DPA Attacks and S-Boxes

Emmanuel Prouff

Oberthur Card Systems
25 rue Auguste Blanche, 92800 Puteaux, France
e.prouff@oberthurcs.com

Abstract. For the power consumption model called *Hamming weight model*, we rewrite DPA attacks in terms of correlation coefficients between two Boolean functions. We exhibit properties of S-boxes (also called (n, m)-functions) relied on DPA attacks. We show that these properties are opposite to the non-linearity criterion and to the propagation criterion. To quantify the resistance of an S-box to DPA attacks, we introduce the notion of *transparency order of an S-box* and we study this new criterion with respect to the non-linearity and to the propagation criterion.

1 Introduction

Block cipher algorithms embedded in cryptographic devices are sensitive to two main kinds of attacks, which are usually investigated in parallel. The first kind relies on the properties of the cryptographic primitives involved in the cryptosystem. The second kind is based on the analysis of the hardware's leakages.

The most well-known attacks against block ciphers algorithms are the *known-plaintext attacks* called *differential cryptanalysis* [2, 13] and *linear cryptanalysis* [19]. Most block cipher algorithms (such as DES or AES) use vectorial functions, also called *S-boxes*, as cryptographic primitives. To protect such cryptosystems against linear and differential attacks, S-boxes are designed to fulfil some cryptographic criteria (balancedness, high nonlinearity or high algebraic degree).

Since electronic components are not usually perfectly tamper-proof, one can obtain sensitive information from side channels such as the timing of operations or the power consumption. In 1996, Kocher successfully used this approach to exhibit a first *side-channel attack* effective enough to recover secret keys in numerous cryptosystems [14]. Since Kocher's original paper, a large number of very efficient attacks has been reported on a wide variety of cryptographic implementations (see for instance [4,5,8,11,20,23]). Among these attacks, the *Differential Power Analysis* (DPA) is one of the most powerfull against iterated block ciphers. DPA are usually used to attack on either the first or the last round but it can sometimes be applied to attack on intern rounds of the block ciphers. It requires the knowledge of either the plaintext or the ciphertext. It relies on a statistical analysis of a large number of samples where the same key operates on different data. For this strategy of attacks, S-boxes involved in the cryptosystems are usually considered by cryptanalysts and also by cryptographers as oracles providing the output corresponding to a given data. So, to withstand DPA attacks, countermeasures are added at the implementation level to make the signals needed for these attacks useless.

The efficiency of DPA attacks is much greater than the efficiency of differential or linear cryptanalysis [1]. Moreover, in the area of embedded cryptography, because of the life expectancy of the device, known-plaintext attacks requiring a large number of pairs plaintext/ciphertext or requiring a large number of encryptions are unpracticable. The difference between the efficiencies of the two categories of attacks is not taken into account to design block ciphers for smart cards. Indeed, nearly all the algorithms embedded in smart cards have been designed to resist at high level to linear, differential and high-order differential attacks, whereas nothing has been done to make them inherently resistant to DPA attacks. Countermeasures against DPA attacks are generally added to the algorithms when implemented on devices. Following this addition, the performances and the code sizes of the resulting embedded algorithms are approximately multiplied by two. This increase is damaging in the area of embedded cryptography where the computation power and the memory capability are limited. The design of DPA-resistant algorithms would make the addition of countermeasures innecessary. Such a design could be done by selecting pertinent S-boxes.

For a very particular power consumption model, Guilley *et al.* studied in [9] the *single-bit* DPA attack in terms of *correlation coefficients* between two Boolean signals, the first one depending on linear combinations of output-bits of S-boxes and the second one depending on consumption. The authors pointed out that the better shielded against linear cryptanalysis an S-box is, the more vulnerable it is to side-channel attacks such as DPA. In this paper, we extend the study of Guilley *et al.* for *multi-bit* DPA attacks and for the power consumption model called *Hamming weight model*. We exhibit the properties of S-boxes related to DPA attacks. We argue that these new properties and the classical cryptographic criteria (such as the high non-linearity or the satisfaction of propagation criteria at high level) cannot be satisfied simultaneously. Since a highly non-linear S-box cannot withstand DPA attacks in an optimal way, we point out that a trade-off between the classical cryptographic criteria and resistance to DPA attacks has to be found. We introduce a new cryptographic criteria, that we call *transparency order of an S-box*, to quantify the resistance of an S-box to DPA attacks. We exhibit lower and upper bounds on it and we study their tightness. We prove in particular that bent functions (and more generally functions satisfying $PC(l)$ for a high level l) cannot by definition resist DPA attacks. To ensure the resistance of an algorithm to these attacks, we argue that the new criterion must be satisfied at a certain level and that this level depends on the amount of noise inside the

[1] For example, a DPA of a software DES without any countermeasure requires between 50 and 200 plaintext/ciphertext pairs, whereas the best non-side-channel attack against DES requires under 64 terabytes of plaintexts and ciphertexts encrypted under a single key

device and/or the number of encryptions that a cryptanalyst can do with the same key.

This paper is organized as follows. In Sect. 2, we recall the basic facts about the main cryptographic properties of S-boxes. In Sect. 3, we give the formal definition of an iterated block cipher and we recall the theory behind DPA attacks. To establish the relationship between these attacks and the cryptographic properties of S-boxes, we rewrite in Sect. 4 the DPA attacks in terms of *correlation coefficients*. After arguing that the efficiency of (*single-bit* or *multi-bit*) DPA attacks relies on the behavior of the so-called *differential trace*, we analyze it in Sect. 5. We use this analysis in Sect. 6 to investigate how S-boxes can withstand DPA attacks. In Sect. 7, we introduce and we briefly study the notion of *transparency order of a function*, whose aim is to quantify the resistance of an S-box to DPA attacks.

2 Notation and Preliminaries

In this paper, we distinguish the additions of integers in $\mathbb{R}$, denoted by $+$, and the additions mod 2, denoted by $\oplus$. For simplicity and because there will be no ambiguity, we denote by $+$ the addition of vectors of $\mathbb{F}_2^n$ (words) with $n > 1$.

We call (n,m)-function any mapping F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^m$. If m equals 1, then the function is called Boolean. If F is an affine (n,m)-function, then we call *direction of* F, the linear (n,m)-function L such that there is a vector $B \in \mathbb{F}_2^m$ for which $F(x) = L(x) + B$, $x \in \mathbb{F}_2^n$.

For every vector $a \in \mathbb{F}_2^n$, $n \in \mathbb{N}$, we denote by $H(a)$ the Hamming weight of a. We denote the all-zero vector (resp. the all-one vector) on $\mathbb{F}_2^m$, by 0_m (resp. by 1_m). The set $\{x \in \mathbb{F}_2^n / F(x) \neq 0_m\}$ is called *support* of F: it is denoted by *Supp* F. An (n,m)-function F is said to be *balanced* if every element $y \in \mathbb{F}_2^m$ admits the same number 2^{n-m} of pre-images by F.

To every (n,m)-function F, we associate the m-tuple $(f_1, \cdots, f_m)$ of Boolean functions on $\mathbb{F}_2^n$, called *the coordinate functions of* F, such that we have $F(x) = (f_1(x), \cdots, f_m(x))$ for every $x \in \mathbb{F}_2^n$. The usual scalar product is denoted by "$\cdot$". We recall that it is defined for every pair of vectors $a = (a_1, \cdots, a_m)$ and $b = (b_1, \cdots, b_m)$ by $a \cdot b = \bigoplus_{i=0}^m a_i b_i$.

To make the study of the properties of F easier, we introduce the *sign function* of F, that is the function $(x,v) \mapsto (-1)^{v \cdot F(x)}$ (if F is Boolean, the sign function is the function $x \mapsto (-1)^{F(x)}$). For every (n,m)-function F and for every vector $v \in \mathbb{F}_2^m$, we have:

$$v \cdot F = \frac{1}{2} - \frac{1}{2}(-1)^{v \cdot F} \ . \tag{1}$$

The Fourier transform of the sign function of an (n,m)-function F (that we call *Walsh transform* of F) is the function W_F defined on $\mathbb{F}_2^n \times \mathbb{F}_2^m$ by the formula:

$$W_F(u,v) = \sum_{x \in \mathbb{F}_2^n} (-1)^{v \cdot F(x) + u \cdot x} \ . \tag{2}$$

As we recall in the following proposition, the balancedness of a function can be characterized through its Walsh transform's coefficients.

Proposition 1. *A (n,m)-function F is balanced if and only if $W_F(0,v)$ equals zero for every vector $v \in \mathbb{F}_2^{m*}$.*

Let n be a positive integer and let f and g be two Boolean functions defined on $\mathbb{F}_2^n$, the *correlation coefficient* of f and g, denoted by $\mathrm{Cor}\,(f,g)$, is defined by:

$$Cor(f,g) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+g(x)} \quad . \tag{3}$$

If the output-bits of two Boolean functions are statistically independent, then their correlation coefficient equals zero.

The nonlinearity of a function F is one of the parameters which quantify the level of *confusion* brought in the system by the function (another such parameter is the degree). The nonlinearity of a vectorial function F is defined as the minimum Hamming distance between the nonzero linear combinations of the coordinate functions of F and the set of all Boolean affine functions (that is functions $x \mapsto a \cdot x \oplus b$, $a, b \in \mathbb{F}_2^n$). Cryptographic functions used in block ciphers must have high nonlinearities to prevent linear attacks (see [19]).

For every (n,m)-function F, the nonlinearity N_F and the Walsh transform W_F satisfy the relation $N_F = 2^{n-1} - \frac{1}{2}\max_{u \in \mathbb{F}_2^n, v \in \mathbb{F}_2^{m*}} |W_F(u,v)|$. The nonlinearity N_F of every (n,m)-function F is upper bounded by $2^{n-1} - 2^{n/2-1}$. If n is even and $m \leq \frac{n}{2}$, then this bound is tight. The functions achieving it are called *bent*.

Another useful tool for quantifying the cryptographic resistance of functions is the notion of *derivative*. The derivative of F with respect to a vector $a \in \mathbb{F}_2^n$ is the (n,m)-function $D_aF : x \mapsto F(x) + F(x+a)$. The notion of derivative is related to differential and higher-order differential attacks [2, 13, 17]. A vector $a \in \mathbb{F}_2^n$ such that D_aF is a constant function is called *linear structure* of F. The space $\{a \in \mathbb{F}_2^n;\ D_aF = \mathrm{cst}\}$ is called *linear space* of F and it is denoted by ε_F. As argued by Evertse in [7], the linear spaces of functions used as cryptographic primitives in iterated block ciphers have be reduced to the null vector in order to protect the systems against differential attacks.

Remark 1. Notice that for every (n,m)-function F and for every pair $(a,v) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$, the correlation coefficient between Boolean functions $x \mapsto v \cdot F(x)$ and $x \mapsto v \cdot F(x+a)$ equals $W_{D_aF}(0,v)$. $\diamond$

The *Strict Avalanche Criterion* (*SAC*) was introduced by Webster and Tavares in [30] and this concept was generalized into the *Propagation Criterion* (*PC*) by Preneel [28]. These properties describe the behavior of a function whenever some input coordinates are complemented. They must be satisfied at high levels, in particular by functions involved in block ciphers. A function F *satisfies* $PC(l)$ if the function D_aF is balanced for every vector a of weight at most l. In [29], Rothaus showed that a function is bent if and only if it satisfies $PC(n)$.

In the next section, our aim is to highlight the role that (n,m)-functions play in DPA attacks on block ciphers.

3 DPA Attacks on Iterated Block Ciphers

3.1 Introduction to Iterated Block Ciphers

To define an iterated block cipher in a formal way, we usually consider a family $(F_K)_{K\in\mathcal{K}}$ of (n,n)-functions indexed by a value $K \in \mathcal{K}$, where $\mathcal{K}$ is called the *round key space*. The *encryption function* of the iterated block cipher with block size n, with R rounds and with round functions F_K is defined by:

$$X^{(i)} = F_{K_i}\left(X^{(i-1)}\right) \text{ for } 1 \leq i \leq R, \tag{4}$$

where $X^{(0)}$ is the plaintext and $X^{(R)}$ is the ciphertext.

The vector $(K_1, \ldots, K_R)$ is called the *key* and its coordinates are the *round keys*.

As recalled in Sect. 2, balancedness is a fundamental property which has to be satisfied by every designed round function F_K, $K \in \mathcal{K}$. A classical way to define the balanced functions F_K is to design or select the coordinates functions of each F_K being pairwisely independent. We assume in this paper that the coordinate functions of every round function F_K are pairwisely independent.

3.2 Introduction to Differential Power Analysis

Differential Power Analysis uses the fact that computers and microchips leak information about the operations they process. Specific methods for analyzing the power consumption measurements to find secret keys from *tamper-resistant* devices have been studied in [3, 15, 23]. In what follows, we use notations introduced in [15]. Moreover, we assume that the set $\mathcal{K}$ equals $\mathbb{F}_2^r$, where r is a positive integer.

Let $(F_K)_{K\in\mathbb{F}_2^r}$ be a family of (n,n)-functions used as round functions in an iterated block cipher embedded in a smart card, the power consumption of the smart card after one round of the encryption of a message $X \in \mathbb{F}_2^n$ using a round key $\dot{K} \in \mathbb{F}_2^r$ is usually (*cf.* [3, 15]) denoted by $C_{\dot{K}}(X)$. Function $C_{\dot{K}}$ is called *power consumption function related to* $\dot{K}$ or *power consumption function* if there is no ambiguity on $\dot{K}$.

To describe the DPA attacks, one usually introduces a Boolean function D called *selection function* and defined for every 3-tuple $(X, K, j) \in \mathbb{F}_2^n \times \mathbb{F}_2^r \times \{1, \cdots, n\}$ as the value of the j^{th} bit of $F_K(X)$.

A DPA attack is done by computing a so-called *differential trace* whose values are related to the selection function and to the power consumption function. In what follows, we recall the definition of the differential trace.

Definition 1. *[15] Let $(F_K)_{K\in\mathbb{F}_2^r}$ be a family of permutations on $\mathbb{F}_2^n$ and let D be a selection function related to this family. Let $(X_i)_{i\leq N}$ be a family of N distinct vectors of $\mathbb{F}_2^n$ (randomly chosen if $N < 2^n$). Then, for every pair*

$(K, \dot{K}) \in \mathbb{F}_2^{r\,2}$ *and for every integer* $j \leq n$, *the* differential trace of K with respect to the 3-tuple $(\dot{K}, N, j)$ *is denoted by* $\Delta_{K,\dot{K}}(N, j)$ *and defined by:*

$$\Delta_{K,\dot{K}}(N, j) = \frac{\displaystyle\sum_{i=1}^{N} D(X_i, K, j) C_{\dot{K}}(X_i)}{\sum_{i=1}^{N} D(X_i, K, j)} - \frac{\displaystyle\sum_{i=1}^{N} (1 - D(X_i, K, j)) C_{\dot{K}}(X_i)}{\sum_{i=1}^{N} (1 - D(X_i, K, j))} \ , \qquad (5)$$

where $C_{\dot{K}}$ *is the power consumption function related to* $\dot{K}$.

For large values N, the value $\Delta_{K,\dot{K}}(N, j)$ approximately equals $\Delta_{K,\dot{K}}(2^n, j)$. To simplify notations, we denote $\Delta_{K,\dot{K}}(2^n, j)$ by $\Delta_{K,\dot{K}}(j)$.

In Relation (5), information about the secret parameter $\dot{K}$ is given by the power consumption function $C_{\dot{K}}$. Each value $C_{\dot{K}}(X)$ can be viewed as the energy to flip bits from a previous state to state $F_{\dot{K}}(X)$. To better understand the kind of information this function can give about the round key $\dot{K}$, a theoretical model for the power consumption of devices must be introduced.

In this paper, we use the *Hamming distance model* introduced in [3] as a generalization of the *Hamming weight model* (*cf.* [1]). In the Hamming distance model, it is assumed that switching a bit from 0 to 1 requires the same amount of energy as switching it from 1 to 0. The average power consumption to switch a bit from 0 to 1 is denoted by c and for every pair $(X, K) \in \mathbb{F}_2^n \times \mathbb{F}_2^r$, one denotes by $\alpha(X, K) \in \mathbb{F}_2^n$ the value of the data which is replaced by $F_K(X)$ on the device. We call *state function* function α. For every pair $(X, K) \in \mathbb{F}_2^n \times \mathbb{F}_2^r$, we assume throughout this paper that the power consumption $C_K(X)$ satisfies the relation $C_K(X) = c \times H(\alpha(X, K) + F_K(X)) + w$, where w denotes a noise.

Remark 2. Due to Relation (1), we have:

$$C_K(X) = \frac{nc}{2} - \frac{c}{2} \times \sum_{\substack{u \in \mathbb{F}_2^n \\ H(u)=1}} (-1)^{u \cdot (\alpha(X,K) + F_K(X))} + w \ . \qquad (6)$$

◇

In the following section, we describe DPA attacks more formally and we rewrite the differential trace in terms of correlation coefficients for balanced S-boxes and for constant noise w.

4 DPA Attacks and Correlations

4.1 Single-bit DPA Attacks

One denotes by $\dot{K}$ the first round key used in an iterated block cipher encrypting messages X. We assume in this section that a cryptanalyst wants to retrieve $\dot{K}$ and that he has measured nearly all the values $C_{\dot{K}}(X)$, X ranges over $\mathbb{F}_2^n$.

In the rest of the paper, since we only consider the restriction $\alpha(\cdot, \dot{K})$ of the state function α, we denote by α the function $X \mapsto \alpha(X, \dot{K})$ to simplify notations.

Let $(\dot{K}, N, j) \in \mathbb{F}_2^r \times \{1, \cdots, 2^n\} \times \{1, \cdots, n\}$ be a fixed 3-tuple. In a DPA attack, coefficients $\Delta_{K,\dot{K}}(N, j)$ are computed for different round keys $K \in \mathbb{F}_2^r$ until one value is significantely greater than the others. Let us denote by K_t the corresponding key. The core of the attack is the following: if $\Delta_{K_t,\dot{K}}(N, j)$ is significantely greater than the other values $\Delta_{K,\dot{K}}(N, j)$, $K \in \mathbb{F}_2^r$, then equality $K_t = \dot{K}$ holds with high probability. Since such an attack uses one single bit (of index j) of the outputs $F_{\dot{K}}(X)$, it is usually called *single-bit DPA attack*.

Currently, the main cryptographic properties of S-boxes (nonlinearity, resiliency, balancedness and propagation criteria) are characterized through the Walsh transform. Therefore, to reveal the properties of balanced S-boxes that are related to DPA attacks, we start rewriting the differential trace of a vector in terms of correlation coefficients.

Lemma 1. *Let $(F_K)_{K \in \mathbb{F}_2^r}$ be a family of (n, n)-functions. Let α denote the state function of a cryptographic system implementing F_K, $K \in \mathbb{F}_2^r$, as round functions and let c denote the average power consumption to switch a bit in the system. If all the functions F_K are balanced, then for every pair $(K, \dot{K}) \in {\mathbb{F}_2^r}^2$ and for every positive integer $j \leq n$, we have :*

$$\Delta_{K,\dot{K}}(j) = \frac{c}{2^n} \sum_{\substack{u \in \mathbb{F}_2^n \\ H(u)=1}} \mathrm{Cor}\left(v \cdot F_K, u \cdot (F_{\dot{K}} + \alpha)\right) \ , \tag{7}$$

where $v = (v_1, \cdots, v_n) \in \mathbb{F}_2^n$ is such that $v_j = 1$ and $v_i = 0$ if $i \neq j$.

Proof. By definition of v, we have $D(X, K, j) = v \cdot F_K(X)$, which implies equalities $\sum_{X \in \mathbb{F}_2^n} D(X, K, j) = \#Supp(v \cdot F_K)$ and $\sum_{X \in \mathbb{F}_2^n}(1 - D(X, K, j)) = 2^n - \#Supp(v \cdot F_K)$. Because we assume that every F_K is balanced, it follows that cardinality of $Supp(v \cdot F_K)$ equals 2^{n-1} for every pair $(v, K) \in \mathbb{F}_2^n \times \mathbb{F}_2^r$, $v \neq 0$. Thus, Relation (5) applied for $N = 2^n$ implies the equality $\Delta_{K,\dot{K}}(j) = \frac{-1}{2^{n-1}} \left(\sum_X (1 - 2(v \cdot F_K(X)))\, C_{\dot{K}}(X)\right)$. Using Relation (1), we obtain $\Delta_{K,\dot{K}}(j) = \frac{-1}{2^{n-1}} \sum_X (-1)^{v \cdot F_K(X)} C_{\dot{K}}(X)$. This equality and Relation (6) imply

$$\begin{aligned}\Delta_{K,\dot{K}}(j) = \frac{-nc - 2w}{2^n} & \sum_{X \in \mathbb{F}_2^n} (-1)^{v \cdot F_K(X)} \\ & + \frac{c}{2^n} \sum_{\substack{u \in \mathbb{F}_2^n \\ H(u)=1}} \sum_{X \in \mathbb{F}_2^n} (-1)^{v \cdot F_K(X) + u \cdot (\alpha(X) + F_{\dot{K}}(X))} \ ,\end{aligned} \tag{8}$$

where we recall that w denotes a constant noise. Due to the balancedness of F_K and Proposition 1, the first summation in Relation (8) is null for every non-zero vector v and for every $K \in \mathbb{F}_2^r$. Because the second summation in Relation (8) equals $\frac{c}{2^n} \sum_{\substack{u \in \mathbb{F}_2^n \\ H(u)=1}} \mathrm{Cor}\left(v \cdot F_K, u \cdot (F_{\dot{K}} + \alpha)\right)$, Relations (8) and (7) are equivalent. $\diamond$

More generally a DPA attack can be done by studying correlations between a non-zero linear combination $v \cdot F_K$ and all the coordinate functions of $F_{\dot{K}}$, when

K ranges over $\mathbb{F}_2^r$.. To take this remark into account, we extend Definition 1 by assuming that the differential trace of a vector K is defined with respect to a pair $(\dot{K}, v)$ by:

$$\Delta_{K,\dot{K}}(v) = \frac{c}{2^n} \sum_{\substack{u \in \mathbb{F}_2^n \\ H(u)=1}} \mathrm{Cor}\left(v \cdot F_K, u \cdot (F_{\dot{K}} + \alpha)\right) \quad . \tag{9}$$

In our model, a single-bit DPA attack on the first round of a block cipher is led by designing, for a vector $v \in \mathbb{F}_2^{n*}$, the set of round keys K such that $|\Delta_{K,\dot{K}}(v)|$ is maximal.

4.2 Multi-bit DPA Attacks

Single-bit DPA attacks were generalized in multi-bit DPA attacks in [3,21,22,24, 27]. Among these generalizations, the multi-bit DPA attack proposed by Brier *et al.* in [3] is the most efficient. It is led by searching for high correlations between functions $X \mapsto H(F_K(X))$, $K \in \mathbb{F}_2^r$, and the power consumption function $X \mapsto C_{\dot{K}}(X)$, where $\dot{K}$ is the expected round key. One can prove as in Lemma 1 (and for the same assumptions on $(F_K)_{K \in \mathbb{F}_2^r}$) that multi-bit DPA attack is done by selecting round keys K which maximize the value $\delta_{\dot{K}}(K)$ defined for every pair $(K, \dot{K}) \in \mathbb{F}_2^{r\,2}$ by:

$$\delta_{\dot{K}}(K) = \Big| \sum_{v \in \mathbb{F}_2^n,\ H(v)=1} \Delta_{K,\dot{K}}(v) \Big| \quad . \tag{10}$$

To better understand how the candidate round keys are selected, we study the differential trace in the next section.

5 Analysis of the Differential Trace

Values $\Delta_{K,\dot{K}}(v)$ (and hence $\delta_{\dot{K}}(K)$) are strongly related to the assumptions which are made on the state function α. Indeed, as noticed in [3, 5, 8], if α is supposed to be unknown and dependent on $F_{\dot{K}}$, then the values taken by $(K, v) \mapsto \Delta_{K,\dot{K}}(v)$ cannot be used to get information about the round key $\dot{K}$. Consequently, it is usually assumed either that functions α and F_K are independent for every round key $K \in \mathbb{F}_2^r$, or that α is constant.

5.1 Functions F_K and Function α Are Independent

To prevent statistical attacks, round functions $(F_K)_{K \in \mathbb{F}_2^r}$ of iterated block ciphers are currently designed such that the coordinates of vectors $Y = F_K(X)$, $X \in \mathbb{F}_2^n$, are statistically independent. Moreover, to withstand differential and statistical attacks, the functions in $(F_K)_{K \in \mathbb{F}_2^r}$ are defined to be as uncorrelated as possible. Then, for every pair of distinct elements $(K, \dot{K}) \in \mathbb{F}_2^{r\,2}$ and for every pair $(u, v) \in \mathbb{F}_2^{n*} \times \mathbb{F}_2^{n*}$, $u \neq v$, one can realistically assume in a cryptographic area that $\mathrm{Cor}(v \cdot F_K, u \cdot F_{\dot{K}})$ equals zero (let us notice that in the particular case $u = v$, it cannot be usually assumed that functions $u \cdot F_K$ and $u \cdot F_{\dot{K}}$ are uncorrelated). This assumption is related to the *hypothesis of wrong-key randomization* [10,16]. Under this assumption, we argue in the following proposition that the differential trace has a very simple behavior.

Proposition 2. *Let $(F_K)_{K\in\mathbb{F}_2^r}$ be a family of (n,n)-functions. Let α denote the state function of a cryptographic system implementing functions F_K, $K \in \mathbb{F}_2^r$, as round functions and let c denote the average power consumption to switch a bit in the system. If for every pair $(K,\dot{K}) \in {\mathbb{F}_2^r}^2$ and for every pair of distinct vectors $(u,v) \in {\mathbb{F}_2^n}^2$ s.t. $H(u) = 1$, functions $u \cdot F_K$ and $v \cdot F_{\dot{K}}$ are independent and if α is independent of the round functions F_K for every $K \in \mathbb{F}_2^r$, then for every 3-tuple $(v,K,\dot{K}) \in {\mathbb{F}_2^n}^* \times {\mathbb{F}_2^r}^2$, coefficient $\Delta_{K,\dot{K}}(v)$ equals $\frac{c(-1)^{v\cdot(F_K+F_{\dot{K}})}}{2^n} W_\alpha(0,v)$ if $v \cdot (F_K + F_{\dot{K}})$ is constant and equals 0 otherwise.*

Proof. Because the functions α and F_K are independent for every $K \in \mathbb{F}_2^r$, the correlation coefficient $\mathrm{Cor}(v \cdot F_K \oplus u \cdot F_{\dot{K}}, u \cdot \alpha)$, $(u,v) \in {\mathbb{F}_2^n}^2$, $H(u) = 1$, equals zero if $v \cdot F_K \oplus u \cdot F_{\dot{K}}$ is not constant and equals $\pm W_{u\cdot\alpha}(0)$ if $v \cdot F_K \oplus u \cdot F_{\dot{K}}$ is constant. We assumed that Boolean functions $v \cdot F_K$ and $u \cdot F_{\dot{K}}$ are independent for every pair $(K,\dot{K}) \in {\mathbb{F}_2^r}^2$ and every pair of distinct vectors (u,v) such that $H(u) = 1$. One deduces that if $v \cdot F_K \oplus u \cdot F_{\dot{K}}$ is constant, then u equals v (and $H(v) = 1$). $\diamond$

5.2 Study of $\boldsymbol{\Delta_{K,\dot{K}}}$ when $\boldsymbol{\alpha}$ Is Constant

It is realistic to assume that during the execution of an algorithm embedded in smart cards, state function α is constant. This can be assigned to the so-called *pre-charged logic* where the bus is cleared between each significant transferred value or when the previous operation concerning the bus is an opcode loading (*cf.* [5]). As explained in [3], another possible reason is that complex architectures implement separated busses for data and addresses, that may prohibit certain transitions.

Proposition 2 was established after assuming in particular that functions $v \cdot F_K \oplus u \cdot F_{\dot{K}}$ and $u \cdot \alpha$ are independent for every pair of distinct nonzero vectors $(u,v) \in {\mathbb{F}_2^n}^2$, $H(u) = 1$, and every pair $(K,\dot{K})$. When α is assumed to be constant, this assumption cannot be satisfied. However when α is constant, Relation (9) can be rewritten $\Delta_{K,\dot{K}}(v) = \frac{c}{2^n} \sum_{\substack{u\in\mathbb{F}_2^n \\ H(u)=1}} (-1)^{u\cdot\beta} \mathrm{Cor}\left(v \cdot F_K, u \cdot F_{\dot{K}}\right)$, after denoting by β the constant value of α. Thus, one straightforwardly deduces the following proposition:

Proposition 3. *Let $(F_K)_{K\in\mathbb{F}_2^r}$ be a family of (n,n)-functions. Let α denote the state function of a cryptographic device implementing functions F_K, $K \in \mathbb{F}_2^r$, as round functions and let c denote the average power consumption to switch a bit in the system. Let us assume that functions $v \cdot F_K$ and $u \cdot F_{\dot{K}}$ are independent for every pair $(K,\dot{K}) \in {\mathbb{F}_2^r}^2$ and for every pair of distinct elements $(u,v) \in {\mathbb{F}_2^n}^2$,*

$H(u) = 1$. *If α is constant, equal to $\beta \in \mathbb{F}_2^n$, then for every 3-tuple $(v, K, \dot{K}) \in \mathbb{F}_2^{n*} \times {\mathbb{F}_2^r}^2$, the differential trace of K with respect to $(\dot{K}, v)$ satisfies:*

$$\Delta_{K,\dot{K}}(v) = \frac{c \times (-1)^{v \cdot \beta}}{2^n} \mathrm{Cor}\left(v \cdot F_K, v \cdot F_{\dot{K}}\right) \quad . \tag{11}$$

5.3 Efficiency of the Discrimination of Round Keys in DPA Attacks

Usually, DPA attacks do not permitt to obtain the expected key $\dot{K}$ immediately but allow to isolate it in a subset of $\mathbb{F}_2^r$. For single-bit DPA attacks (resp. multi-bit DPA attacks), the elements of this subset correspond to *ghost peaks* in the distribution of the values of the function $K \in \mathbb{F}_2^r - \{\dot{K}\} \mapsto |\Delta_{K,\dot{K}}(v)|$ (resp. $K \in \mathbb{F}_2^r - \{\dot{K}\} \mapsto |\delta_{\dot{K}}(K)|$). Clearly, the greater the number of ghost peaks, the smaller the efficiency of the attack. Indeed, wrong guesses have to be tested again.

Under assumptions done in Propositions 2 and 3, the set of round keys selected in a single-bit DPA attack with respect to a pair $(v, \dot{K})$ contains the set $\{K \in \mathbb{F}_2^r |\ v \cdot (F_K + F_{\dot{K}}) = \mathrm{cst}\}$. Indeed, when the state function is constant or independent of functions F_K, then the value $|\Delta_{K,\dot{K}}(v)|$ is maximal for every K belonging to $\{K \in \mathbb{F}_2^r |\ v \cdot (F_K + F_{\dot{K}}) = \mathrm{cst}\}$. For multi-bit DPA attacks on a device with random (or null) state function, the set of selected round keys admits the set $\{K \in \mathbb{F}_2^r |\ F_K + F_{\dot{K}} \in \{0_n, 1_n\}\}$ as a subset.

6 Resistance of *S*-Boxes to DPA Attacks when Round Keys Are Introduced by Addition

In many iterated block ciphers such as DES [25] or AES [26], the round key is introduced *by addition*. In this case, we have $r = n$ and, for every round key $K \in \mathbb{F}_2^n$, the round function F_K is the function $X \mapsto F(X + K)$, where F is a robust cryptographic permutation on $\mathbb{F}_2^n$. In such a system we call *S-box* the function F.

In the rest of the paper, we assume that the round keys are introduced by addition. Under this assumption, Propositions 2 and 3 imply the following corollary:

Corollary 1. *Let F be an (n, n)-function whose coordinate functions are pairwisely independent and let α be the state function of a cryptographic device in which F is embedded as an S-box. If α is independent of F or constant, then the number of round keys selected after a single-bit DPA attack with respect to the vector $v \in \mathbb{F}_2^n$ (resp. after a multi-bit DPA attack) is greater than or equal to $\#\varepsilon_{v \cdot F}$ (resp. $\#\varepsilon_F$).*

One cannot withstand multi-bit DPA attacks by increasing the size of the linear space ε_F of F, since the elements of $\dot{K} + \varepsilon_F$ act in a very similar way in the cryptosystem. Indeed, by definition of ε_F, for every element K in $\dot{K} + \varepsilon_F$, there exists a constant $C \in \mathbb{F}_2^n$ such that $X \mapsto F(X+\dot{K})$ and $X \mapsto F(X+K)+C$ are equal.

We showed in Sect. 4.2 that only vectors K such that $\delta_{\dot{K}}(K)$ is maximal have to be stored as good candidate round keys. In practice, because of the imperfections of the measurements (and also because the values of $K \mapsto \Delta_{K,\dot{K}}(N,j)$, $N \leq 2^n$ and $j \leq n$ fixed, are not computed for $N = 2^n$ but for large $N \ll 2^n$), every tested vector such that $\delta_{\dot{K}}(K)$ is significantely high, is stored as a good candidate key (even if $\delta_{\dot{K}}(K)$ is not the maximal value achieved). For this reason, it is difficult to mount an efficient DPA attack when the amplitude of the peaks in the distribution of the values $\delta_{\dot{K}}(K)$, $K \in \mathbb{F}_2^n$, are not high enough (*cf.* [5,6]). Indeed, let us denote by σ the assumed margin of error on the computation of values $\delta_{\dot{K}}(K)$. We argued in Sect. 4 that under some realistic assumptions, the value $\delta_{\dot{K}}(K)$ is always maximal for $K = \dot{K}$. Thus, if the average value

$$D(\dot{K}) = \frac{1}{2^n - 1} \sum_{K \in \mathbb{F}_2^n - \{\dot{K}\}} \Big(\delta_{\dot{K}}(\dot{K}) - \delta_{\dot{K}}(K)\Big) \tag{12}$$

is smaller than σ, then the peaks in the distribution of values $\delta_{\dot{K}}(K)$ could not be identified by an attacker because of the imperfections of the measurements. Reciprocally, if Difference (12) is significantly higher than σ, then the peak corresponding to $\delta_{\dot{K}}(\dot{K})$ will clearly appear in the distribution of values $\delta_{\dot{K}}(K)$ when K ranges over $\mathbb{F}_2^n$.

Let us develop the computation of $D(\dot{K})$ for α independent of F and for α constant.

Lemma 2. *Let F be a (n,n)-function whose coordinate functions are pairwisely independent and let α be the state function of a cryptographic system implementing F as an S-box.*
If α is independent of F, then for every element $\dot{K} \in \mathbb{F}_2^n$ we have :

$$D(\dot{K}) = \frac{c}{2^n} \Big| \sum_{\substack{v \in \mathbb{F}_2^n \\ H(v)=1}} W_\alpha(0,v) \Big| - \frac{1}{2^n - 1} \sum_{K \in \mathbb{F}_2^n - \{\dot{K}\}} \delta_{\dot{K}}(K) \ . \tag{13}$$

If α is constant and equals $\beta \in \mathbb{F}_2^n$, then for every element $\dot{K} \in \mathbb{F}_2^n$ we have :

$$D(\dot{K}) = c|n - 2H(\beta)| - \frac{c}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} \Big| \sum_{\substack{v \in \mathbb{F}_2^n \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0,v) \Big| \ . \tag{14}$$

Proof. Due to Proposition 2, if α and F are independent, then the summations $\sum_{v \in \mathbb{F}_2^n,\ H(v)=1} \Delta_{\dot{K},\dot{K}}(v)$ and $\frac{c}{2^n} \sum_{v \in \mathbb{F}_2^n,\ H(v)=1} W_\alpha(0,v)$ are equivalent: one straightforwardly deduces Relation (13). If the function α equals the constant value β, coefficient $W_\alpha(0,v)$ in Relation (13) equals $(-1)^{v \cdot \beta} \times 2^n$. Moreover, due to Remark 1 and Relation (11), one has

$$\Delta_{K,\dot{K}}(v) = \frac{c}{2^n} \times (-1)^{v \cdot \beta} W_{D_{K+\dot{K}}F}(0,v) \ . \tag{15}$$

From Relations (1), (10), (13) and (15), one deduces Relation (14). $\diamond$

Remark 3.
1. More generally, one can rewritte Relation (14) for (n, m)-functions as:

$$D(\dot{K}) = c|m - 2H(\beta)| - \frac{c}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} | \sum_{\substack{v \in \mathbb{F}_2^n \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0, v)| \ . \quad (16)$$

Moreover, due to Relation (1), summation $\sum_{\substack{v \in \mathbb{F}_2^n \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0, v)$ is also equal to $[n2^n - 2\sum_{X \in \mathbb{F}_2^n} H(\beta + D_a F(X))]$. 2. For every vector $t \in \mathbb{F}_2^n$, let τ_t denotes the function $X \in \mathbb{F}_2^n \mapsto X + t$. Since $W_{D_{K+\dot{K}} F}(0, v)$ equals the function $X \mapsto \mathrm{Cor}(v \cdot F, v \cdot F \circ \tau_{K+\dot{K}})$, Relation (15) relates the differential trace function to the *cross-correlation function* of the coordinate functions of F viewed as binary sequences (see for instance [12] for more details about the cross-correlation function of binary sequences). ◇

As we recalled in Sect. 5.2, it is realistic to assume that during the execution of an algorithm running in a smart card environment, state function α is constant. For such a case, we introduce a new notion, that we call transparency order of a function, to quantify the resistance of an S-box to (single bit or multi-bit) DPA attacks.

7 Transparency Order of S-Boxes

7.1 Definition

Let us assume that the state function is constant. Usually, one cannot presuppose the constant value taken by α, which depends on the implementation. Thus, to thwart DPA attacks on one round of an iterated block cipher, the $D(\dot{K})$ values have to be small enough not only for any round key $\dot{K}$ but also for every possible value β. This remark leads us to introduce a new criterion on S-boxes. In order to be as general as possible, we introduce the notion for (n, m)-functions and not only for permutations on $\mathbb{F}_2^n$.

Definition 2. *Let n and m be two positive integers and let F be an (n, m)-function. The transparency order of F, denoted by $\mathcal{T}_F$, is defined by:*

$$\mathcal{T}_F = \max_{\beta \in \mathbb{F}_2^m} (|m - 2H(\beta)| - \frac{1}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} | \sum_{\substack{v \in \mathbb{F}_2^m \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0, v)|). \quad (17)$$

The smaller the transparency order of an S-box, the higher its resistance to DPA attacks. Indeed, to make the peak corresponding to $\delta_{\dot{K}}(\dot{K})$ undistinguishable from noise of measurements, value $\delta_{\dot{K}}(\dot{K})$ must be approximately equal to the average amplitude $\delta_{\dot{K}}(K)$ when K ranges over $\mathbb{F}_2^n$. Thus, the greatest transparency order that an S-box can achieve without compromising its resistance to DPA attacks depends on the quality of the measurements an attacker can achieve [2].

[2] By adding Hardware's countermeasures to the device, it is possible to ensure a minimal margin of error for any measurement of the power consumption.

7.2 Study of Transparency Order of S-Boxes

In order to determine what a reasonably high transparency order is, there is a need for an upper bound on the transparency order of (n, m)-functions. In what follows, we introduce an upper bound and a lower bound on the transparency order of a function. We show that these bounds can be achieved.

Theorem 1. *Let n and m be two positive integers, transparency order $\mathcal{T}_F$ of every (n, m)-function F satisfies the following relation:*

$$0 \le \mathcal{T}_F \le m \ . \tag{18}$$

If every coordinate function of F is bent, then $\mathcal{T}_F = m$. Moreover, $\mathcal{T}_F$ is null if and only if F is an affine function, whose direction L satisfies $Im(L) \subseteq \{0_m, 1_m\}$.

Remark 4. Since n-variables bent functions only exist for n even, the tightness of the upper bound in Relation (18) is still an open problem for n odd. $\diamond$

Being unbalanced, bent functions are never used as cryptographic primitives. However, due to their properties recalled in Sect. 2 (optimal non-linearity and only balanced non-zero derivatives), they resist in an optimal way to linear and differential cryptanalysis. By showing that bent functions are the weakest possible functions from DPA attacks viewpoint, Theorem 1 establishes that it is impossible to design a function that can resist in an optimal way to linear, differential and DPA attacks. In the following proposition, we show more generally that the functions satisfying $PC(l)$ for a large (but not necessarily optimal) order l do not have a good transparency order.

Proposition 4. *Let m and n be two positive integers such that $m \le n$. Let F be a (n, m)-function. Let $l \le n$ be a positive integer. If F satisfies the $PC(l)$ criteria, then the transparency order of F satisfies:*

$$\mathcal{T}_F \ge m \left(1 - \frac{2^n - \sum_{j=0}^{l} \binom{n}{j}}{2^n - 1}\right) \ . \tag{19}$$

Proof. Because function F satisfies $PC(l)$, then function $D_a F$ is balanced for every vector a s.t. $H(a) \le l$. Due to Proposition 1, one deduces that for every vector a such that $H(a) \le l$ and for every non-zero vector $v \in \mathbb{F}_2^m$, we have $W_{D_a F}(0, v) = 0$. Thus, if F satisfies $PC(l)$, then for every vector $\beta \in \mathbb{F}_2^m$, we have

$$\sum_{a \in \mathbb{F}_2^{n*}} | \sum_{\substack{v \in \mathbb{F}_2^m \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0, v)| = \sum_{\substack{a \in \mathbb{F}_2^{n*} \\ H(a)>l}} | \sum_{\substack{v \in \mathbb{F}_2^m \\ H(v)=1}} (-1)^{v \cdot \beta} W_{D_a F}(0, v)| \ .$$

The cardinality of the set $\{a \in \mathbb{F}_2^n,\ H(a) > l\}$ is $2^n - \sum_{j=0}^{l} \binom{n}{j}$. Moreover, since every value $W_{D_a F}(0, v)$ is lower than or equal to 2^n, then the inequality

$|\sum_{\substack{v\in\mathbb{F}_2^m\\ H(v)=1}}(-1)^{v\cdot\beta}W_{D_aF}(0,v)|\leq m2^n$ is satisfied for every $\beta\in\mathbb{F}_2^m$. One deduces the following relation for every vector $\beta\in\mathbb{F}_2^m$:

$$\sum_{a\in\mathbb{F}_2^{n*}}|\sum_{\substack{v\in\mathbb{F}_2^m\\ H(v)=1}}(-1)^{v\cdot\beta}W_{D_aF}(0,v)|\leq m2^n\left(2^n-\sum_{j=0}^{l}\binom{n}{j}\right)\ . \tag{20}$$

From Relations (17) and (20) and the fact that $\max_\beta|m-2H(\beta)|$ is maximal for $\beta\in\{0_m,1_m\}$, one deduces Inequality (19). $\diamond$

In the next proposition, we investigate the transparency order of affine (n,m)-functions. In particular, we argue that the transparency of an affine function is related to the *weight enumerators* of the *cosets* of $\mathrm{Im}(L)$, where $\mathrm{Im}(L)$ is seen as a *binary linear code.*

Proposition 5. *Let n and m be two positive integers. Let F be an affine (n,m)-function admitting L for direction, then its transparency order satisfies the following relation:*

$$\mathcal{T}_F=\max_{\beta\in\mathbb{F}_2^m}\left(\frac{2^n}{2^n-1}|m-2H(\beta)|-\frac{1}{2^n-1}\sum_{j=0}^{m}|m-2j|\mathcal{N}_{j,\beta}\right)\ , \tag{21}$$

where $\mathcal{N}_{j,\beta}$ denotes the cardinality of the set $\{a\in\mathbb{F}_2^n;\ H(L(a)+\beta)=j\}$.
Moreover, if F is balanced, then its transparency order satisfies:

$$\mathcal{T}_F=\begin{cases}\frac{2^n}{2^n-1}\left(m-\frac{m}{2^m}\binom{m}{\frac{m}{2}}\right) & \text{if } m \text{ is even}\\ \frac{2^n}{2^n-1}\left(m-\frac{2m}{2^m}\binom{m-1}{\frac{m-1}{2}}\right) & \text{if } m \text{ is odd}\end{cases}\ . \tag{22}$$

Remark 5.
1. In Proposition 5, the set $\beta+\mathrm{Im}\ (L)$ can be viewed as a coset of a linear code. Let C denotes this code. If β belongs to C, then $\beta+\mathrm{Im}(L)=\mathrm{Im}(L)$ and values $\mathcal{N}_{j,\beta}$, $j\leq m$, are the coefficients of the *weight enumerator* of C (see for instance [18] for more details about weight enumerators of codes).
2. We recall that for m even and due to *Stirling's formula*, we have $\binom{m}{m/2}\simeq 2^m/\sqrt{\frac{m}{2}\pi}$ for large values of m. Thus for large values m and for balanced affine (n,m)-functions F, the transparency order of F equals approximately $\frac{2^n}{2^n-1}(m-\sqrt{\frac{2m}{\pi}})$ if m is even and to $\frac{2^n}{2^n-1}\left(m-\frac{m}{\sqrt{\frac{(m-1)\pi}{2}}}\right)$ if m is odd. $\diamond$

Due to Proposition 5 and to Remark 5, the transparency order of balanced affine functions is not close to 0 for high values m. Moreover, Relation (21) relates the problem of the construction of affine functions with small transparency order to the problem of defining linear codes whose elements have a Hamming weight either close to 0 or close to m.

8 Conclusion

The study of DPA attacks in terms of correlation coefficients enables us better to understand these attacks. It allows us to characterize the properties of S-boxes related to DPA attacks. To quantify the information leakage of devices involving S-boxes, we introduced the notion of transparency order. We established a spectral characterization of the transparency order of S-boxes and we exhibit its upper and lower bounds. We proved that the lower bound is achieved by particular affine functions and we proved that the transparancy order of bent functions achieves the upper bound. The construction of highly-nonlinear S-boxes with small transparency order (close to 0) is an open problem. The definition of such S-boxes would allow the design of specific block cipher algorithms for smart cards which are less resistant to linear or differential attacks but are inherently resistant to DPA attacks. To make up for this security loss, such algorithms can be implemented in smart cards without the high penalties due to DPA-countermeasures.

References

1. M.-L. Akkar, R. Bévan, P. Dischamp, and D. Moyart. Power Analysis, What is Now Possible. In T. Okamoto, editor, *ASIACRYPT 2000*, volume 1976 of *LNCS*, pages 489–502. Springer, 2000.
2. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
3. E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *CHES 2004*, volume 3156 of *LNCS*, pages 16–29. Springer, 2004.
4. S. Chari, C. Jutla, J. Rao, and P. Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In M. Wiener, editor, *CRYPTO '99*, volume 1666 of *LNCS*, pages 398–412. Springer, 1999.
5. C. Clavier, J.-S. Coron, and N. Dabbous. Differential power analysis in the presence of hardware countermeasures. In Ç. Koç and C. Paar, editors, *CHES 2000*, volume 1965 of *LNCS*, pages 252–263. Springer, 2000.
6. J.-S. Coron, P. Kocher, and D. Naccache. Statistics and secret leakage. In Y. Frankel, editor, *Financial Cryptography – FC 2000*, volume 1962 of *LNCS*. Springer, 2000.
7. J. Evertse. Linear structures in blockciphers. In D. Chaum and W. Price, editors, *EUROCRYPT '87*, volume 304 of *LNCS*, pages 249–266. Springer, 1987.
8. L. Goubin and J. Patarin. DES and Differential Power Analysis – The Duplication Method. In Ç. Koç and C. Paar, editors, *CHES '99*, volume 1717 of *LNCS*, pages 158–172. Springer, 1999.
9. S. Guilley, P. Hoogvorst, and R. Pascalet. Differential power analysis model and some results. In J.-J. Quisquater, P. Paradinas, Y. Deswarte, and A. E. Kalam, editors, *Smart Card Research and Advanced Applications VI – CARDIS 2004*, pages 127–142. Kluwer Academic Publishers, 2004.
10. C. Harpes. Cryptanalysis of iterated block ciphers. In *ETH Series in Information Processing*, volume 7. Hartung-Gorre Verlag, 1996.

11. A. A. Hasan. Power analysis attacks and algorithmic approaches to their countermeasures for Koblitz cryptosystems. In Ç. Koç and C. Paar, editors, *CHES 2000*, volume 1965 of *LNCS*, pages 93–108. Springer, 2000.
12. T. Helleseth and P. V. Kumar. Sequences with low correlation. In *Handbook of coding theory, Vol. II*, pages 1765–1853. North-Holland, 1998.
13. L. Knudsen. Truncated and Higher Order Differentials. In B. Preneel, editor, *Fast Software Encryption – FSE '94*, volume 1008 of *LNCS*, pages 196–211. Springer, 1994.
14. P. Kocher. Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In N. Koblitz, editor, *CRYPTO '96*, volume 1109 of *LNCS*, pages 104–113. Springer, 1996.
15. P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In M. Wiener, editor, *CRYPTO '99*, volume 1666 of *LNCS*, pages 388–397. Springer, 1999.
16. Z. Kukorelly. On the validity of certain hypotheses used in linear cryptanalysis. In *ETH Series in Information Processing*, volume 13. Hartung-Gorre Verlag, 1999.
17. X. Lai. Higher order derivatives and differential cryptanalysis. In *Symposium on Communication, Coding and Cryptography*, 1994. en l'honneur de J.L. Massey à l'occasion de son 60ème anniversaire.
18. F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes*. North-Holland Publishing Co., 1977. North-Holland Mathematical Library, Vol. 16.
19. M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *EUROCRYPT '93*, volume 765 of *LNCS*, pages 386–397. Springer, 1993.
20. R. Mayer Sommer. Smartly Analyzing the Simplicity and the Power of Simple Power Analysis on Smartcards. In Ç. Koç and C. Paar, editors, *CHES 2000*, volume 1965 of *LNCS*, pages 78–92. Springer, 2000.
21. T. Messerges. *Power Analysis Attacks and Countermeasures for Cryptographic Algorithms*. PhD thesis, University of Illinois, 2000.
22. T. Messerges, E. Dabbish, and R. Sloan. Investigations of Power Analysis Attacks on Smartcards. In *the USENIX Workshop on Smartcard Technology (Smartcard '99)*, pages 151–161, 1999.
23. T. Messerges, E. Dabbish, and R. Sloan. Power Analysis Attacks of Modular Exponentiation in Smartcard. In Ç. Koç and C. Paar, editors, *CHES '99*, volume 1717 of *LNCS*, pages 144–157. Springer, 1999.
24. T. Messerges, E. Dabbish, and R. Sloan. Examining Smart-Card Security under the Threat of Power Analysis Attacks. *IEEE Transactions on Computers*, 51(5), May 2002.
25. National Bureau of Standards. *FIPS PUB 46: The Data Encryption Standard*, January 1977.
26. National Institute of Standards and Technology. *FIPS PUB 197: Advanced Encryption Standard*, 2001.
27. E. Oswald. *On Side-Channel Attacks and the Application of Algorithmic Countermeasures*. PhD thesis, Institute for Applied Information Processing and Communications - Graz University of Technology, May 2003.
28. B. Preneel, R. Govaerts, and J. Vandewalle. Boolean functions satisfying higher order propagation criteria. In F. Pichler, editor, *EUROCRYPT '85*, volume 219 of *LNCS*, pages 141–152. Springer, 1985.
29. O. S. Rothaus. On bent functions. In *Journal of Combinatorial Theory*, volume 20a, pages 300–305. Academic Press, 1976.
30. A. Webster and S. Tavares. On the design of S-boxes. In H. Wiliams, editor, *CRYPTO '85*, volume 218 of *LNCS*, pages 523–534. Springer, 1985.

A Proofs of Theorem 1 and of Proposition 5

A.1 Proof of Theorem 1

Proof. The value of $|m-2H(\beta)|$ is upper bounded by m and equals m for $\beta = 0_m, 1_m$. On the other hand, values taken by the summation in Relation (17) belong to $[0;m]$. One straightforwardly deduces Inequality (18).
$\mathcal{T}_F$ equals m if and only if $\beta \in \{0_m, 1_m\}$. In this case, the value of the summation $\sum_{a\in\mathbb{F}_2^{n*}} |\sum_{\substack{v\in\mathbb{F}_2^m\\ H(v)=1}} (-1)^{v\cdot\beta} W_{D_aF}(0,v)|$ is null if and only if the summation $\sum_{\substack{v\in\mathbb{F}_2^m\\ H(v)=1}} W_{D_aF}(0,v)$ is null for every non-zero vector a. On the other hand, if every coordinate function of F is bent, then for every $a\in\mathbb{F}_2^n$ and every $v\in\mathbb{F}_2^m$ such that $H(v)=1$, the function $D_a(v\cdot F)$ is balanced and (due to Proposition 1) satisfies $W_{D_aF}(0,v)=0$. One concludes that such functions F, $\mathcal{T}_F$ is maximal and equals m.
Now, we show that if $\mathcal{T}_F$ is null, then F is an affine function, whose direction L satisfies $\mathrm{Im}(L)\subseteq\{0_m,1_m\}$. By definition, $\mathcal{T}_F$ is greater than or equal to each value

$$|m-2H(\beta)| - \frac{1}{2^n(2^n-1)} \sum_{a\in\mathbb{F}_2^{n*}} | \sum_{v\in\mathbb{F}_2^m, H(v)=1} (-1)^{v\cdot\beta} W_{D_aF}(0,v)| \ ,$$

$\beta\in\mathbb{F}_2^m$, which implies (for $\beta\in\{0_m,1_m\}$):

$$m - \frac{1}{2^n(2^n-1)} \sum_{a\in\mathbb{F}_2^{n*}} | \sum_{v\in\mathbb{F}_2^m,\ H(v)=1} W_{D_aF}(0,v)| \leq \mathcal{T}_F \ . \tag{23}$$

The left-hand side of Relation (23) being always positive or null, if $\mathcal{T}_F$ equals 0, then $m - \frac{1}{2^n(2^n-1)} \sum_{a\in\mathbb{F}_2^{n*}} |\sum_{v\in\mathbb{F}_2^m,\ H(v)=1} W_{D_aF}(0,v)|$ must equal 0, which is equivalent to:

$$\sum_{a\in\mathbb{F}_2^{n*}} | \sum_{v\in\mathbb{F}_2^m,\ H(v)=1} W_{D_aF}(0,v)| = m2^n(2^n-1) \ . \tag{24}$$

Relation (24) is satisfied if and only if $|W_{D_aF}(0,v)|$ equals 2^n for every pair $(a,v)\in\mathbb{F}_2^{n*}\times\mathbb{F}_2^m$, $H(v)=1$, which implies that F is affine. Let L denote the direction of F, then Relation (24) is equivalent to

$$\sum_{a\in\mathbb{F}_2^{n*}} | \sum_{v\in\mathbb{F}_2^m,\ H(v)=1} (-1)^{v\cdot L(a)}| = m(2^n-1) \ ,$$

and the equality holds if and only if $\sum_{v\in\mathbb{F}_2^m,\ H(v)=1}(-1)^{v\cdot L(a)}$ (that is the value $m-2H(L(a)))$ equals $\pm m$ i.e. if and only if $L(a)$ equals 0_m or 1_m. One deduces that if $\mathcal{T}_F$ equals 0, then F is an affine function whose direction L satisfies $\mathrm{Im}(L)\subseteq\{0_m,1_m\}$. Let us prove now that this necessary condition is a sufficient one.

Let F be an affine function whose direction L satisfies $\mathrm{Im}(L) \subseteq \{0_m, 1_m\}$. Then summation $|\sum_{v\in\mathbb{F}_2^n, H(v)=1} W_{D_aF}(0,v)|$ equals $2^n|m-2H(\beta)|$ if $L(a) = 0_m$ and equals $2^n|m-2H(\beta+1_m)|$ if $L(a) = 1_m$. Since one has $|m-2H(\beta+1_m)| = |m-2H(\beta)|$, one deduces the equality

$$\sum_{a\in\mathbb{F}_2^{n*}} |\sum_{v\in\mathbb{F}_2^n, H(v)=1} W_{D_aF}(0,v)| = 2^n(2^n-1)|m-2H(\beta)| \ ,$$

and hence, that $\mathcal{T}_F$ is null. $\diamond$

A.2 Proof of Proposition 5

Before providing proof of Proposition 5, let us first introduce the following technical lemma:

Lemma 3. *For every positive integer m, the following relation is satisfied:*

$$\sum_{j=0}^{m} |m-2j| \binom{m}{j} = \begin{cases} m\binom{m}{\frac{m}{2}} & \text{if } m \text{ is even} \\ 2m\binom{m-1}{\frac{m-1}{2}} & \text{if } m \text{ is odd} \end{cases} . \tag{25}$$

Using Lemma 3, a proof of Proposition 5 is:

Proof. Function L being the direction of F, for every pair $(a,v) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$, coefficient $W_{D_aF}(0,v)$ equals $2^n(-1)^{v\cdot L(a)}$. Thus, for every $\beta \in \mathbb{F}_2^m$, summation $\sum_{\substack{v\in\mathbb{F}_2^m \\ H(v)=1}} (-1)^{v\cdot\beta} W_{D_aF}(0,v)$ equals $2^n\left(m - 2H\left(\beta + L(a)\right)\right)$. Hence, from Relation (17) one deduces:

$$\mathcal{T}_F = \max_{\beta\in\mathbb{F}_2^m} \left(\frac{2^n}{2^n-1} |m-2H(\beta)| - \frac{1}{2^n-1} \sum_{a\in\mathbb{F}_2^n} |m-2H\left(\beta+L(a)\right)| \right) . \tag{26}$$

Because summation $\sum_{a\in\mathbb{F}_2^n} |m-2H\left(\beta+L(a)\right)|$ can be rewritten on the form $\sum_{j=0}^{m} \sum_{a\in\mathbb{F}_2^n, H(\beta+L(a))=j} |m-2j|$, Relation (21) is satisfied.
If F is balanced, then $\mathrm{Im}(L) = \beta + \mathrm{Im}(L) = \mathbb{F}_2^m$ and $\mathcal{N}_{j,\beta}$ equals $2^{n-m} \times \binom{m}{j}$ for every vector β and every integer $j \leq m$. In this case, summation $\sum_{a\in\mathbb{F}_2^n} |m-2H\left(\beta+L(a)\right)|$ equals $2^{n-m}\sum_{j=0}^{m} |m-2j| \binom{m}{j}$. By applying Lemma 3, one deduces that for every balanced affine (n,m)-function, Relations (22) and (26) are equivalent. $\diamond$