k}_4(W_{C_i}, l)$ is given by the following formula [39]:

$$\begin{aligned}
\widehat{k}_4(W_{C_i}, l) = &\frac{N_w + 2}{N_w(N_w - 1)} \sum_{t=l}^{l+N_w-1} W_{C_i}^4(t) \\
&- \frac{3}{N_w(N_w - 1)} \sum_{t_1,t_2=l}^{l+N_w-1} W_{C_i}^2(t_1) W_{C_i}^2(t_2). \quad (5)
\end{aligned}$$

B. Comparison With the Noise Variance Subtraction Method

A standard noise reduction technique in signal processing is to calculate the noise variance and then subtract it. As the noise B is independent of the signal S_{C_i}, the power of S_{C_i} is the power of W_{C_i} minus the power of B

$$E\left[S_{C_i}^2\right] = E\left[W_{C_i}^2\right] - E[B^2]. \quad (6)$$

To illustrate this technique, we use the same sliding window and compute the power of the signal W_{C_i} in each window. Then we estimate the noise variance[1] and subtract it from the power of W_{C_i}. We obtain the power of the useful signal S_{C_i} as presented in the third curve of Fig. 1. While comparing the second and the third curves of Fig. 1, we observe that the contrast between the signal and the noise of the cumulant signal (the second one) is greater than that of the power signal (the third one). This can be explained by two points. First, the noise variance subtraction method requires an estimation of the noise variance while the cumulant method suppresses the noise itself. If this noise estimation is not exact, the useful information for DPA can be modified and the efficiency of the key detection may be reduced.

[1]We consider the signal between two consecutive peaks as noise.

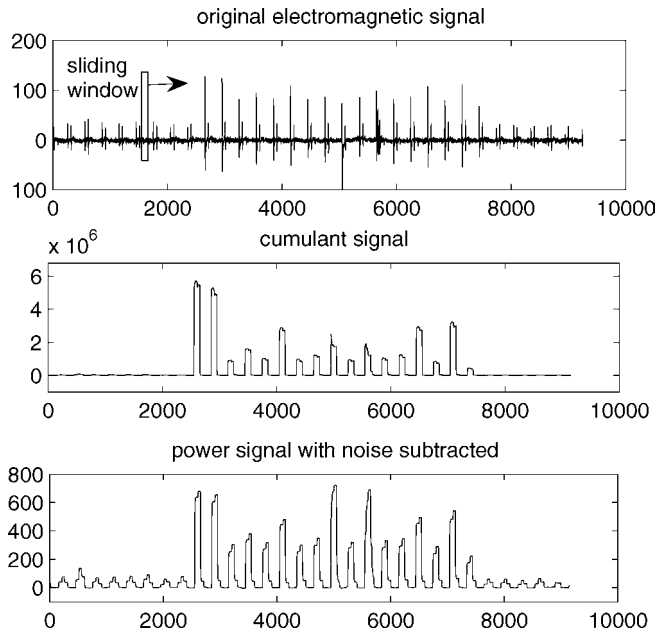

Fig. 1. Horizontal axes represent the time sampling proportional to clock cycles. The first vertical axis represents the voltage value on the output of an electromagnetic sensor (mV), the second one represents its fourth-order cumulant signal, and the third one represents the power signal with noise subtracted. The two last signals are obtained by sliding a window of 100 samples on the upper signal.

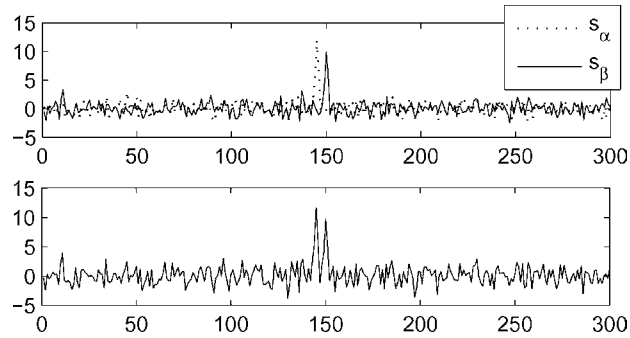

Fig. 2. Upper figure: misaligned signals s_α and s_β. Lower figure: sum of two signals $s = s_\alpha + s_\beta$.

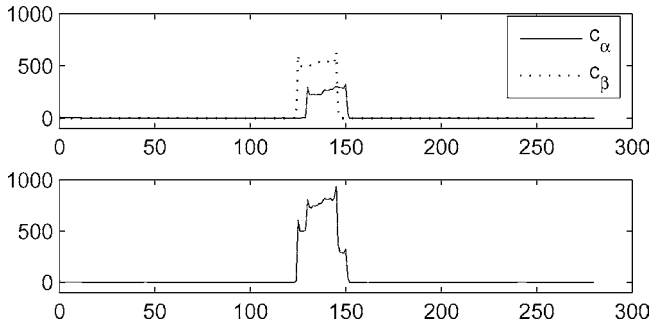

Fig. 3. Upper figure: misaligned cumulant signals c_α and c_β. Lower figure: sum of two cumulant signals $c = c_\alpha + c_\beta$.

Second, the efficiency of our method using the fourth-order cumulant is based on the fact that the useful signal is impulsive (strongly superGaussian). This property of the signal is characterized by high values of its kurtosis (i.e., the normalized fourth-order cumulant)

$$\mathcal{K}(S_{C_i}) = \frac{\kappa_4(S_{C_i})}{E\left[S_{C_i}^2\right]^2} \gg 1 \Leftrightarrow \kappa_4(S_{C_i}) \gg E\left[S_{C_i}^2\right]^2. \quad (7)$$

From (7), we deduce that for impulsive signals with high kurtosis values, its fourth-order cumulant $\kappa_4(S_{C_i})$ is superior to its power $E[S_{C_i}^2]$. Consequently, the relative amplitude of the cumulant signal to noise is greater than that of the power signal to noise. The dynamics of cumulant signals allows us to detect the correct key more easily.

C. Temporal Misalignment Correction

Like any attack based on a sliding window [25], the attack using cumulant signals makes it possible to minimize the effect of the lack of temporal synchronization. Note that the temporal misalignment in our paper does not refer to countermeasures, such as RPI or random order executions, but to the imprecision of measurements or the clock jittering. We consider two signals s_α and s_β which are not well aligned as represented in the upper figure of Fig. 2. The summed signal $s = s_\alpha + s_\beta$ is shown by the lower curve of Fig. 2. We clearly observe that the information contained in s_α and s_β is dispersed in two distinct peaks of s. The temporal misalignment of side channel signals reduces the attack effectiveness. If we use the cumulant signals c_α and c_β (upper figure of Fig. 3), the information in both signals c_α and c_β is then accumulated into the signal $c = c_\alpha + c_\beta$ (lower figure of Fig. 3). Hence, useful data of s_α and s_β converge in c and the effect due to the temporal misalignment can be reduced.

V. Theoretical Evaluation

A theoretical model was proposed in [41] to determine the effect of hardware countermeasures against DPA (noise adding and random disarrangement of the instant τ). However, this evaluation is only dedicated to original signals (power consumption and electromagnetic signals). The goal of this section is to provide a theoretical study which makes it possible to evaluate the DPA methods using different signal types: the original DPA [2], the cumulant DPA, and two other methods based on the sliding window technique: the integration DPA [25] and the energy DPA [42].

In [25], the sliding window concept was used to collect peaks distributed over consecutive cycles. This technique can also be applied when the peak is distributed over consecutive samples. In regard to the energy DPA proposed in [42], we make two remarks. First, the differential signal of DPA is the difference between two mean signals. If we use the energy signals instead of the original signals, after the subtraction, the noise variance of two mean signals will be removed. Therefore, the DPA method using energy signals suppresses implicitly the noise variance. Second, as the power of a signal is its energy divided by the signal length, the energy signals can be replaced by the power signals. In our case, the signal length is the sliding window length, which is fixed. Therefore, the energy DPA method is nothing other than the DPA using power signals presented in Section IV-B. Hereafter, we call this method the power DPA.

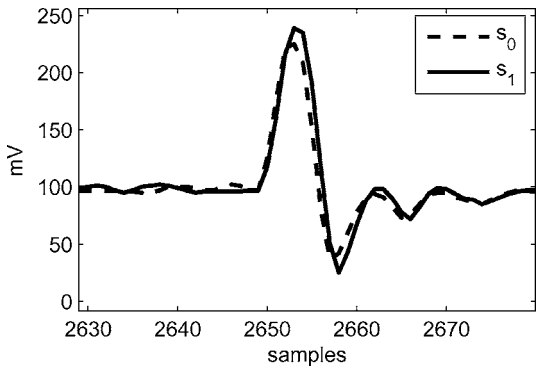

Fig. 4. Two electromagnetic signals s_1 and s_0.

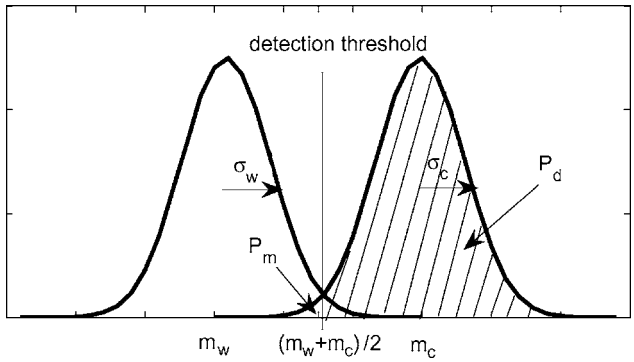

Fig. 5. Probability of detection P_d and the probability of a miss P_m.

We consider the monobit analysis where one bit b is examined. Side channel signals are distributed in two groups corresponding to the value of b ($b = 0$ or $b = 1$). If the bit b is handled to 1, the corresponding side channel signal is denoted by s_1, and if b is handled to 0, the corresponding side channel signal is denoted by s_0. Fig. 4 shows an example of two electromagnetic signals s_1 and s_0, which are used in the theoretical evaluation.

A. First Index: Probability of Detection

In our context, the probability of detection represents the capacity of correctly detecting the secret key among key hypotheses. This parameter is computed at the instant τ when the examined bit b is handled. In order to simplify the problem without loss of generality, we consider two hypotheses: the correct hypothesis $\mathcal{H}_c$ and the wrong hypothesis $\mathcal{H}_w$. We choose $m_c + m_w/2$ as the detection threshold. The key hypothes, whose peak is higher than $m_c + m_w/2$, is considered to be the correct key. Let us denote:

- h_c as the height of the detection peak of $\mathcal{H}_c$;
- h_w as the height of the detection peak of $\mathcal{H}_w$;
- m_c and m_w as the expectations of h_c and h_w;
- σ_c and σ_w as the standard deviations of h_c and h_w. As differential signals are computed from the same elementary signals, we can consider that the distributions of h_c and h_w are Gaussian with the same standard deviation, that is $\sigma_c = \sigma_w = \sigma'$ (Fig. 5).

The probability of detection is $P_d = 1 - P_m$, where P_m is the probability of a miss

$$P_m = \int_{-\infty}^{(m_w+m_c)/2} \frac{1}{\sqrt{2\pi}\sigma'} e^{-\frac{1}{2}\left(\frac{x-m_c}{\sigma'}\right)^2} dx.$$

The probability of detection can be written as

$$P_d = 1 - \frac{1}{2} erfc\left(\frac{m_c - m_w}{2\sqrt{2}\sigma'}\right) \tag{8}$$

where $erfc(x)$ is the complementary error function $erfc(x) = 1 - erf(x) = (2/\sqrt{\pi}) \int_x^\infty e^{-t^2} dt$. In order to compute P_d, we must calculate the values m_c, m_w, and σ' for each method: the original DPA, the integration DPA, the power DPA, and the cumulant DPA. These values depend on the signals s_0, s_1, and the noise level (see the Appendix).

TABLE I
THEORETICAL VALUES OF THE DPA PEAK HEIGHT AND THE NOISE LEVEL

Method	DPA peak	Noise	SNR
Original	$s_1(\tau) - s_0(\tau)$	$2\sigma\sqrt{N}$	$\frac{(s_1(\tau)-s_0(\tau))}{2\sigma\sqrt{N}}$
Integration	$\bar{s}_1 - \bar{s}_0$	$2\sigma/\sqrt{N_w N}$	$\frac{\bar{s}_1-\bar{s}_0}{2\sigma/\sqrt{N_w N}}$
Power	$E(s_1^2) - E(s_0^2)$	$2\sqrt{2}\sigma^2/\sqrt{N_w N}$	$\frac{E(s_1^2)-E(s_0^2)}{2\sqrt{2}\sigma^2/\sqrt{N_w N}}$
Cumulant	$\hat{k}_4(s_1) - \hat{k}_4(s_0)$	$2\sqrt{24}\sigma^4/\sqrt{N_w N}$	$\frac{\hat{k}_4(s_1)-\hat{k}_4(s_0)}{2\sqrt{24}\sigma^4/\sqrt{N_w N}}$

B. Second Index: SNR

In many cases, we do not have any knowledge about the instant τ when the examined bit is handled. The detection peak of the correct key cannot be observed because it is covered with noise. We thus define the second parameter, which is the SNR of the differential curve corresponding to the correct key. The detection peak is considered to be the signal and the other parts of the curve are defined as noise. The SNR of each method is

$$\text{SNR} = \frac{\text{height of the detection peak}}{\text{standard deviation of noise}}. \tag{9}$$

The theoretical values of DPA peak height and the standard deviation of noise are given in Table I. ($\bar{s}$ denotes the mean of s). The calculations of noise are given in the Appendix.

Note that $\hat{k}_4(s) = \mathcal{K}(s).E(s^2)^2$, where $\mathcal{K}(s)$ is the kurtosis of the signal s. We develop SNR_{cum} of the cumulant DPA method as follows:

$$\begin{aligned}\text{SNR}_{\text{cum}} &= \frac{\hat{k}_4(s_0) - \hat{k}_4(s_1)}{2\sqrt{24}\sigma^4/\sqrt{N_w N}} \\ &= \frac{\mathcal{K}(s_1).E\left(s_1^2\right)^2 - \mathcal{K}(s_0).E\left(s_0^2\right)^2}{2\sqrt{2}\sigma^2/\sqrt{N_w N}} . \frac{1}{\sqrt{12}\sigma^2}\end{aligned}.$$

As $\mathcal{K}(s_1) \approx \mathcal{K}(s_0)$, we have

$$\begin{aligned}\mathrm{SNR}_{\mathrm{cum}} &\approx \frac{\mathcal{K}(s_0)}{\sqrt{12}\sigma^2}.\frac{E\left(s_1^2\right)^2 - E\left(s_0^2\right)^2}{2\sqrt{2}\sigma^2/\sqrt{N_w N}} \\ &= \frac{\mathcal{K}(s_0)\left(E\left(s_1^2\right) + E\left(s_0^2\right)\right)}{\sqrt{12}\sigma^2}.\frac{E\left(s_1^2\right) - E\left(s_0^2\right)}{2\sqrt{2}\sigma^2/\sqrt{N_w N}} \\ &= \frac{\mathcal{K}(s_0)}{\sqrt{3}}.\frac{E\left(s_1^2\right) + E\left(s_0^2\right)}{2\sigma^2}.\mathrm{SNR}_{\mathrm{power}}.\end{aligned}$$

The value σ^2 represents the noise variance of s_0 and s_1, so $E(s_1) > \sigma^2$, $E(s_0) > \sigma^2$, or $(E(s_1^2) + E(s_0^2)/2\sigma^2) > 1$. As signal s_0 (and s_1) is impulsive,[2] we have $(\mathcal{K}(s_0)/\sqrt{3}) \gg 1$. We obtain:

$$\mathrm{SNR}_{\mathrm{cum}} > \mathrm{SNR}_{\mathrm{power}}. \tag{10}$$

The previous demonstration confirms the advantage of the cumulant method compared to the power one when the examined signal is impulsive (i.e., its kurtosis is highly superior to 1).

C. About CPA Using Fourth-Order Cumulant Signals

If the Hamming weight or the Hamming distance model is adopted $S_{C_i}(\tau) = aH_{i,R}(\tau)+b$, where τ is the instant when the data are handled, $H_{i,R}(\tau)$ is the Hamming weight or Hamming distance of the data and a and b are constant values. The side channel $W_{C_i}(t)$ at the instant τ becomes

$$W_{C_i}(\tau) = aH_{i,R}(\tau) + b + B(\tau). \tag{11}$$

As $B(t)$ is Gaussian noise, it will disappear after the cumulant computation. We can write $\widehat{k_4}(W_{C_i}) = a_3 H_{i,R}(\tau) + [a_0 H_{i,R}^4(\tau) + a_1 H_{i,R}^3(\tau) + a_2 H_{i,R}^2(\tau) + a_4]$, where a_0, a_1, a_2, a_3, and a_4 are constant values computed from a and b. As only the term $a_3 H_{i,R}(\tau)$ contributes in the correlation factor between $\widehat{k_4}(W_{C_i})$ and $H_{i,R}$, this correlation factor is exactly the one between W_{C_i} and $H_{i,R}$ multiplied by a constant. The probability of detection of CPA with cumulant signals, calculated at the instant τ, is equivalent to that of CPA with original signals. However, using cumulant signals, the SNR of CPA will clearly be enhanced because of the signal denoising. Consequently, the key detection is more efficient.

D. Discussion

The criteria defined previously allow us to evaluate the performance of an attack. A method is powerful if both the probability of detection P_d and the SNR of the DPA signal are high. It is obvious that P_d and SNR depend on the number of side channel signals N and the standard deviation of noise in a side channel signal σ. However, these dependences are simple: when N increases, P_d and SNR increase and when σ increases, P_d and SNR decrease. In this section, we present only the variation of P_d given by (8) (Fig. 6), and the variation of SNR given in Table I (Fig. 7) according to the sliding window length N_w. The number of side channel signals is set to $N = 200$ and the noise level of an elementary side channel signal is $\sigma = 20$ mV.

[2]For example, if we truncate an electromagnetic peak (the first curve of Fig. 1) by a window of 100 samples, its kurtosis is about 13.

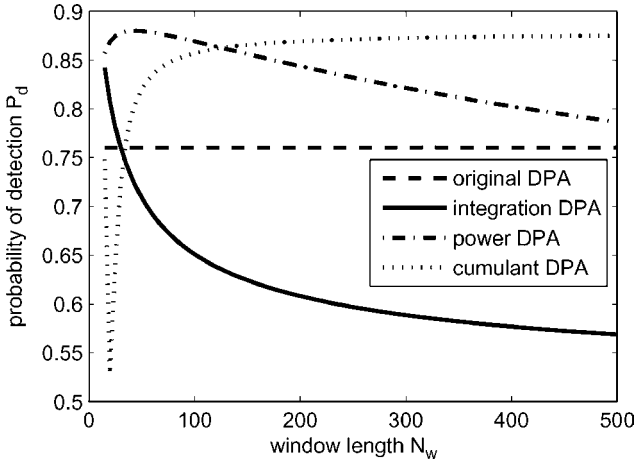

Fig. 6. Variation of the detection probability in function of N_w.

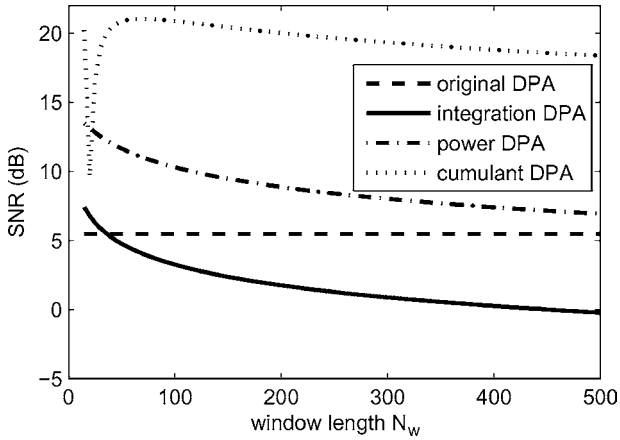

Fig. 7. Variation of SNR in a function of N_w.

The P_d and the SNR of the original method, which does not use the sliding window technique, are independent of N_w. They are thus represented by horizontal lines (Figs. 6 and 7). For the integration method, the longer the window is, the greater the noise is added to the window. Consequently, its P_d and SNR decrease rapidly when the window length increases. The decrease of SNR of the integration DPA was explained in [25]. This method is even worse than the original one if N_w is large ($N_w > 100$). It means that the integration method is only applicable to weak misalignments [i.e., the peak is distributed over a small number of consecutive samples (or cycles)]. Regarding the power DPA, the fact that the noise variance is removed by the subtraction of two mean signals makes it better than the original and the integration methods.

The variations of P_d and SNR corresponding to the cumulant method presents a fall when $N_w = 20$. It can be explained by two reasons. On one hand, when the sliding window is too small, the signal cannot be considered as impulsive. Accordingly, its kurtosis is not high, and the values of P_d and SNR are close to 0. On the other hand, if the window is not large enough, the assumption about the Gaussian noise may not hold and the cumulant of noise can be different from 0. When the window is large, the conditions of impulsive signal and of Gaussian noise hold.

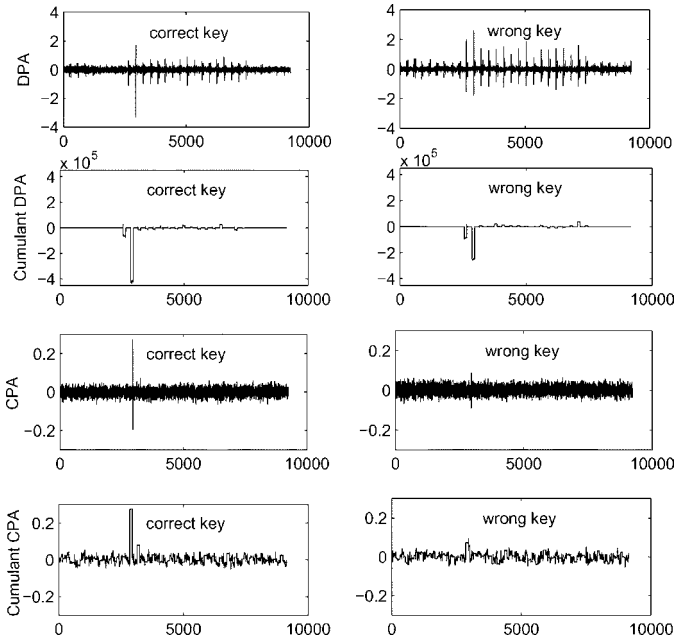

Fig. 8. DPA, cumulant DPA, CPA and cumulant CPA signals. Left column: signals corresponding to the correct key. Right column: signals corresponding to a wrong key.

Therefore, the cumulant DPA performs better than the power DPA and it becomes the best method in both criteria: the probability of detection and the SNR.

VI. Experimental Results

For a real experiment, the probability of detection is replaced by index i_1, which is easier to compute. It is defined as the ratio between the DPA/CPA peak corresponding to the correct key (expected peak) and the highest DPA/CPA peak resulting from incorrect keys (ghost peaks). These peaks are observed at the same time location τ when the data are handled. If this index is greater than 1, the expected peak is higher than any ghost peak and the key detection is reliable. In contrast, if this index is smaller than 1, a ghost peak exists which is higher than the expected peak and the method is not effective. The second index, denoted i_2, is the signal-to-noise ratio of the DPA/CPA signal corresponding to the correct key. This index is the SNR defined in the previous section.

A. Experimental Validation of the Cumulant-Based Analysis

In our experiment, we measure the electromagnetic emanations of a synthesized application-specific integrated circuit (ASIC) during a DES operation. The sampling rate is 612.5 MHz and the clock rate is 2.1 MHz. We obtain an electromagnetic signal from each random message used in the input (upper curve of Fig. 1). Here, the notation $W_{C_i}(t)$ represents the voltage value at the output of our electromagnetic sensor corresponding to the message C_i.

In the first experiment, we used 3000 messages to test 64 key assumptions. The DPA and CPA signals were computed using

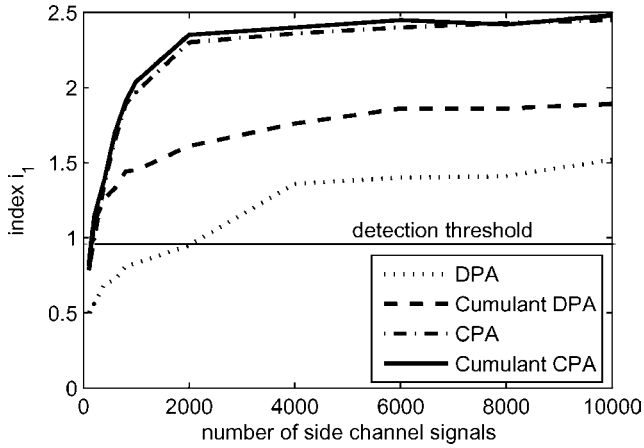

Fig. 9. Variation of the index i_1.

(1) and (2), respectively.[3] For DPA, a selection function based on 1-b Hamming distance was used. For the CPA method, we examined 4 b. The cumulant signals were collected by sliding a window of $N_w = 100$ samples with a step $p = 1$. The choice of $N_w = 100$ is verified by the theoretical evaluation in Section V. It corresponds to the high values of P_d and SNR (Figs. 6 and 7).

Fig. 8 represents from top to bottom the DPA, cumulant-based DPA, CPA, and cumulant-based CPA signals corresponding to the correct key (left column) and a wrong key resulting in the highest ghost peak (right column). First, the results show that all four methods allow the retrieval of the correct key. It means that the cumulant operation does not eliminate the useful information for DPA and CPA in the electromagnetic signals. Second, thanks to the high dynamic of cumulant signals, the peaks at other instants than τ of DPA signals (the secondary peaks), which appear frequently in monobit DPA, are clearly reduced using the cumulant method. Third, we observe a high level of noise in the CPA signal. Index i_2 gives a good measure of the noise problem.

B. Performance Evaluation

The variation of index i_1, when the number of cipher messages varies from 100 to 10000 message, is illustrated in Fig. 9. This figure shows that the cumulant-based DPA method performs much better than the original DPA. This improvement is explained by the fact that the cumulant operation removes the Gaussian noise impact, corrects the misalignment, and keeps the difference of power dissipation to manipulate one bit to 1 or to 0. When comparing CPA and cumulant-based CPA, we see that the latter method still works but its improvement is not significant.

The evaluation of index i_2 is depicted in Fig. 10. It shows that the SNR of DPA and cumulant-based DPA signals are always good. Index i_2 of CPA- and cumulant-based CPA methods is low because of the normalization of CPA [13].

[3]The main goal here is not to compare DPA and CPA but to investigate the effect of cumulant computation. In the experimental result, DPA is performed with 1-b weighting and CPA with 4-b weighting. Thus, CPA will give better results.

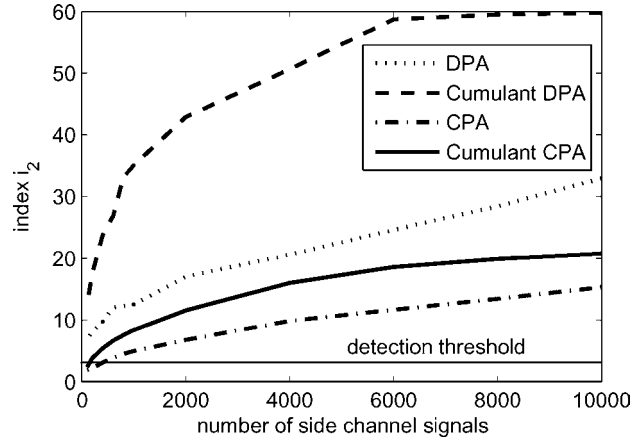

Fig. 10. Variation of the index i_2.

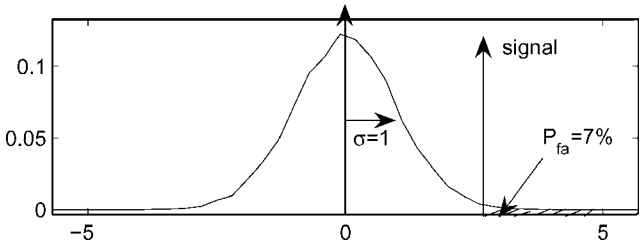

Fig. 11. Choice of i_2.

The key detection depends on both indexes i_1 and i_2. It is feasible and reliable if the two following conditions are satisfied $i_1 > 1$ and $i_2 > 3$. The first condition is trivial. The choice of i_2 depends on the probability of false alarm (see Fig. 11). For a centered normalized Gaussian noise and a signal of 3, the SNR is equal to 3, then the corresponding probability of false alarm (i.e., noise amplitude $>$ signal amplitude) is about 7%.

According to Figs. 9 and 10, the DPA method needs about 2500 messages and CPA needs about 400 messages to detect the correct key. By using the cumulant tool, our proposed methods require only 200 messages to retrieve the encryption key. Fig. 12 confirms our conclusion about the required number of messages. The left column signals correspond to the experiment with 400 cipher messages. We see the appearance of many unexpected peaks in DPA signals (i.e., the encryption key cannot be uncovered). Meanwhile, CPA, the cumulant-based DPA, and CPA methods are effective. If the number of messages is reduced to 200, only the cumulant-based DPA/CPA methods allow detection of the secret key.

The experimental results show that our cumulant based methods are more powerful than the original ones. The cumulant application improves DPA significantly in terms of the number of messages. Instead of using 2500 messages, the cumulant based DPA needs only 200 messages.

Note that in this experiment, the misalignment of signals is relatively weak (about 3, 4 samples). If the misalignment becomes more important, the key detection of the original DPA and CPA methods will be reduced. The performance of cumulant methods, which use the sliding window technique, is not affected. In this case, the attack efficiency is much more remarkable.

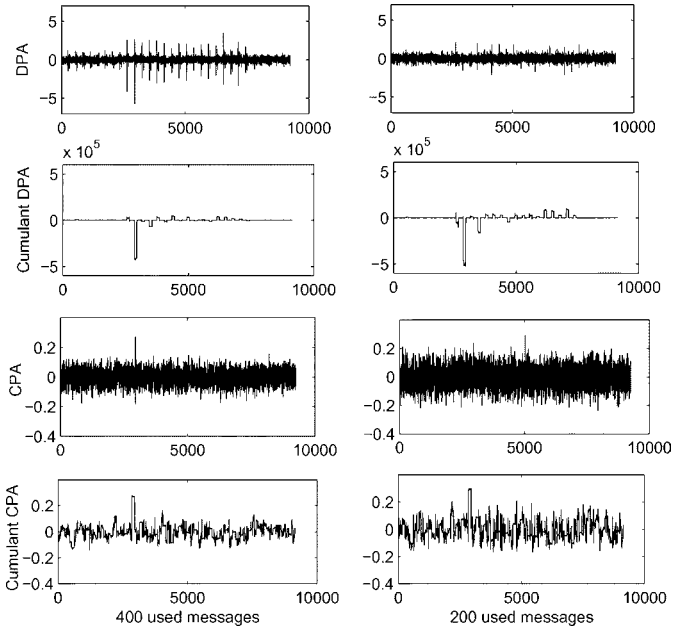

Fig. 12. DPA, cumulant DPA, CPA, and cumulant CPA signals. Left column: 400 used messages. Right column: 200 used messages.

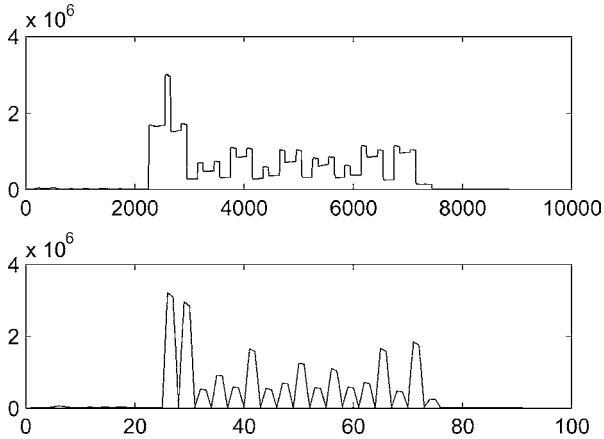

Fig. 13. Upper: a cumulant signal with $N_w = 400$ and $p = 1$. Lower: a cumulant signal with $N_w = 100$ and $p = 100$.

C. Choice of the Window Length and the Sliding Step

As we observed in the previous paragraph, the cumulant method gives good values of index i_2; hence, the detection efficiency is related to i_1. It depends strongly on the choice of the window size N_w and the sliding step p. One should note that the relevant information from the side channel signals of DES operation is located around 16 peaks corresponding to 16 rounds of DES. In our case, the distance d between two consecutive peaks is about 300 samples. If we choose $N_w > d$, some positions of the sliding window exist that contain two consecutive peaks. After performing cumulant calculation, the information included in two consecutive peaks will be merged into one large cumulant peak as depicted in the upper curve of Fig. 13 ($N_w = 400$, $p = 1$). The 16 original peaks in the side channel signal are completely deformed, and the effectiveness of the cumulant DPA will be degraded.

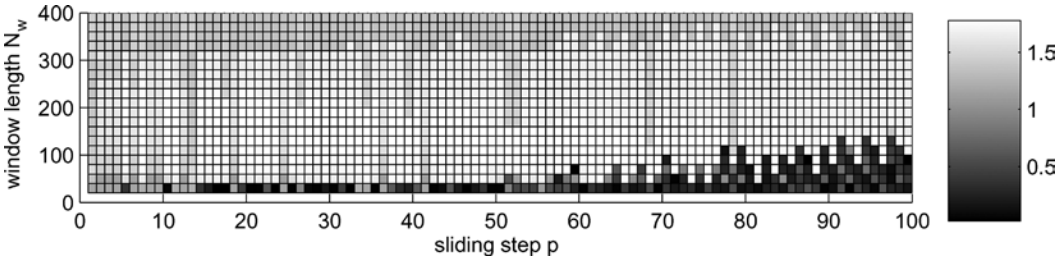

Fig. 14. First index evaluation with 1000 messages.

The cumulant signals used in Section VI are computed by sliding the window with a step $p = 1$ which makes it possible to examine exhaustively the behavior of side channel signals. However, when $p = 1$, the cumulant information of each electromagnetic peak will be repeated N_w times in cumulant signals. We can reduce this redundancy by selecting a greater value of p. Note that step p should be smaller than or equal to the sliding window size ($p \leq N_w$) to avoid the loss of information. When $p = N_w$, electromagnetic signals are truncated into pieces of size N_w. The lower curve of Fig. 13 represents a cumulant signal with $N_w = 100$ and $p = 100$. We observe that the signal size is divided by $p = 100$ and the peaks are still well distinguished.

In order to investigate the performance fluctuation of the cumulant-based DPA method in the function of (N_w, p), we observe the variation of index i_1 as presented in Fig. 14 when varying p from 1 to 100 and N_w from 20 to 400. This result is obtained by using 1000 side channel signals. First, the result confirms that when N_w is superior to the distance between two electromagnetic peaks ($N_w > 300$), or when N_w is too small ($N_w < 60$), index i_1 decreases. The best values for the window length are about $80 < N_w < 200$, which are consistent with Figs. 6 and 7. Second, when $N_w < p$, we observe the bad values of i_1 due to the loss of information. Third, the performance of a cumulant-based DPA method is reduced when the values of p become dividers of τ. In our case, $\tau = 2952$ is the position where data are handled, thereby for $p = 1$, 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, 41, the dividers of 2952, i_1 is small. It can be explained by the fact that the position τ will be found at the edge of a sliding window. This is called the edge effect (see Fig. 3). It affects the cumulant calculation which then affects the effectiveness of the cumulant-based DPA. In conclusion, the values (N_w, p) which satisfy the conditions 1) $N_w < d$, 2) $p < N_w$, and 3) p not being a divider of τ, will allow the detection of the correct key with a good attack-efficient index i_1.

VII. Conclusion

In this paper, we have proposed a new method to reduce the Gaussian noise of signals in DPA and CPA attacks using the fourth-order cumulant of side channel signals. We have given the theoretical evaluation based on two criteria—the probability of detection and the SNR. The formulas to calculate these parameters have been given under a general form with flexible parameters, such as the noise level, the number of side channel signals, and the length of sliding window. They can be applied to any type of side channel signals for selecting the most suitable parameters. The proposed cumulant method has been validated by real experiments with electromagnetic signals of an ASIC. The cumulant method is a powerful solution for the noise suppression and the temporal misalignment correction in a side channel attack.

Appendix

In this appendix, we want to show how the values m_c, m_w, and σ' defined in Section V-A can be computed from the signals s_1, s_0, and the standard deviation of noise σ of an elementary electromagnetic signal. We first present the common part of all methods. Then, the calculations dedicated to each method (the integration DPA, the power DPA, and the cumulant DPA) are separately developed. The calculations of noise are shown in the last section of the Appendix.

We assume that for the correct hypothesis H_c, N signals are correctly distributed in two groups G_0, G_1 and the number of signals in each group is $N/2$. In the case of H_w, for each group G_0 and G_1, there are n signals that are wrongly distributed and only $(N/2-n)$ signals are correctly placed in their group. To obtain the results shown in Section V, we have chosen $n = (N/10)$. The value of n does not change the relative results between the examined methods.

The probability of detection is $P_d = 1 - (1/2)erfc((m_c - m_w/2\sqrt{2}\sigma'))$. As the differential signals are computed from elementary signals, to obtain m_c, m_w and σ', we have to calculate the expectation and the variance of the peak at the instant τ of each elementary signal. If we note m_1 and σ_1, the expectation and the standard deviation of the peak of the signal s_1, m_0, and σ_0, respectively, for s_0, it is easy to demonstrate that

$$m_c = m_1 - m_0 \tag{12}$$

$$m_w = \left(1 - \frac{4n}{N}\right)(m_1 - m_0) \tag{13}$$

$$\sigma_c = \sigma_w = \sigma' = \sqrt{\frac{\sigma_1^2}{N/2} + \frac{\sigma_0^2}{N/2}}. \tag{14}$$

The calculations of m_0 and m_1 (or σ_0 and σ_1) are similar. Hence, we consider a signal $S(t)$ (which can be s_1 or s_0) of length N_w, added by a centered i.i.d Gaussian noise $B(t)$ of variance σ^2. After the integration, power or cumulant operations in the window N_w, $S(t)$ is replaced by a value v. We calculate the mathematical expectation and the variance of v in three cases: integration, power, and cumulant operations

Integration operation

$$v = \frac{1}{N_w}\sum_{t=1}^{N_w} S(t) + B(t).$$

The mathematical expectation and the variance of v are

$$\begin{aligned} E[v] &= \frac{1}{N_w}\sum_{t=1}^{N_w} E\left(S(t)+B(t)\right) = \frac{1}{N_w}\sum_{t=1}^{N_w} S(t) \\ var[v] &= \frac{1}{N_w^2}\sum_{t=1}^{N_w} var\left(B(t)\right) = \frac{1}{N_w^2} N_w\sigma^2 = \frac{\sigma^2}{N_w} \end{aligned}.$$

So for the intergration method, m_c, m_w, and σ' can be calculated by replacing in (12)–(14) $m_0 = (1/N_w)\sum_{t=1}^{N_w} s^0(t)$, $m_1 = (1/N_w)\sum_{t=1}^{N_w} s^1(t)$, $\sigma_0^2 = \sigma_1^2 = (\sigma^2/N_w)$.

Power operation

$$\begin{aligned} v &= \frac{1}{N_w}\sum_{t=1}^{N_w} \left(S(t)+B(t)\right)^2 \\ E[v] &= \frac{1}{N_w}\sum_{t=1}^{N_w} E\left[S(t)^2 + 2S(t)B(t) + B(t)^2\right] \\ &= \frac{1}{N_w}\sum_{t=1}^{N_w} S(t)^2 + E\left[B(t)^2\right] = \frac{\sum_{t=1}^{N_w} S(t)^2}{N_w} + \sigma^2 \\ &\mathrm{var}\left[\sum_{t=1}^{N_w}\left(S(t)+B(t)\right)^2\right] \\ &= E\left[\sum_{t_1=t_2=1}^{N_w}\left(S(t_1)+B(t_1)\right)^2\left(S(t_2)+B(t_2)\right)^2\right] \\ &\quad + E\left[\sum_{t_1\neq t_2=1}^{N_w}\left(S(t_1)+B(t_1)\right)^2\left(S(t_2)+B(t_2)\right)^2\right] \\ &\quad - \sum_{t_1=t_2=1}^{N_w} E\left[\left(S(t_1)+B(t_1)\right)^2\right] E\left[\left(S(t_2)+B(t_2)\right)^2\right] \\ &\quad - \sum_{t_1\neq t_2=1}^{N_w} E\left[\left(S(t_1)+B(t_1)\right)^2\right] E\left[\left(S(t_2)+B(t_2)\right)^2\right] \\ &= \sum_{t=1}^{N_w} E\left[\left(S(t)+B(t)\right)^4\right] - \sum_{t=1}^{N_w}\left(E\left[\left(S(t)+B(t)\right)^2\right]\right)^2 \\ &= \sum_{t=1}^{N_w}\left(S(t)^4 + 6S(t)^2\sigma^2 + 3\sigma^4\right) - \left(S(t)^2+\sigma^2\right)^2 \\ &= \sum_{t=1}^{N_w}\left(4S(t)^2\sigma^2 + 2\sigma^4\right). \end{aligned} \quad (15)$$

Let us denote $M_r = (1/N_w)\sum_{t=1}^{N_w} S(t)^r$, we can write

$$\mathrm{var}[v] = \frac{4M_2\sigma^2 + 2\sigma^4}{N_w}.$$

Cumulant operation

$$v = \widehat{k}_4\left(S(t)+B(t)\right).$$

By performing the same calculation steps of the power operation, we can obtain the expectation and the variance of the cumulant. As these calculations are quite complex, we present the final results here

$$E[v] = \widehat{k}_4\left(S(t)\right).$$

Denote that $M_r = (1/N_w)\sum_{t=1}^{N_w} S(t)^r$, $\mathrm{var}[v]$ is given by

$$\begin{aligned} \mathrm{var}[v] = &\frac{\sigma^2\left(16M_6 - 96M_4M_2 + 144M_2^3\right)}{N_w} \\ &+ \frac{\sigma^4\left(72M_4 - 72M_2^2\right) + 96\sigma^6M_2 + 24\sigma^8}{N_w}. \end{aligned}$$

Calculations of the noise level: In each case, the variance of noise is given by the value of $\mathrm{var}[v]$ with $S(t) = 0$ for all t (noise only, no signal). To obtain the standard deviation of noise, we extract the root of the variance $\mathrm{var}[v]$. In consequence, we have

$$\begin{aligned} \sigma_{\mathrm{noise,integration}} &= \sqrt{\sigma^2/N_w} = \sigma/\sqrt{N_w} \\ \sigma_{\mathrm{noise,power}} &= \sqrt{2\sigma^4/N_w} = \sqrt{2}\sigma^2/\sqrt{N_w} \\ \sigma_{\mathrm{noise,cumulant}} &= \sqrt{24\sigma^8/N_w} = \sqrt{24}\sigma^4/\sqrt{N_w}. \end{aligned}$$

REFERENCES

[1] P. Kocher, "Timing attack on implementation of Diffie-Hellman, RSA, DSS and other systems," in *Proc. Advances Cryptology*, CA, 1996, pp. 104–113.

[2] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Proc. CRYPTO*, CA, 1999, pp. 388–397.

[3] *Data Encryption Standard (DES)*, Std. FIPS PUB 46-3, U.S. Dept. Commerce Nat. Inst. Standards Technol., 1999.

[4] K. Gandolfi, C. Mourtel, and F. Olivier, "Electromagnetic Attacks: Concrete Results," in *Proc. CHES* , Paris, France, 2001, pp. 252–261.

[5] J. Quisquater and D. Samyde, "Electromagnetic Analysis (EMA): Measures and countermeasures for smart cards," in *Proceedings e-Smart*, Sophia Antipolis, Greece, 2001, pp. 200–210.

[6] J. Rao and P. Rohatgi, EMpowering side-channel attacks. Cryptology ePrint archive Rep. 2001/037, 2001. [Online]. Available: http://www.eprint.iarc.org/.

[7] L. Goubin and J. Patarin, "DES and differential power analysis: The duplication method," in *Proc. CHES*, MA, 1999, pp. 158–172.

[8] M. Akkar and C. Giraud, "An Implementation of DES and AES secure against some attacks," in *Proc. CHES*, Paris, France, 2001, pp. 309–318.

[9] M. Akkar and L. Goubin, "A generic protection against high-order differential power analysis," in *Proc. FSE*, Lund, Sweden, 2003, pp. 192–205.

[10] J. Coron and L. Goubin, "On boolean and arithmetic masking against differential power analysis," in *Proc. CHES*, MA, 2000, pp. 231–237.

[11] R. Bevan and E. Knudsen, "Ways to enhance DPA," in *Proc. ICISC*, Seoul, Korea, 2002, pp. 327–342.

[12] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, "Examining smart-card security under the threat of power analysis attacks," *IEEE J. Comput.*, vol. 51, no. 5, pp. 541–552, May 2002.

[13] T.-H. Le, J. Clédière, C. Canovas, C. Servière, J.-L. Lacoume, and B. Robisson, "A proposition for correlation power analysis enhancement," in *Proc. CHES*, Yokohama, Japan, 2006, pp. 174–186.

[14] R. Bevan, "Estimation statistique et sécurité des cartes à puce, évaluation d'attaques dpa évolués," Ph.D. dissertation, Dept. Faculty Sci., Univ. Paris-Sud 11, Supélec, France, 2004.

[15] A. Chari, J. Rao, and P. Rohatgi, "Template attacks," in *Proc. CHES*, San Francisco, CA, 2002, pp. 13–28.

[16] C. Rechberger and E. Oswald, "Practical template attacks," in *Proc. WISA*, 2004, pp. 440–456.
[17] D. Agrawal, J. Rao, P. Rohatgi, and K. Schramm, "Templates as master keys," in *Proc. CHES*, Edinburgh, U.K., 2005, pp. 15–29.
[18] W. Schindler, K. Lemke, and C. Paar, "A stochastic model for differential side channel cryptanalysis," in *Proc. CHES*, Edinburgh, U.K., 2005, pp. 30–46.
[19] E. Sangfelt and L. Persson, "Experimental performance of some higher-order cumulant detectors for hydroacoustic transients," in *Proc. IEEE Signal Processing Workshop H.O.S*, South Lake Tahoe, CA, Jun. 1993, pp. 182–186.
[20] B. Porat and B. Friedlander, "Performance analysis of cumulant based detection of non-gaussian signals," *J. Adapt. Control Signal Process.*, vol. 10, pp. 99–112, 1996.
[21] N. E. Weste and K. Eshraghian*, Principles of CMOS VLSI Design*. Reading, MA: Addison-Wesley, 1994.
[22] A. V. D. Ziel*, Noise in Solid State Devices and Circuits*. New York: Wiley, 1986.
[23] P. Lowenborg and H. Johansson, "Quantization noise in filter bank analog-to-digital converters," in *Proc. Circuits Systems*, Geneva, Switzerland, 2001, vol. 2, pp. 601–604.
[24] C. C. Tiu, "A new frequency-based side channel attack for embedded systems," M.Eng. dissertation, Dept. Elect. Comput. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2005.
[25] C. Clavier, J. Coron, and N. Dabbous, "Differential power analysis in the presence of hardware countermeasures," in *Proc. CHES*, Worcester, MA, 2000, pp. 252–263.
[26] C. Gebotys, S. Ho, and A. Tiu, "EM analysis of Rijndael and ECC on a wireless java-based PDA," in *Proc. CHES*, Edinburgh, U.K., 2005, pp. 250–264.
[27] N. Homma, S. Nagashima, Y. Imai, T. Aoki, and A. Satoh, "High-resolution side-channel attack using phase-based waveform matching," in *Proc. CHES*, Yokohama, Japan, 2006, pp. 187–200.
[28] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Proc. CHES*, Cambridge, MA, 2004.
[29] C. Canovas and J. Clédière, What do s-Boxes Say in differential side channel attacks? Cryptology ePrint archive Rep. 20085/311, 2005. [Online]. Available: http://www.eprint.iarc.org/.
[30] R. Mayer-Sommer, "Smartly analysing the simplicity and the power of simple power analysis on smartcards," in *Proc. CHES*, Worcester, MA, 2000, pp. 78–92.
[31] J. Coron, P. Kocher, and D. Naccache, "Statistics and secret leakage," in *Proc. Financial Cryptography*, Anguilla, British West Indies, 2001, pp. 157–173.
[32] M. Dogan and J. Mendel, "Cumulant-based blind optimum beamforming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 722–741, Jul. 1994.
[33] Y. Chen and Y. S. Lin, "Fourth-order cumulant matrices for DOA estimation," in *Proc. IEEE Radar, Sonar Navigation*, 1994, vol. 141, pp. 144–148.
[34] M. Feng and K.-D. Kammeyer, "Suppression of gaussian noise using cumulants: A quantitative analysis," *Proc. ICASSP*, vol. 5, pp. 3813–3816, 1997.
[35] X. Fan and N. Younan, "Asymptotic analysis of the cumulant-based MUSIC method in the presence of sample cumulant errors," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 799–802, Mar. 1995.
[36] A. Swami and J. Mendel, "Cumulant-based approach to harmonic retrieval and related problems," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1099–1109, May 1991.
[37] J. Cardoso and A. Souloumiac, "An efficient technique for the blind separation of complex sources," in *Proc. IEEE Signal Processing Workshop Higher-Order Statistics*, 1993, pp. 275–279.
[38] J. Lacoume and M. Gaeta, "The general source separation problem," in *Proc. 5th ASSP Workshop Spectrum Estimation and Modeling*, 1990, pp. 154–158.
[39] M. Kendall and A. Stuart*, The Advanced Theory of Statistics*, 2nd ed. London, U.K.: Charles Griffin, 1963.
[40] J. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theorical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.
[41] S. Mangard, "Hardware countermeasure against DPA—A statistical analysis of their effectiveness," in *Proc. CT-RSA*, 2004, vol. 2964, Lecture Notes Comput. Sci., pp. 222–235.
[42] T.-H. Le, J. Clédière, C. Servière, and J.-L. Lacoume, "Efficient solution for misalgnment of signal in side channel analysis," presented at the ICASSP, Honolulu, HI, Apr. 2007.