What do S-boxes Say in Differential Side Channel Attacks?

Cécile Canovas and Jessy Clédière

CEA-LETI
17 avenue des martyrs, 38 054 Grenoble Cedex 9, France
{cecile.canovas, jessy.clediere}@cea.fr

Abstract. Cryptographic devices are vulnerable against the now well-known side channel leakage analysis. Secret data, such as keys, can be revealed by attacks like DPA, DEMA, CPA. However, this kind of attacks also exhibits wrong keys, this phenomenon being known as the "ghost peaks" problem and has been briefly explained in CPA. We give here a comprehension and analysis of the ghost peak problem that occurs in differential analysis regarding to different power consumption model and various weighting techniques.

Keywords. side channel, differential power analysis, DPA, DEMA, CPA, DES, AES, S-box, correlation, ghost peaks

1 Introduction

Differential analysis on compromising signals have been set up by Kocher et al. [16, 17] on DES algorithm and well detailed by Messerges et al. [19, 21]. Power consumption signals (DPA) of CMOS chips were first used, giving good results to retrieve key values by difference of mean curves selected on defined criteria. Electromagnetic radiation signals (DEMA), acquired by different kinds of sensors, were then successfully used by several authors [24, 13, 25]. Differential analysis has been applied on various cryptographic algorithms, including DES and AES, and several countermeasures have been proposed to secure them from first and high order differential attacks (see for example [14, 10, 3, 4]). We give here an approach and some good results to better understand the influence of S-boxes on differential signals and their leading role in the **ghost peak** phenomenon.

2 Modelling the power consumption

In the past, several power consumption models have been proposed. The first and most simple one is given by Messerges in [20] and is essentially based on the Hamming weight:

$$P(t) = \epsilon.d(t) + L \tag{1}$$

where ϵ and L are real constant numbers and $d(t)$ denotes the Hamming weight of the data being processed at time t. As said by Messerges, this simple model can

be successfully used to understand the behaviour of first and second order DPA attacks. Some other authors presented models based on the Hamming distance, i.e. taking into account the previous state of the internal registers of the CMOS device. Such a model has been proposed by Brier et al.[8] and is given very concisely by:

$$W = a.H(D \oplus R) + b \tag{2}$$

where a and b are real constant numbers, D and R represent respectively the current and previous states processed in the microchip, and $H()$ stands for the Hamming weight. Bevan gives in [6] and [7] a more detailed model:

$$C(t) = \sum_{i=1}^{m} (1 - \alpha_i)\beta_i c_{01}(t) + \alpha_i(1 - \beta_i)c_{10}(t) + Crest(t) \tag{3}$$

where m is the machine size, c_{01} represents the current consumption of a bit to flip from 0 to 1, c_{10} represents the current consumption of a bit to flip from 1 to 0, α_i and β_i the bit value respectively at time $t-1$ and t, and $Crest$ the current consumption independent of the data being processed. These last two models rely on the Hamming distance, which is relevant to represent power consumption of CMOS devices.

All these models may look a little rough at first glance, but many experiments show that they give good indications of what actually happens. To achieve our study, we just simplify the model given in (2) by Brier et al., considering $a = 1$ and $b = 0$:

$$W = H(D \oplus R) \tag{4}$$

With $R = 0$, (4) just simplifies the original Messerges model based on Hamming weight, and as we will see, gives good results to qualify the DES S-boxes in differential analysis.

3 S-boxes

Our approach is independent of the chosen algorithm. Whatever S-boxes can be studied with the following simulation technique. We choose here the well-known DES algorithm ([1, 2]) to explain and illustrate our results, because the obtained tables are remarkable. Moreover, as pointed out by Courtois et al. in [12], the DES algorithm has still many years to live on commercial products.

3.1 Modelling a DES S-box

A DES S-box is modelled, according to (4), by the Hamming distance of its output for a given input. Let's note C six permuted bits of the plaintext or ciphertext, K six bits of the genuine key value, SBox() a DES S-box function which takes six input bits $C \oplus K$ and gives four output bits; let R be the initial value of these four output bits. The compromising signal corresponding to the computation of the considered S-box becomes:

$$W(C) = H(\text{SBox}(C \oplus K) \oplus R) \tag{5}$$

3.2 Weighting, single bit and multi-bits

Kocher et al. introduce in [17] the DPA selection function $D(\mathcal{C}, \mathcal{B}, K_s) \in \{0, 1\}$, where $\mathcal{C}$ is a ciphertext, K_s the guess of six key bits, and $\mathcal{B} \in [0 \dots 31]$ the bit number used for the weighting. The value of the selection function $D(\mathcal{C}, \mathcal{B}, K_s)$ allows to split the curves in two sets. Correlation between the bit $\mathcal{B}$ and the compromising signal takes place with a peak when the difference of the mean of two sets of curves is done. This correlation concept has been greatly extended in many papers, including [9, 18, 11, 8]. However, all these correlations can be rewritten like Kocher's one with an extended selection function $D(\mathcal{C}, K_s)$ and the differential analysis ends up with a difference of the mean of two sets of curves.[1]

Our selection function $D(C, b, K_s)$ is defined for one S-box and a single output bit $b \in [1 \dots 4]$ with six bits C of the cipher or plaintext and a six key guess bits K_s as follow:

$$D(C, b, K_s) = H(\text{SBox}(C \oplus K_s)_b \oplus R_b) \tag{6}$$

where $\text{SBox}()_b$ denotes the bit b value of the considered S-box output, and R_b the bit b value of this S-box output initial state. When $K_s = K$, $C \oplus K_s$ represents the six real input bits of the S-box. This selection function can be easily defined for multi-bits : $D(C, K_s) = H(\text{SBox}(C \oplus K_s) \oplus R)$. With such a selection function, CPA can be rewritten the Kocher's way, extending it with multi-bits and Hamming distance but with no normalisation factors (see footnote 1). Here, for a single bit, (6) reduces to: $D(C, b, K_s) = \text{SBox}(C \oplus K_s)_b \oplus R_b$.

3.3 Differential Analysis by simulation

Let us note $In = C \oplus K$ and $Of = K \oplus K_s$. In represents the six real input bits of the S-box and Of the offset to the real input. With these notations, we now obtain for the modelling:

$$W(C) = H(\text{SBox}(In) \oplus R) \tag{7}$$

and the selection function:

$$D(C, b, K_s) = \text{SBox}(In \oplus Of)_b \oplus R_b \tag{8}$$

In order to simulate the differential analysis, the input of the considered S-box has to be explored by all the possible values, $In \in [0 \dots 63]$. For a given bit b and an offset Of, 64 values C_i, $i \in [1 \dots 64]$, have to contribute to the differential trace defined by Kocher et al. in [17]:

$$\Delta_D = \frac{\sum_{i=1}^{64} D(C_i, b, K_s) W(C_i)}{\sum_{i=1}^{64} D(C_i, b, K_s)} - \frac{\sum_{i=1}^{64} (1 - D(C_i, b, K_s)) W(C_i)}{\sum_{i=1}^{64} (1 - D(C_i, b, K_s))} \tag{9}$$

[1] Some improvement in the differential signal can be realised using normalisation factors, see for example [11, 8]. However, we point out that these normalisation factors, although increasing the signal dynamic, do not change the correlation process, which is still based on a difference of mean curves. This point will be published later.

This theoretical value is reaching when enough curves are available. In general, when we observe that $|\Delta_D| \gg 0$, K_s is the correct key value. Otherwise, $\Delta_D \simeq 0$ and K_s is a wrong guess. This principle allows to learn 6 key bits. This core of attack is the main idea of DPA [17].

4 Hardware versus software

For a software DES, the R value might be the address of a data word or an opcode of an instruction (see [8] for more details). So it is constant for a given implementation and does not depend on the plaintext or ciphertext. For a hardware DES, the R value is the previous state of a register or a wire that holds the output of the S-boxes and can be dependent of In. In order to take into account this dependence of R on In, we would need to know or guess the hardware implementation used. But to ignore R in the selection function is equivalent to a boolean masking, thus the first order differential attack will not be successful.

5 S-boxes properties

In the rest of this paper, we assume that the initial value R is constant. This assumption results in a simplification of (9): the number of input values In, such as $\text{SBox}(In \oplus Of)_b = 1$, is equal to the number of those with $\text{SBox}(In \oplus Of)_b = 0$. So the number of input value i, such as $D(C_i, b, K_s) = \text{SBox}(In \oplus Of)_b \oplus R_b = 1$, is equal to the value number of those with $D(C_i, b, K_s) = 0$. Therefore we obtain:

$$\Delta_D = \frac{1}{32} \sum_{i|D(C_i,b,K_s)=1} W(C_i) - \frac{1}{32} \sum_{i|D(C_i,b,K_s)=0} W(C_i) \tag{10}$$

5.1 Hamming weight model

First we consider the Hamming weight approach ($R = 0$), so we have the following equation:

$$32\Delta_D = \sum_{In|\text{SBox}(In\oplus Of)_b=1} H(\text{SBox}(In)) - \sum_{In|\text{SBox}(In\oplus Of)_b=0} H(\text{SBox}(In)) \tag{11}$$

For example, for a zero offset ($Of = 0\text{x}00$, we guess the real key) $32\Delta_D = 32$. In an ideal world, for a non zero offset ($Of \neq 0\text{x}00$, the key guess is incorrect), the values $H(\text{SBox}(In))$ and $\text{SBox}(In \oplus Of)_b$ are independent. So the probability that $H(\text{SBox}(In)) = X$, knowing $\text{SBox}(In \oplus Of)_b = 1$, should be equal to the probability that $H(\text{SBox}(In)) = X$ knowing $\text{SBox}(In \oplus Of)_b = 0$ and we should have $32\Delta_D = 0$. However S-boxes are not perfect and we can observe that Δ_D sometimes differs from 0. This may lead to wrong detection of the correct key in DPA, referred to as **ghost peaks**. It has often been reported (see for example [6, 22, 8, 15, 5]).

Table 1 gives the result $32\Delta_D$ computed for all offset values Of, for the four bits of S-box 1. These tables correspond to a simulation of the differential analysis for the considered S-box with a one bit weighting. They only depend on

Table 1. $32\Delta_D$ ($R = 0$) as a function of the offset Of = 0x< `Row` >< `Column` > for the four bits of S-box 1

bit 1	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-12	-12	2	6	-2	-8	14	-12	12	8	**-18**	-8	-4	-2	4
1	-14	2	4	8	4	-2	0	2	-6	-2	0	12	6	-4	2	-12
2	-2	10	14	-8	-2	6	2	4	2	-8	-4	-2	-8	4	-2	-6
3	8	-10	**-16**	14	-2	-6	2	-6	-12	12	10	-4	0	4	10	-4

bit 2	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	4	-10	-4	-8	-2	-4	10	-8	4	-4	-4	-12	0	14	-4
1	4	-8	0	10	-4	-6	-2	8	2	6	-12	-8	2	-2	10	-4
2	-12	-2	0	8	10	2	0	-2	-4	6	2	2	10	-6	-10	-12
3	2	-2	10	-4	6	12	-4	4	-8	-8	14	0	-4	2	-12	0

bit 3	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-10	**-18**	2	-4	**20**	10	-4	-14	14	12	**-16**	-6	**-16**	-8	14
1	-14	-2	8	-2	6	**-24**	-12	**16**	10	4	-12	10	-10	**18**	**20**	**-24**
2	-2	8	**16**	6	**22**	**-16**	-10	14	0	**-16**	-10	2	**-20**	14	4	**-16**
3	10	**-16**	**-18**	12	-4	6	-2	-14	-14	**18**	**22**	-14	**16**	-2	-10	14

bit 4	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-6	**-20**	12	**-18**	8	14	0	-14	**22**	0	**-22**	10	**-24**	0	14
1	-8	-4	12	-12	14	**-16**	**-30**	14	2	-8	-4	14	-10	**24**	**20**	**-16**
2	**-24**	8	**22**	**-26**	2	-4	-12	12	14	**-18**	-12	**26**	-6	12	12	-10
3	4	-8	-8	6	-12	**24**	**16**	**-24**	-2	10	6	-2	**24**	**-20**	**-24**	14

S-box values. For the first bit (leftmost) and Of = 0x00, we have $32\Delta_D = 32$, which corresponds to the differential peak for the correct key guess. For other values, $32\Delta_D$ fluctuates from 0 to 18 in absolute value. Value -18 corresponds to a correlation between $H(\text{SBox}(In))$ and $\text{SBox}(In \oplus \text{0x0B})_b$. This correlation gives a peak amplitude greater than one half of the peak value corresponding to the correct key (Of = 0x00). So it is still likely that the correct key K is found by DPA, although the second best results will probably be given by $K \oplus \text{0x0}B$. Second bit of S-box 1 has a smoother property. In Table 1 for bit 2, no peak of amplitude more than 14 is observed. In fact, as it is reported by Brier et al. in [8], this second bit of S-box 1 has a good behaviour in differential side channel analysis. Third and fourth bits of S-box 1 have not at all good properties in

differential analysis since high values (up to 30) appear in the table for non-zero offsets. So it becomes difficult to find the correct key by DPA.

Table 2. $32\Delta_D$ ($R = 0$) as a function of the offset Of = 0x< `Row` >< `Column` > for the four bits of S-box 4

bit 1	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-8	4	4	0	-4	24	-12	12	0	-20	-12	-12	4	-12
1	-12	0	-8	-4	0	0	-4	-24	8	-8	4	12	8	8	0	20
2	-12	12	4	-12	-4	-16	-4	-12	16	-4	0	4	4	0	0	**32**
3	4	-16	0	16	12	12	0	16	-16	-4	0	0	-4	0	0	-28

bit 2	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	0	-4	4	0	-4	-16	-12	4	-16	4	-12	-4	12	12
1	-28	0	0	4	0	0	-4	16	16	0	12	-12	16	0	-16	-4
2	-12	-4	-12	12	-20	0	12	12	24	4	0	-4	4	8	0	**-32**
3	20	0	8	-8	12	-4	-8	-8	-24	4	0	0	-4	8	0	12

bit 3	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-8	-4	4	0	-4	-24	-12	-12	0	20	-12	12	4	12
1	-12	0	-8	4	0	0	-4	24	8	8	4	-12	8	-8	0	-20
2	-12	-12	4	12	-4	16	-4	12	16	4	0	-4	4	0	0	**-32**
3	4	16	0	-16	12	-12	0	-16	-16	4	0	0	-4	0	0	28

bit 4	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	0	4	4	0	-4	16	-12	-4	-16	-4	-12	4	12	-12
1	-28	0	0	-4	0	0	-4	-16	16	0	12	12	16	0	-16	4
2	-12	4	-12	-12	-20	0	12	-12	24	-4	0	4	4	-8	0	**32**
3	20	0	8	8	12	4	-8	8	-24	-4	0	0	-4	-8	0	-12

The other S-boxes have various properties with other ranges for $32\Delta_D$. Moreover, for some S-boxes we may observe very big values like for S-box 7 (see Table 3). The value of $32\Delta_D$ for the third bit of the S-box output and an offset Of = 0x35 is equal to 38, so more than the correct key. Therefore, **the wrong key $K \oplus$ 0x35 will give better DPA results than the correct key K, even with a large number of messages**. This property explains the ghost peak that may appear in the DPA signal for the guess key $K_s = K \oplus$ 0x35. For this offset, the heavy Hamming weights contribute more in the left sum of $32\Delta_D$. This property does not appear for the other output bits of S-box 7. Using another bit for this S-box enables to complete the analysis with success.

This is not the case for S-box 4 (see Table 2) which shows a big value for the offset Of = 0x2F, for all output bits. However, we may still differentiate this

Table 3. $32\Delta_D$ ($R = 0$) as a function of the offset for Of = 0x< `Row` >< `Column` > for the third bit of S-box 7

bit 3	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-6	4	-2	-6	-10	-14	10	-14	14	10	-2	-8	4	0	-8
1	-16	8	-8	14	14	-14	-6	4	12	-10	-12	-12	-10	18	22	-8
2	-10	10	6	-2	-4	-6	8	-16	10	-2	0	20	6	-4	-16	-8
3	-4	-8	-2	-10	-12	**38**	14	-6	4	-6	-4	-2	14	-18	-10	20

incorrect key from the correct one looking at the four tables for each output bit of S-box 4: the polarity of the peaks changes for the wrong key.

5.2 Hamming distance model

Now we consider the initial value R constant and not equal to 0 as an Hamming distance approach. Equation (10) becomes:

$$\begin{aligned} 32\Delta_D = & \sum_{In|\mathrm{SBox}(In\oplus Of)_b\oplus R_b=1} H(\mathrm{SBox}(In)\oplus R) \\ & - \sum_{In|\mathrm{SBox}(In\oplus Of)_b\oplus R_b=0} H(\mathrm{SBox}(In)\oplus R) \end{aligned} \tag{12}$$

The previous tables are computed again with all the possible R values. These new tables also reveal ghost peaks that occur for incorrect key guesses but with different offset values. For example with R = 0x01, the new computed table for the third output bit of S-box 7 (not represented here) does not exhibit a ghost peak at offset value Of = 0x35, because its amplitude has decreased from 38 to 2, but an other ghost peak appears at the offset value Of = 0x3A with amplitude -24. We have previously said that the second bit of S-box 1 has a good behaviour for differential side channel analysis. Although it is not shown here, we note that this property is slightly deteriorated for non-zero initial values.

In the case of S-box 4, whatever value R we choose, we still observe a ghost peak for the same offset value (0x2F). This peak appears for the four output bits of S-box 4, but its polarity changes according to the output bit and the R value. For example, Table 4 with the previous value R = 0x04 shows four ghost peaks which have the same polarity and level than the ones of the correct key.

Some other bits keep their properties for all the possible initial values. To illustrate this point, we take the first bit of S-box 6. It is "the leftmost bit of the fifth S-box" used to illustrate *Fact B.* in [8]. For any initial value R, with offset Of = 0x24 (36 in decimal value in [8]) the obtained amplitude is 24.[2] See for example Table 5 for the initial value R = 0x09.

The second bit of S-box 2 is also noteworthy. For all possible initial values R, this bit gives a peak of amplitude from -26 to -30 for offset Of = 0x20. The

[2] The given ratio 56/64 in [8] does not take into account the weight of the output of the S-box. Some inputs of the S-box, that give the same value for the chosen output bit, do not weight as much.

Table 4. $32\Delta_D$ as a function of the offset Of = 0x< `Row` >< `Column` > for the four bits of S-box 4 and a initial value R = 0x04

bit 1	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-8	4	4	0	-4	16	-12	-4	0	-4	-12	4	4	-12
1	-12	0	-8	-4	0	0	-4	-16	8	0	4	12	8	0	0	4
2	-12	4	4	-12	-4	0	-4	-12	16	-4	0	4	4	-8	0	**32**
3	4	0	0	8	12	4	0	8	-16	-4	0	0	-4	-8	0	-12

bit 2	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-16	4	-20	0	12	-8	-4	-12	16	20	-4	12	-12	-12
1	12	0	-16	-4	0	0	12	8	-8	-16	-12	12	-8	16	16	-20
2	-12	-12	12	-4	20	16	-12	-4	-8	12	0	-20	4	-16	0	**32**
3	-20	16	16	-8	12	-12	-16	-8	8	12	0	0	-4	-16	0	12

bit 3	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-16	4	-20	0	12	-8	-4	-12	16	20	-4	12	-12	-12
1	12	0	-16	-4	0	0	12	8	-8	-16	-12	12	-8	16	16	-20
2	-12	-12	12	-4	20	16	-12	-4	-8	12	0	-20	4	-16	0	**32**
3	-20	16	16	-8	12	-12	-16	-8	8	12	0	0	-4	-16	0	12

bit 4	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	0	-8	4	4	0	-4	16	-12	-4	0	-4	-12	4	4	-12
1	-12	0	-8	-4	0	0	-4	-16	8	0	4	12	8	0	0	4
2	-12	4	4	-12	-4	0	-4	-12	16	-4	0	4	4	-8	0	**32**
3	4	0	0	8	12	4	0	8	-16	-4	0	0	-4	-8	0	-12

same example given in *Fact B.* of [8] but with the second bit of S-box 2 and the shifted input 0x20 is:

100101111100001101100010110010010110100000111100100111111000010110
011010000011110010011111000101101001011111000011011000101100100 1
111111111111111111111101110111111111111111111111111111110111011111

where the first line lists the second output bit values of $\mathrm{SBox}_2(In)$ and the second line lists those of $\mathrm{SBox}_2(In \oplus 0\text{x}20)$. The third line is their bitwise XOR; it contains only four bits of value 0, pointing out that the offset 0x20 gives a good inverse prediction for the key guess. This approach based on output bit correlation, has been developped by Akkar in [5].

5.3 Multi-bits and CPA models

We consider now the correlation factor used by Brier et al. for the key research [8] with an initial value R that is known and constant and the simplified consumption model given by (4):

Table 5. $32\Delta_D$ ($R = 0x09$) as a function of the offset $Of = 0x< \texttt{Row} >< \texttt{Column} >$ for the first bit of S-box 6

bit 1	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-18	-10	2	-12	6	6	2	-20	12	-8	14	10	2	10	-24
1	0	2	-12	8	-16	14	20	-12	12	-20	10	-8	-2	6	-20	14
2	-8	4	12	4	**24**	-18	-14	2	6	0	14	-26	-22	12	-4	18
3	-12	10	24	-14	2	0	-16	8	-6	4	-26	14	12	-16	14	-2

$$\hat{\rho}_{WH}(R) = \frac{N\sum W_i H_{i,R} - \sum W_i \sum H_{i,R}}{\sqrt{N\sum {W_i}^2 - (\sum W_i)^2}\sqrt{N\sum {H_{i,R}}^2 - (\sum H_{i,R})^2}} \tag{13}$$

In the same way, we simulate the correlation power analysis, by exploring all the possible values of the S-box input $In \in [0 \dots 63]$. As the distribution is uniform (13) becomes:

$$\begin{aligned}\hat{\rho}_{WH}(R) &= \frac{64\sum H(\text{SBox}(In \oplus Of) \oplus R)H(\text{SBox}(In) \oplus R) - 128 * 128}{\sqrt{64 * 320 - 128^2}\sqrt{64 * 320 - 128^2}} \\ &= \frac{1}{64}(\sum H(\text{SBox}(In \oplus Of) \oplus R)H(\text{SBox}(In) \oplus R) - 256)\end{aligned} \tag{14}$$

The obtained tables (see for example[3] Table 6) have lower values than those calculted with Δ_D, but they always reveal ghost peaks. For example, for the S-box 4 and an initial value $R = 0x04$, the offset $Of = 0x2F$ gives a peak of the same amplitude as the peak corresponding to the correct key.

Table 6. $\simeq 32\hat{\rho}_{WH}(0x00)$ as a function of the offset $Of = 0x< \texttt{Row} >< \texttt{Column} >$ for the S-box 6

S-box 6	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-13	-4	-3	-8	-1	9	-9	-12	14	0	-5	-6	-1	-8	16
1	**-17**	10	1	9	5	-3	-15	8	7	-12	0	2	3	6	13	**-18**
2	-7	0	15	-15	**18**	-5	-4	-2	-1	-6	-9	**19**	-11	14	-3	-3
3	1	2	-11	7	**-17**	10	3	5	4	5	10	-14	13	**-20**	-1	3

5.4 Location of the weighting

As we have seen, **all the output bits of a DES S-box are not equivalent in differential analysis**. S-boxes are responsible for ghost peaks for incorrect key

[3] We should note that the amplitudes in this table are still multiplied by 32 in order to be comparable with the previous tables.

guesses due to the correlation with the shifted input values. This phenomenon is highly dependent on the consumption model. In order to reduce this phenomenon, we locate the weighting at the output of the XOR of the right and left part of the DES algorithm (this technique was used by Bevan in [6]). Such a one bit weighting can average the incorrect correlation thanks to an extra independent bit of the input plain or ciphertext. With such a weighting, Table 3 is computed again to give Table 7.[4] We remark in this new table that the peak with the amplitude of 38 (Of = 0x35) has been partially averaged and the amplitude lowered.

Table 7. $\simeq 32\Delta_D$ $(R = 0)$ as a function of the offset Of = 0x< `Row` >< `Column` > for the third bit of S-box 7 xor-ed to the left part of the DES

bit 3	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*A*	*B*	*C*	*D*	*E*	*F*
0	**32**	-8	4	4	-12	0	-4	0	**-16**	8	4	-4	0	0	-4	0
1	-12	0	-12	8	12	-4	0	0	4	0	0	-8	-4	4	8	0
2	-12	4	4	-8	4	-4	0	0	12	0	-4	**16**	-4	0	0	-8
3	0	0	4	-4	-4	**8**	4	-4	0	-8	-4	0	4	0	-4	8

However we point out that, even bringing a new bit of the plain or ciphertext, such a weighting still leave some strong correlations. This fact is present in a peak of amplitude 16 in Table 7 and in other values in the 31 remaining tables not represented here.[5] As we can see here, the proposed modelling provides a quick way to verify the properties and pertinence of a weighting.

6 Experimental results

We have previously showed that DES S-box 4 exhibits an unusual behaviour. The correct key guess and the incorrect key guess of offset 0x2F have similar differential signals. To illustrate the soundness of our approach, we show two differential curves in Fig. 1 obtained with the power consumption of a DES software implementation on a smart-card product. The weighting used is an Hamming weight of the second output bit of S-box 4. Top curve of Figure 1 corresponds to the correct key guess, value 0x2D. The first peaks (around 200) correspond to the computation of S-box 4 output, the four following groups of peaks (around 650, 800, 950 and 1050) are the signature of the P permutation inside the f function of DES. Bottom curve of Figure 1 corresponds to the incorrect key guess of value 0x02 = 0x2D $\oplus$ 0x2F. The amplitude of the peaks

[4] For the same reason, the amplitudes in this new table are still multipied by 32, when $\sum_{i|D(C_i,b,K_s)=1} 1 = 512$ for this weighting.

[5] A value of 28 can be found in the table corresponding to the second output bit of S-box 2.

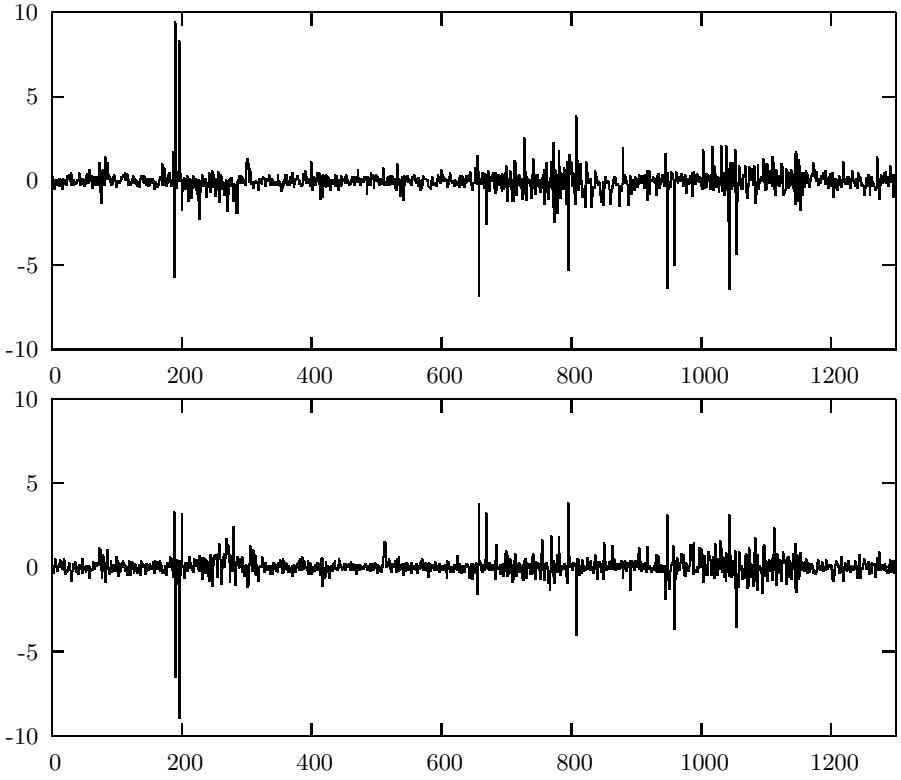

Fig. 1. Differential curves obtained with an Hamming weight on second output bit of S-box 4 for correct key guess 0x2D (top) and for incorrect key guess 0x02 (bottom). Horizontally: time sampling proportional to clock cycles, vertically: 0.01 mA for one graduation

in the two curves is similar and their polarity is opposite. Although it is not shown here, the polarity of the ghost peaks corresponding to the incorrect key guess of value 0x02 changes in accordance to the position of the output bit used for the weighting. Performing the differential analysis separately with the four output bits of S-box 4 enables to eliminate ghost peaks with large amplitude and changing polarity, and at last to identify the correct key guess.

7 Conclusion

The ghost peak phenomenon for incorrect key guesses has been reported by several authors (see for example [6, 22, 8, 15, 5]), but without a clear quantification. **The obtained tables quantify this phenomenon for a given power consumption model**. Values can be given for the amplitude of the peaks corresponding to incorrect keys. Contrarily to a common belief, ghost peaks do not

result from noise in the experiments but **from well understood, non-random behavior of S-boxes**.

From simple simulations on Hamming distance or Hamming weight, we obtained some properties of the DES S-boxes in differential side channel analysis. Some S-boxes are more friendly than others. For example, with an Hamming weight approach, S-box 4 output exhibits a ghost peak for an incorrect key guess with the same amplitude as the peak corresponding to the correct key guess (but with changing polarity). Third output bit of S-box 7 also exhibits a ghost peak for an incorrect key guess with the same polarity and a bigger amplitude than the peak of the correct key guess.

For an initial value $R = 0$, the number of peaks in a table with an amplitude greater than one half of the correct key peak amplitude is on average 8.63 among 64 and varies from 1 to 19 according to the chosen output bit. For other initial values R, the properties of S-boxes are changed. This initial value has a strong impact on the behaviour of a differential side channel attack. For six output bits of four S-boxes, a peak of amplitude over 24 remains whatever initial value R may be. Thanks to these S-Box properties, we can optimise the choice of output bits (single or multi-bits) and the definition of the extended selection function.

Moreover, this work quantifies the ghost peaks that appear due to correlations with shifted value at DES S-boxes input for incorrect key guesses. Other phenomena can also be revealed with this approach, for example the interaction of the adjacent S-boxes due to the E expansion of the DES algorithm.

Introducing imbalance conception effects in the modelling could perform some improvements. For example, real numbers c_{01} and c_{10} indexed with the bit number i (see Bevan's power model in [6] and given by (3)) could increase the soundness of the approach. Such imbalance effects have been proposed, noticed and reported in [26, 15].

This simulation improves the differential analysis by optimising the weighting used for the attack. For single bit weighting, the choice of the bit position in the DES algorithm is essential for the performance of the analysis. For multi-bits weighting, the balance between the bits values can be greatly optimised. This kind of approach is also very useful to quickly ratify the relevance of protected implementations of DES computation before performing experimental tests.

In the future, using a formalism as defined by Prouff in [23], our approach could be used to define S-boxes proof against differential attacks.

8 Acknowledgement

We would like to thanks Raphaël Bauduin, Frédéric Valette, Frédéric Müller, Guillaume Poupard, Jean-Luc Rainard, Pierrick Vignard and Amaël Broustet for very useful discussions.

References

1. Data Encryption Standard (DES). *Federal Information Processing Standards Publication (FIPS PUB) 46, National Bureau of Standards, Washington, DC*, 1977.
2. Data Encryption Standard (DES). *Federal Information Processing Standards Publication (FIPS PUB) 46-3, National Bureau of Standards, Gaithersburg, MD*, `http://csrc.nist.gov/publications/fips/fips46-3/fips46-3.pdf`, 1999.
3. M.L. Akkar, C. Giraud: An Implementation of DES and AES Secure Against Some Attacks. *In proceedings of CHES 2001*, LNCS 2162, pp. 309-318, Springer, 2001.
4. M.L. Akkar, L. Goubin: A Generic Protection Against High-Order Differential Power Analysis. *In proceedings of FSE 2003*, LNCS 2887, pp. 192 - 205, Springer, 2003.
5. M.L. Akkar: Attaques et méthodes de protections de systèmes cryptographiques embarqués. *Rapport de thèse*, 2004.
6. R. Bevan, E. Knudsen: Ways to Enhance DPA. *In proceedings of ICISC 2002*, LNCS 2587, pp.327-342, Springer, 2003.
7. R. Bevan: Estimation statistique et sécurité des cartes à puce, évaluation d'attaques DPA évolués. *Rapport de thèse*, 2004
8. E. Brier, C. Clavier, F. Olivier: Correlation Power Analysis with a Leakage Model, *In proceedings of CHES 2004*, LNCS 3156, pp. 16-29, Springer, 1999.
9. S. Chari, C.S. Jutla, J.R. Rao, P. Rohatgi: Towards Sound Approaches to Counteract Power Analysis Attacks. *In proceedings of CRYPTO 1999*, LNCS 1666, pp. 348-412, Springer, 1999.
10. J.S. Coron, L. Goubin: On Boolean and Arithmetic Masking Against Differential Power Analysis. *In proceedings of CHES 2000*, LNCS 1965, pp. 231-237, Springer.
11. J.S. Coron, P. Kocher, D. Naccache: Statistics and Secret Leakage. *In proceedings of Financial Cryptography*, LNCS 1972, pp. 157-173, Springer, 2000.
12. N.T. Courtois, G. Castagnos, L. Goubin: What do DES S-boxes Say to Each Other? *Cryptology ePrint Archive*, `http://eprint.iacr.org/`, Report 2003/184, 2003.
13. K. Gandolfi, C. Mourtel, F. Olivier: Electromagnetic Attacks: Concrete Results. *In proceedings of CHES 2001*, LNCS 2162, pp. 252-261, Springer, 2001.
14. L. Goubin, J. Patarin: DES and Differential Power Analysis: The Duplication Method. *In proceedings of CHES 1999*, LNCS 1717, pp. 158-172, Springer, 1999.
15. S. Guilley, P. Hoogvorst, R. Pacalet: Differential Power Analysis Model and some Results. *In proceedings of CARDIS 2004*, Kluwer Academic Publishers, pp. 127-142, 2004.
16. P. Kocher, J. Jaffe, B. Jun: Introduction to Differential Power Analysis and related attacks. `http://www.cryptography.com`.
17. P. Kocher, J. Jaffe, B. Jun: Differential Power Analysis. *In proceedings of CRYPTO 1999*, LNCS 1666, pp. 388-397, Springer, 1999.
18. R. Mayer-Sommer: Smartly Analysing the Simplicity and the Power of Simple Power Analysis on Smartcards. *In proceedings of CHES 2000*, LNCS 1965, pp. 78-92, Springer, 2000.

19. T. S. Messerges, E. A. Dabbish, R. H. Sloan: Investigations of Power Analysis Attacks on Smartcards. *In proceedings of the USENIX Workshop on Smart Card Technology 1999*, `http://www.usenix.org/`, 1999.
20. T. S. Messerges: Using Second-Order Power Analysis to Attack DPA Resistant Software. *In proceedings of CHES 2000*, LNCS 1965, pp. 238-251, Springer, 2000.
21. T. S. Messerges, E. A. Dabbish, R. H. Sloan: Examining Smart-Card Security under the Threat of Power Analysis Attacks. *IEEE Transactions on Computers*, Vol. 51, N5, pp. 541-552, May 2002.
22. E. Oswald: On Side-Channel Attack and the application of Algorithmic Countermeasures. *PhD Thesis*, 2003
23. E. Prouff: DPA attacks and S-boxes. *In proceedings of FSE 2005*, LNCS 3557, pp. 424 - 441, Springer, 2005.
24. J.J. Quisquater, D. Samyde: Electromagnetic Analysis (EMA): Measures and Countermeasures for Smart Cards. *In proceedings of E-smart 2001*, LNCS 2140, pp. 200-210, Springer, 2001.
25. J. R. Rao, P. Rohatgi: EMpowering Side-Channel Attacks. *Cryptology ePrint Archive*, `http://eprint.iacr.org/`, Report 2001/037, 2001.
26. J. R. Rao, P. Rohatgi, H. Scherzer, S. Tinguely : Partitioning Attacks : Or How to Rapidly Clone Some GSM Cards. *In proceedings of the 2002 IEEE Symposium on Security and Privacy*, pp. 31-41, IEEE Computer Society, 2002.