Stochastic Side-Channel Leakage Analysis
via Orthonormal Decomposition

Sylvain Guilley[1,2]([✉]), Annelie Heuser[3], Tang Ming[4], and Olivier Rioul[2]

[1] Secure-IC S.A.S., Cesson-Sévigné, France
`sylvain.guilley@secure-ic.com`
[2] Telecom-ParisTech, LTCI, Université Paris-Saclay, Paris, France
[3] CNRS, IRISA, Rennes, France
[4] Wuhan University, Wuhan, China

Abstract. Side-channel attacks of maximal efficiency require an accurate knowledge of the leakage function. Template attacks have been introduced by Chari et al. at CHES 2002 to estimate the leakage function using available training data. Schindler et al. noticed at CHES 2005 that the complexity of profiling could be alleviated if the evaluator has some prior knowledge on the leakage function. The initial idea of Schindler is that an engineer can model the leakage from the structure of the circuit. However, for some thin CMOS technologies or some advanced countermeasures, the engineer intuition might not be sufficient. Therefore, inferring the leakage function based on profiling is still important. In the state-of-the-art, though, the profiling stage is conducted based on a linear regression in a non-orthonormal basis. This does not allow for an easy interpretation because the components are not independent.

In this paper, we present a method to characterize the leakage based on a Walsh-Hadamard orthonormal basis with staggered degrees, which allows for direct interpretations in terms of bits interactions. A straightforward application is the characterization of a class of devices in order to understand their leakage structure. Such information is precious for designers and also for evaluators, who can devise attack bases relevantly.

Keywords: Side-channel analysis · Stochastic attacks · Leakage model · Pseudo-Boolean functions · Orthonormal bases · Leakage characterization

1 Introduction

The existence of side-channels weakens the security of embedded devices, as it allows an attacker to retrieve information about secret keys. The best attacks require the best possible knowledge about the leakage function. A first method in this direction consists of exhaustive characterizations, referred to as *templates* by Chari et al. [5]. Templates are asymptotically perfect estimations of the model, but as pointed out by Schindler [15], they may be inaccurate when there is only a limited amount of profiling traces. Therefore, Schindler has suggested to simplify

the characterization using *stochastic* attacks. While the template method consists in profiling leakage values for all configurations of intermediate variables, which Schindler describes as a projection over a full basis, stochastic attacks consist in characterizing the leakage over a basis of smaller dimensionality.

Leakage characterization does not only benefit to actual attacks. As shown by Kasper et al. [11], it is also a *constructive* feature: when the basis is able to describe the switching activity of the circuit, the estimated weights (basis coefficients) highlight specific exploitable security flaws in the implementation. In their case study, the absolute value of the weight corresponding to one specific bit showed that is was leaking in an outstanding way, and this could be connected to the underlying hardware components (that bit was driving a multiplexer network).

Another motivation is for implementing masking countermeasures. The sensitive data is split into shares which should not interfere physically. Stochastic characterization of the leakage of a *bit pairs* (and in general, of a *bit tuples*) belonging to different shares can reveal flaws in the implementation.

Additionally, stochastic characterization can also benefit to the analysis of unprotected implementations. Recent works showed that, if the *linear basis* describing the switching activity of each bit independently is extended to a *nonlinear basis* which also includes *interactions* between bits, then attacks are more successful in terms of success rate (see e.g., [8,13]). Interestingly, while we know that the consideration of nonlinear bases improves the attack, no sound explanations have been given about what precise information is captured by these nonlinear basis vectors. In [10,13] the authors mention *cross-talk* and *glitch* effects as one possible reason. Up to now, these effects could not be precisely accounted for. One possible reason is that a badly chosen nonlinear basis extension, made with products of bits (i.e., *monomials*), is neither normalized nor orthogonal. As a result, the estimated weights cannot be compared to each other and it seems difficult to draw conclusions about the influence of either individual bits or bit interactions. While the basis *normalization* can be easily carried out (see e.g., [10]), any unstructured *orthogonalization* procedure comes at the expense of the loss of its interpretability in terms of bit interactions, due to the underlying complex change of basis.

Contributions. The goal of this paper is to describe the best possible basis decomposition that is able to isolate leakage from a given coupling of pairs, triples, ..., tuples of bits, independently of the others. We conduct an extensive study of the underlying basis and find a surprisingly simple method to compute the orthonormalized basis. Our method does not only give a feasible solution to interpret the results but it also helps avoid stability problems that occur using standard procedures [16, Sect. 4.2]. The practicability of our methods is tested using simulations and measurements where a leakage is attributed to a tuple of interacting bits.

Outline. The remainder of the paper is organized as follows. Section 2 provides mathematical background for stochastic profiling. Our contribution starts at

Sect. 3, where we derive a novel basis for leakage function decomposition which allows for an easy interpretation in terms of degrees. The method consists in applying a Gram-Schmidt transform on the monomial basis, ordered according to monomial degrees. In Sect. 4 we investigate the leakage estimation in the new basis, together with a fast computation based on the Fourier transform. Practical validation on simulated and real-world traces is shown in Sect. 5. Finally Sect. 6 concludes. Appendix A shows how to estimate projections, and gives an exemple of a "bad" projection into a non-orthogonal basis.

2 Stochastic Profiling

2.1 Leakage Model

Consider a leaking device which manipulates some secret key k. The cryptographic operations involve xoring k with some (plain or cipher) text T. The attacker focuses on manageable parts of the text and key, and T is taken as an n-bit byte (typically $n = 8$). Thus the leakage function f applies to $T \oplus k$ together with some additive noise N, modeled as a normal random variable $N \sim \mathcal{N}(0, \sigma^2)$. The resulting leakage X is given by the equation

$$X = f(T \oplus k) + N. \tag{1}$$

The purpose of this paper is to characterize f which maps the finite set $\mathbb{F}_2^n = \{0,1\}^n$ to the set of real numbers $\mathbb{R}$. A simple example would be the Hamming weight $f = w_H$. Often, f is taken as the composition of some cryptographic function, such as a substitution box $S : \{0,1\}^n \to \{0,1\}^n$, and a leakage function, such as the Hamming weight w_H. This is represented in Fig. 1. In practice, the mapping from $S(T \oplus k) \in \{0,1\}^n$ to $\mathbb{R}$ can be more complex.
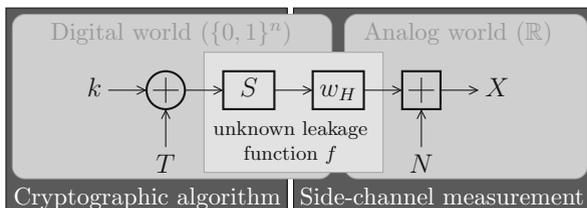

Fig. 1. Setup considered in this paper: f is the unknown

In the following, we consider several independent and identically distributed (i.i.d.) realizations of T, N and X, denoted by $(t_1, \ldots, t_Q) = (t_q)_{q \in \{1, \ldots, Q\}}$, $(n_q)_{q \in \{1, \ldots, Q\}}$ and $(x_q)_{q \in \{1, \ldots, Q\}}$, respectively, where Q denotes the number of queries.

2.2   Notations for Sums and Products

Sum notations will differ depending on whether the considered variables lie in $\mathbb{F}_2^n$ or $\mathbb{R}$. Let $t \in \mathbb{F}_2^n$ be any n-bit vector with bits $t_0, t_1, \ldots, t_{n-1}$. We let $t_i \oplus t_j$ be the exclusive-or addition of bits t_i and t_j in $\mathbb{F}_2$, such that $1 \oplus 1 = 0$, while the usual sum notation $t_i + t_j$ refers to the addition in $\mathbb{R}$, where $1 + 1 = 2$. For the product, there is no such complication. Letting $\wedge$ be the 'and' operator for multiplication in $\mathbb{F}_2$ and $\times$ be the usual multiplicative product in $\mathbb{R}$, we have in fact $t_i \wedge t_j = t_i \times t_j$ for any two bits t_i and t_j in $\{0, 1\}$. Therefore, we will simply denote this product by $t_i t_j$, and use the notation $\prod_{i=0}^{n-1} t_i$ to denote the conjunction of all bits of bit vector t.

2.3   Template and Stochastic Attacks

Template attacks [5] consist in an offline estimation of Eq. (1) for all values t of realizations of T and all choices of the secret key k. This profiling phase is followed by an online application of the maximum likelihood principle to uncover the unknown key. However, template attacks cannot provide an analytic characterization of the leakage. For instance, templates cannot answer the question: "*are bits* 2 *and* 3 *of* T *leaking together?*". We will show in Fig. 4(b) and (c) that our leakage characterization can give a quantitative answer.

While template attacks are *data-driven*, stochastic attacks are *model-driven*: They assume authoritatively that Eq. (1) can be considered to belong to a specific subset of functions $\mathbb{F}_2^n \to \mathbb{R}$. However, the classical approach is to assume some basis for f based on the engineer's intuition. In contract, we aim to find a method to select the most suitable basis for the representation of f.

2.4   Bases and Orthonormality

Let $\mathcal{E}$ be the set of so-called *pseudo-Boolean* [4, Sect. 2.1] functions $\mathbb{F}_2^n \to \mathbb{R}$, which forms a Euclidean vector space over $\mathbb{R}$ of dimension 2^n. The *scalar product* between two vectors f_0 and f_1 in $\mathcal{E}$ is $\langle f_0 | f_1 \rangle = \sum_{t \in \mathbb{F}_2^n} f_0(t) f_1(t)$ and the corresponding *norm* is $||f||_2 = \sqrt{\langle f | f \rangle}$. Any linearly independent family of 2^n vectors $(\psi_u)_{u \in \mathbb{F}_2^n}$ form a *basis* of $\mathcal{E}$. This basis is *orthonormal* if $\langle \psi_u | \psi_v \rangle = 0$ for all $u \neq v$ and $=1$ if $u = v$. In this case an arbitrary pseudo-Boolean function $f \in \mathcal{E}$ can be written as the sum of orthogonal projections

$$f = \sum_{u \in \mathbb{F}_2^n} a_u \psi_u \quad \text{where} \quad a_u = \langle f | \psi_u \rangle \in \mathbb{R} \tag{2}$$

The leakage function $f : \mathbb{F}_2^n \to \mathbb{R}$ is an element of $\mathcal{E}$ that we would like to characterize through a convenient vector basis of $\mathcal{E}$. Two requirements are:

– the basis should somehow relate to bit combinations to make an easy interpretation of the leakage structure in terms of the interactions between bits;
– the basis should be orthonormal so that the characterization of each basis vector is uncorrelated to the other basis vectors.

Appendix A provides an analysis which explains why the use of a non-orthogonal basis is misleading for the interpretation of bit interactions. Appendix A.1 details how coordinates in an orthonormal basis can be estimated with a *correlation method*, and Appendix A.2 shows that the blind application of this method to a non-orthogonal basis yields erroneous results.

2.5 Canonical and Monomial Bases; Degree

The *canonical basis* $(\delta_u)_{u \in \mathbb{F}_2^n}$ of $\mathcal{E}$ is defined by

$$\delta_u(t) = \prod_{i=0}^{n-1} (t_i \oplus u_i) = \begin{cases} 1 & \text{if } t = u, \\ 0 & \text{otherwise,} \end{cases}$$

while the *monomial basis* $(\phi_u)_{u \in \mathbb{F}_2^n}$ of $\mathcal{E}$ is defined by

$$\phi_u(t) = \prod_{i | u_i = 1} t_i = \prod_{i=0}^{n-1} t_i^{u_i}. \tag{3}$$

where the power notation is simply $t_i^0 = 1$ and $t_i^1 = t_i$.

Definition 1 (Degree). *The* degree *of the monomial $\phi_u(t) = \prod_{i=0}^{n-1} t_i^{u_i}$ is the number of coordinates involved in the product, that is, the Hamming weight $w_H(u) = \sum_{i=0}^{n-1} u_i$ of u.*

The degree $\deg(f)$ of any pseudo-Boolean function $f : \mathbb{F}_2^n \to \mathbb{R}$ is the maximum value of the degrees of the monomials ϕ_u in the decomposition of f over the monomial basis.

A function of unit degree is simply a linear combination of bit values, also referred to as Unevenly Weighted Sum of Bits (UWSB) in the side-channel literature [9,17]. A function of degree >1 has *interacting bits* in its decomposition. For example, when the degree is two, product of bits $t_i t_j$ for $i \neq j$ are involved. The degree represents the maximum number of interacting bits.

2.6 Why Canonical and Monomial Bases Are Not Suitable

Properties of the canonical and monomial bases in terms of orthogonality and degree are as follows.

Lemma 2. *The canonical basis is orthonormal, but all vectors have degree n.*

Proof. Clearly $\|\delta_u\| = 1$ and $\langle \delta_u | \delta_v \rangle$ vanishes when $u \neq v$ since the supports of δ_u and δ_v are disjoint. This shows orthonormality. Regarding the degree, we have, for all $t, u \in \mathbb{F}_2^n$:

$$\delta_u(t) = \prod_{i / u_i = 1} t_i \prod_{j / u_j = 0} (1 - t_j).$$

Expending this sum we see that it includes the term $(+1)^{w_H(u)}(-1)^{n - w_H(u)}$ $\phi_{(1,\dots,1)}$, where $(1, \dots, 1)$ is the all-one n-bit vector. Since the latter has Hamming weight equal to n, the corresponding $\phi_{(1,\dots,1)}$, and so δ_u, has degree n. $\qquad\square$

As a consequence, the canonical functions δ_u, albeit simple, are not of practical interest since being all of degree n they are not easily interpretable in terms of "interactions between bits".

On the other hand, the monomial basis is considered in the seminal paper on stochastic side-channel analysis by Schindler *et al.* [15, Eq. (23)], and is customary in side-channel analysis and well understood by engineers because the basis functions have staggered degrees $0, 1, \ldots, n$: While ϕ_0 is the constant 1, the basis vector ϕ_u simply represents the interactions between those bits t_i for which $u_1 = 1$. These basis functions, however, are not even orthogonal:

Lemma 3. *Any monomial basis function ϕ_u has degree equal to $w_H(u) \in \{0, 1, \ldots, n\}$, but the monomial basis is not orthonormal (not even orthogonal):*

$$\langle \phi_u | \phi_v \rangle = 2^{n - w_H(u \vee v)}$$

where $u \vee v$ denotes the bitwise inclusive 'or' of u and v.

Proof. By definition $\deg(\phi_u) = w_H(u)$. We have

$$\langle \phi_u | \phi_v \rangle = \sum_t \phi_u(t)\phi_v(t) = \sum_{t_0, \ldots, t_{n-1}} \prod_{i=0}^{n-1} t_i^{u_i + v_i} \tag{4}$$

$$= \prod_{i=0}^{n-1} \left(\sum_{t_i} t_i^{u_i + v_i} \right) = \prod_{i | u_i = v_i = 0} 2 \tag{5}$$

which is always nonzero.   □

3   Orthonormalizing the Monomial Basis

The monomial basis is ordered by increasing degree (or Hamming weight). For exemple for $n = 3$, the basis vectors are enumerated in the following *weighting order*: $\phi_{(0,0,0)}$, $\phi_{(1,0,0)}$, $\phi_{(0,1,0)}$, $\phi_{(0,0,1)}$, $\phi_{(1,1,0)}$, $\phi_{(1,0,1)}$, $\phi_{(0,1,1)}$ and $\phi_{(1,1,1)}$. Vectors of same weight represent the same number of interacting bits. We proceed to carry out an orthonormalization process that preserves the weight ordering.

3.1   Gram-Schmidt Orthonormalization in Weighting Order

The new orthonormal basis ordered by degree is obtained from the monomial basis by the well-known *Gram-Schmidt orthonormalization*, yielding an orthonormal basis $(\psi_u)_{u \in \mathbb{F}_2^n}$ which can be constrained to *preserve the degree* (as we shall prove in Proposition 4). Algorithm 1 below is Gram-Schmidt procedure operating on vectors ϕ_u with u sorted by weighting order. We write interchangeably $u = (u_0, \ldots, u_{n-1}) \in \mathbb{F}_2^n$ and its equivalent $u = \sum_{i=0}^{n-1} u_i 2^i$ in $\{0, \ldots, 2^n - 1\}$. As the set $\{0, \ldots, 2^n - 1\}$ is totally ordered, this induces the natural *lexicographical* order on $\mathbb{F}_2^n$.

Input   : $(\phi_u)_{u \in \mathbb{F}_2^n}$, a basis of $\mathcal{E}$
Output : $(\psi_u)_{u \in \mathbb{F}_2^n}$, an orthonormal basis of $\mathcal{E}$

`// Creation of the weighting order` ....

1 $W \leftarrow \emptyset$
2 **for** $w = 0$ **to** n **do**
3   **for** $j = 0$ **to** $2^n - 1$ **do**
4   **if** $w_H(j) = w$ **then**
5              $W \leftarrow W \cup \{j\}$

`// Orthonormalization using Gram-Schmidt process` ....

6 **for** $j = 0$ **to** $2^n - 1$ **do**
7       $\xi_{W[j]} \leftarrow \phi_{W[j]} - \sum_{i=0}^{j-1} \frac{\langle \phi_{W[j]} | \xi_{W[i]} \rangle}{\langle \xi_{W[i]} | \xi_{W[i]} \rangle} \xi_{W[i]}$
8       $\psi_{W[i]} \leftarrow \frac{\xi_{W[j]}}{||\xi_{W[j]}||_2}$
9 **return** $(\psi_u)_{u \in \mathbb{F}_2^n}$

Algorithm 1. Gram-Schmidt orthonormalization in weighting order

Proposition 4 (Degree Preservation of the Gram-Schmidt Orthonormalization in Weighting Order). *Let $(\phi_u)_{u \in \mathbb{F}_2^n}$ be a basis of $\mathcal{E}$, such that $\deg(\phi_u) \leq \deg(\phi_v)$ if u is smaller than v with respect to the weighting order (that is $w_H(u) \leq w_H(v)$). Then the Gram-Schmidt orthonormalization process in weighting order (Algorithm 1) applied on $(\phi_u)_{u \in \mathbb{F}_2^n}$ yields a new basis $(\psi_u)_{u \in \mathbb{F}_2^n}$ where $\deg(\psi_u) = \deg(\phi_u)$, for all $u \in \mathbb{F}_2^n$.*

Proof. The weighting order is computed in Algorithm 1 between its lines 1 and 5. It consists in a permutation W of $\{0, \ldots, 2^n - 1\}$, which is such that:

$$\forall j, j' \in \{0, \ldots, 2^n - 1\}, \quad j \leq j' \implies w_H(W[j]) \leq w_H(W[j']). \qquad (6)$$

In Algorithm 1, the first vector fetched from the monomial basis is ϕ_0, which has degree zero. Thus, the degree of $\psi_0 = \phi_0/||\phi_0||_2$ is also zero. Then, by induction on the loop index j (see line 6 of Algorithm 1), we see that the degree of $\psi_{W[j]}$ is equal to that of $\phi_{W[j]}$. Indeed:

– at line 7, we see that $\xi_{W[j]}$ is equal to $\phi_{W[j]}$ minus terms of lower (or equal) degree, owing to the weighting ordering of $W[j]$ (recall Eq. (6));
– at line 8, we see that the degree of $\psi_{W[j]}$ is the same as that of $f_{W[j]}$, because $\psi_{W[j]}$ is the unitary scaling of f_j, operation which keeps the degree unchanged. □

The application of Algorithm 1 on $(\phi_u)_{u \in \mathbb{F}_2^n}$ thus yields a new basis $(\psi_u)_{u \in \mathbb{F}_2^n}$ which meets our requirements: it is orthonormal and ordered by degree.

3.2   Link to Walsh-Hadamard Matrix or Fourier Transform

The Walsh-Hadamard matrices of dimension 2^n for $n \in \mathbb{N}^+$ are given by the recursive formula:

$$H(2^n) = \begin{bmatrix} +H(2^{n-1}) & +H(2^{n-1}) \\ +H(2^{n-1}) & -H(2^{n-1}) \end{bmatrix} \qquad (n > 1)$$

where the lowest order of Walsh-Hadamard matrix is

$$H(2) = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}.$$

A matrix built according to this definition is also referred to as a *lexicographical ordered Walsh-Hadamard matrix*. Walsh-Hadamard matrices are specific square matrices with dimensions of some power of 2, entries of ± 1, and the property that the dot product of any two distinct rows (or columns) is zero.

It is well known that the Walsh-Hadamard matrix H_n is of the form $H_n = 2^{n/2}(\psi_u(t))_{u \in \mathbb{F}_2^n, t \in \mathbb{F}_2^n}$, where u and t are listed in *lexicographical order* (that is, $u \in \mathbb{F}_2^n$ ordered by increasing values of $\sum_{i=0}^{n-1} u_i 2^i$), and where

$$\psi_u(t) = \frac{1}{2^{n/2}}(-1)^{u \cdot t}$$

(where $u \cdot t = \bigoplus_{i=0}^{n-1} u_i t_i$ is the dot product of bitvectors u and t) forms a basis of $\mathcal{E}$ known as the *Fourier basis*.

Theorem 5 (Main Theoretical Result of the Paper). *The basis $(\psi_u)_{u \in \mathbb{F}_2^n}$, obtained by Algorithm 1 from the monomial basis $(\phi_u)_{u \in \mathbb{F}_2^n}$, coincides with the Fourier basis.*

Proof. Let $u \in \mathbb{F}_2^n$. We have that

$$\psi_u(t) = \frac{1}{2^{n/2}}(-1)^{u \cdot t} = \frac{1}{2^{n/2}} \prod_{i=0}^{n-1} (1 - 2t_i)^{u_i}.$$

The development of the product yields a sum of monomials of degrees at most $w_H(u)$. The (only) monomial of degree $w_H(u)$ is $c\phi_u(t)$, where the constant c is equal to $\frac{1}{2^{n/2}}(-2)^{w_H(u)}$. Thus, we have that:

$$\psi_u(t) = c\phi_u(t) - \underbrace{\text{monomials of degree strictly smaller than that of } \psi_u}.$$
$$\text{orthogonal projection of } \phi_u \text{ on } \psi_{u'},$$
$$\text{for each } u' \text{ is smaller than } u \text{ in the weighting order.}$$

This is exactly the procedure of the Gram-Schmidt orthonormalization process in weighting order (line 7 in Algorithm 1). $\square$

Therefore, we have proven that using the Fourier basis $(\psi_u)_{u \in \mathbb{F}_2^n}$ for the projection of the leakage function, the evaluator keeps the mapping between:

– the basis vector $\psi_u : t \mapsto \frac{1}{2^{n/2}}(-1)^{u \cdot t}$, and
– the bit lines which interact (namely, the bits $\{0 \leq i < n, \text{ s.t. } u_i = 1\}$).

Therefore, the leakage can be directly interpreted from the orthonormal projection of the leakage on ψ_u. and the corresponding coefficients a_u of $f : \mathbb{F}_2^n \to \mathbb{R}$ are those on the Fourier basis:

$$f(t) = \sum_u \langle f | \psi_u \rangle \psi_u(t) = \frac{1}{2^{n/2}} \sum_u a_u (-1)^{t \cdot u} \qquad \text{(Eq. (2) in Fourier basis), (7)}$$

which is a *Fourier transform*. The coefficients a_u can be recovered as:

$$a_u = \frac{1}{2^{n/2}} \sum_t f(t)(-1)^{t \cdot u}, \qquad (8)$$

which is the corresponding *inverse Fourier transform*. Notice that direct (Eq. (7)) and inverse (Eq. (8)) Fourier transforms are the same in characteristic two (because $\forall u \in \mathbb{F}_2^n, -u = u$); put differently, the Fourier transform is involutive.
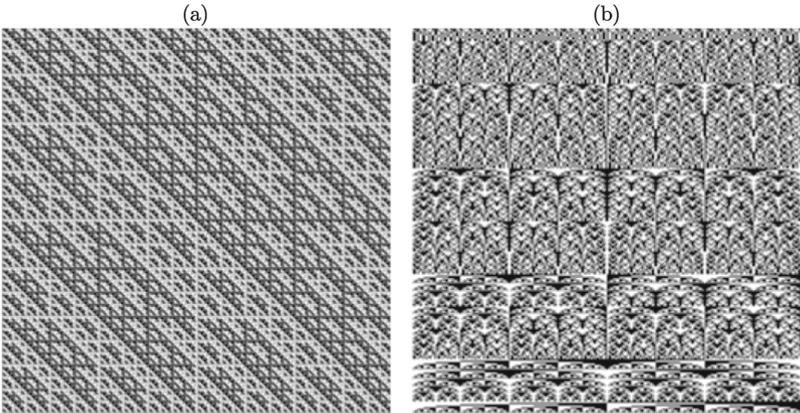

Fig. 2. (a) Walsh-Hadamard 256×256 matrix representation, (b) Truth table of Fourier basis (multiplied by $\sqrt{256} = 16$), in weighting order.

Application to the Case $n = 8$. In the case of byte-oriented block ciphers, such as the AES, the manipulated data are bytes of $n = 8$ bits. The $H(256)$ Walsh-Hadamard matrix is illustrated in Fig. 2(a). Dark pixels are -1 whereas white pixels are $+1$ values. The truth table of the Fourier basis (without the scaling factor of $2^{-n/2}$), represented in weighting order, is depicted in Fig. 2(b). This second matrix is simply the Walsh-Hadamard matrix where lines have been permuted to match the weighting order. One can see that the $H(256)$ matrix is symmetrical. In contrast, the truth table of the Fourier basis is structured as 9 horizontal stripes, comprising 1 (resp. 8, 28, 56, 70, 56, 28, 8 and 1) lines, corresponding to Hamming weight 0 (resp. 1, 2, 3, 4, 5, 6, 7 and 8). It is not immediate visually from Fig. 2(b) that the projection vectors have the same degrees in each "stripe".

3.3   Attribution of Leakage Using the Fourier Basis

Owing to the above properties, the attribution of the leakage using Fourier basis is straightforward:

- build a bitvector $u \in \{0,1\}^n$ where the bits $= 1$ are those we intend to test the interaction in terms of leakage. For instance, to extract the amount of leakage of the Least Significant Bit (LSB), use $u = (1, 0, 0, \ldots, 0)$. Or to test the joint amount of leakage of bits 0 and 1, use $u = (1, 1, 0, \ldots, 0)$;
- compute the projection of the leakage on vector ψ_u (see next section for an estimation method).

4   Estimation of the Projection onto the Fourier Basis

4.1   Exact Solution for the Estimation of the Basis Coefficients

Suppose we have Q leakage values $(x_1, \ldots, x_Q) \in \mathbb{R}^Q$ and let $a = (a_u)_{u \in \mathbb{F}_2^n} \in \mathbb{R}^{2^n}$ be the basis coefficients to be found. Due to the Gaussian nature of the noise, the minimum likelihood determination of a is the following convex optimization problem [10], which happens to be a linear regression problem:

$$\min_{a \in \mathbb{R}^{2^n}} \sum_{q=1}^{Q} \left(x_q - 2^{-n/2} \sum_{u \in \mathbb{F}_2^n} a_u (-1)^{u \cdot (t_q \oplus k)} \right)^2 = \min_{a \in \mathbb{R}^{2^n}} ||x - aG||^2, \qquad (9)$$

where in this case $|| \cdot ||$ is the norm-2 over $\mathbb{R}^Q$, and where G is a $2^n \times Q$ matrix, whose elements are $G[u, q] = 2^{-n/2}(-1)^{u \cdot (t_q \oplus k)}$.

Proposition 6. *The optimal value in Eq. (9) is $a = xG^{\mathsf{T}}(GG^{\mathsf{T}})^{-1}$.*

Proof. This is standard; see [1].

4.2   Fast (Approximate) Solution for the Estimation of $(a_u)_{u \in \mathbb{F}_2^n}$

The expression of Proposition 6 is well known to be a *Moore-Penrose pseudo-inverse*, see e.g. [16, p. 491]. However, it has never been explained in the field of side-channel analysis that the coefficients a_u can be estimated with the following fast formula (in the limit of the low of large numbers), which is an (inverse) *Fourier transform*:

Theorem 7 (Second Main Result of the Paper). *Given Q traces $(x_1, \ldots, x_Q)$ and the Q corresponding texts $(t_1, \ldots, t_Q)$, where the texts are assumed uniformly distributed over $\mathbb{F}_2^n$, the estimation of a_u in the law of large numbers is:*

$$a_u \approx \frac{2^{n/2}}{Q} \sum_{t \in \mathbb{F}_2^n} \left(\sum_{q/t_q = t} x_q \right)(-1)^{u \cdot (t \oplus k)} \quad when \ Q \to \infty. \qquad (10)$$

Proof. Let us notice that xG^T is a vector of length 2^n, whose value at index $u \in \{0,1\}^n$ is $2^{-n/2} \sum_{q=1}^Q x_q (-1)^{u \cdot (t_q \oplus k)}$. Using the reordering of sums put forward in [12], this quantity is also $2^{-n/2} \sum_{t \in \mathbb{F}_2^n} \left(\sum_{q/t_q=t} x_q \right) (-1)^{u \cdot (t \oplus k)}$. Now, assuming that T is uniformly distributed on $\{0,1\}^n$, the $2^n \times 2^n$ matrix GG^T has coefficient at position $(u,v) \in \{0,1\}^n \times \{0,1\}^n$ equal to

$$\frac{2^{-n}}{Q} \sum_{q=1}^Q (-1)^{(u \oplus v) \cdot (t_q \oplus k)} = 2^{-n} \sum_{t \in \mathbb{F}_2^n} \left(\frac{1}{Q} \sum_{q/t_q=t} 1 \right) (-1)^{(u \oplus v) \cdot (t \oplus k)} \xrightarrow[Q \to +\infty]{} \frac{1}{2^n} I_{u,v},$$

by the law of large numbers, where $I_{u,v}$ is the element at position (u,v) in the identity matrix. The limit comes from the fact that $\frac{1}{Q} \sum_{q/t_q=t} 1 \approx \frac{1}{2^n}$ when $Q \to +\infty$, hence the limit using Proposition 7 of [4]. Therefore GG^T is inversed trivially.   □
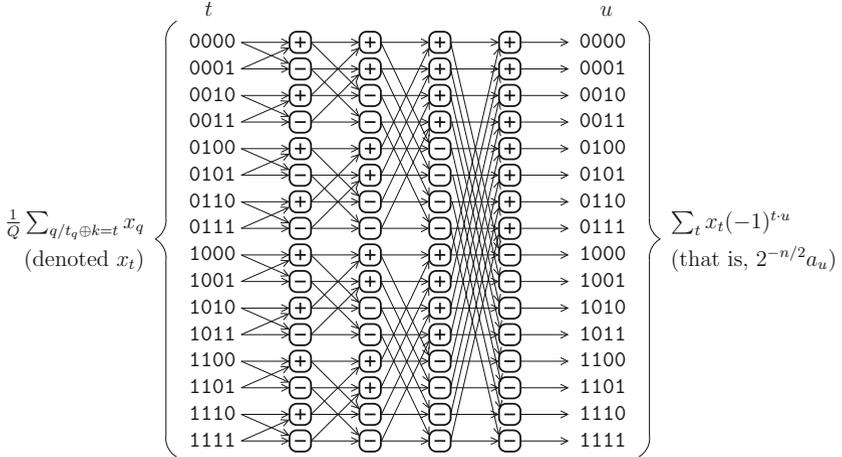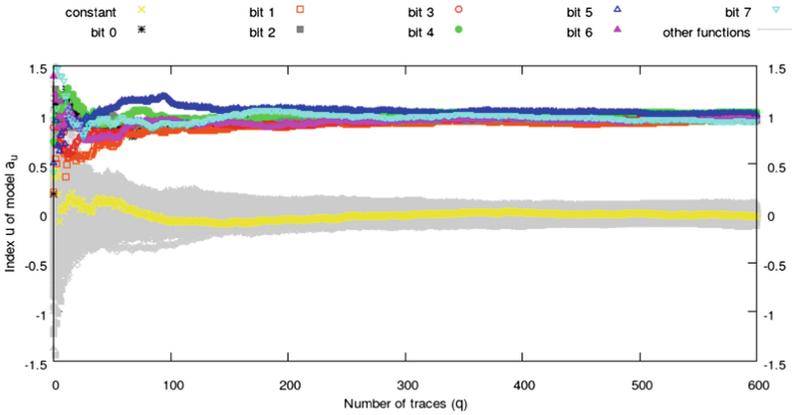

Fig. 3. Butterfly algorithm to compute a_u from the average $\frac{1}{Q} \sum_{q/t_q=t} x_q$ using (10)
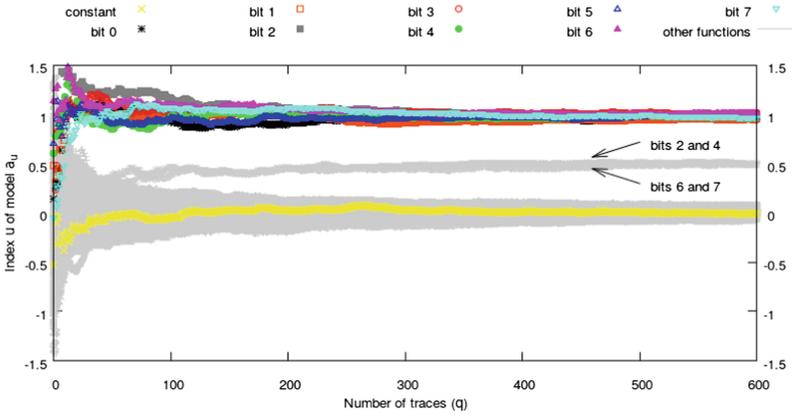
The expression of a_u given in Eq. (10) is (proportional to) the *(inverse) Fourier transform* of the average of leakage traces in each class $(x_q)_{q/q_t=t}$. It is easily computed as follows:

1. sum the traces per value of t, which yields the vector $(\sum_{q/t_q=t} x_q)_{t \in \mathbb{F}_2^n}$,
2. multiply this vector by the Walsh-Hadamard matrix $\frac{2^{n/2}}{Q} H(2^n)$.
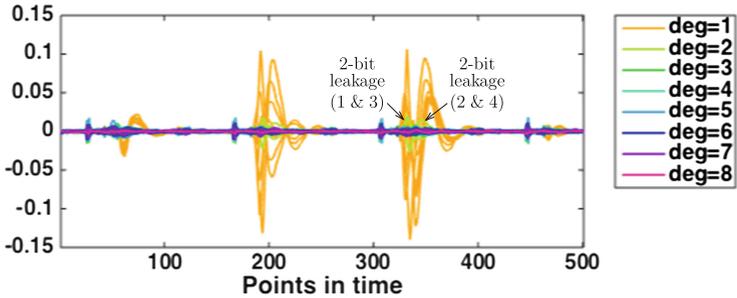
The second step can be optimized with the classical *butterfly* FFT algorithm, which is sketched in Fig. 3 for $n = 4$. Overall, the complexity of the computation of $(a_u)_{u \in \mathbb{F}_2^n}$ from the pairs $(x_q, t_q)_{1 \le q \le Q}$ is $\mathcal{O}(Q + n \cdot 2^n)$.

(a) Centered Hamming weight

(b) Centered Hamming weight with two second-order leakages (of half amplitude)

(c) Dpacontest v4 traces (unprotected scenario) [degree 0 is not represented]

Fig. 4. Estimation of coefficients a_u using Fourier transform

5   Application of the Results

We first consider a simple example from synthetic traces with a linear model and centered Hamming Weight (HW), i.e. $w_H(t) = \frac{n}{2} - \frac{1}{2} \sum_{i=0}^{n-1} (-1)^{t_i}$, and Gaussian noise of variance $\sigma^2 = 2$. Figure 4 shows the coefficients a_u^2 for all $u \in \mathbb{F}_2^n$ and a varying number of profiling traces. One can observe in Fig. 4a that indeed the coefficients are all converging to the same value due to the HW model. Next, we change our model to additionally capture two second order terms, namely $\frac{1}{4}(-1)^{t_2+t_4}$ and $\frac{1}{4}(-1)^{t_6+t_7}$, which are clearly observable in Fig. 4b (in grey). Moreover, these results show that the estimation of a_u is already reasonable stable using only a small number of profiling traces (approximatively 200).

Additionally, we compute a_u^2 for all $u \in \mathbb{F}_2^n$ in the case of almost linear model from real measurement traces. For this purpose, we use the traces from the DPA contest v4 (knowing the mask). Figure 4c shows indeed that in this practical scenario mostly first order coefficients are visible with a minor contribution of second order terms. As these examples show, using our basis we can clearly identify when higher order leakages are present, and directly pinpoint them.

6   Conclusion

In this paper, we have discussed the suitability of "classical" (canonical and monomial) bases for side-channel leakage characterization by stochastic analysis. We show that classical bases are not suitable for this purpose: The canonical basis is of few interest to the evaluator because all elements have maximum degree. The monomial basis, employed in all papers discussing stochastic attacks [6,7,10,11, 14,15] is neither interesting since it is not orthonormal: extracted contributions of bit tuples in the leakage function overlap. Of course, the monomial basis can still be used to attack, since the goal is to extract the key (the linear span of a non-orthogonal basis is equal to that of its orthogonalized basis). By the use of Gram-Schmidt orthonormalization of the monomial basis, we have found that the Fourier basis with vectors ordered in Hamming weight first and lexicographical second is the suitable basis. We explain that leakage characterization can be computed fast using a Fourier transform on partially accumulated traces.

A   Estimations of the Projections

A.1   Estimation of Coordinates in an Orthonormal Basis

We consider a profiling situation where the attacker knows the secret key k, but does not know the model f in Eq. (1). Thanks to an orthonormal basis $(\psi_u)_{u \in \mathbb{F}_2^n}$, the model f can be profiled easily from $(x_q)_{1 \leq q \leq Q}$ measurements, corresponding to $(t_q)_{1 \leq q \leq Q}$ (uniformly distributed) plaintexts.

Lemma 8. *Decompose the unknown function f as $f = \sum_{u \in \mathbb{F}_2^n} a_u \psi_u$, where $a_u = \langle f | \psi_u \rangle$. For every $u \in \mathbb{F}_2^n$, a_u is consistently estimated as $\widehat{a}_u$, the empirical correlation[1] between X and $\psi_u(T \oplus k)$:*

$$\widehat{a}_u = \frac{2^n}{Q} \sum_{q=1}^{Q} x_q \psi_u(t_q \oplus k).$$

Proof. By the law of large numbers,

$$\frac{1}{Q} \sum_{q=1}^{Q} x_q \psi_u(t_q \oplus k) \xrightarrow[Q \to +\infty]{} \mathbb{E}(X \psi_u(T \oplus k)).$$

But from Eq. (1),

$$
\begin{aligned}
\mathbb{E}\left(X \psi_u(T \oplus k)\right) &= \mathbb{E}\left((f(T \oplus k) + N)\psi_u(T \oplus k)\right) \quad &(11)\\
&= \mathbb{E}(f(T \oplus k)\psi_u(T \oplus k)) + \underbrace{\mathbb{E}(N\psi_u(T \oplus k))}_{0}\\
&= \mathbb{E}(f(T \oplus k)\psi_u(T \oplus k))\\
&= \frac{1}{2^n} \sum_{t \in \mathbb{F}_2^n} f(t)\psi_u(t) = \frac{1}{2^n}\langle f | \psi_u \rangle = \frac{1}{2^n} a_u,
\end{aligned}
$$

where the noise term disappeared because N is centered and independent from T, and where the first expectation term is a balanced sum over t because T is uniformly distributed. $\qquad\square$

This theoretical result justifies rigorously why it is customary in the side-channel literature to make use of correlation (or the sibling *covariance* tool) to profile a leakage model [3].

A.2   Incorrect Estimation of Coordinates in a Nonorthogonal Basis

We illustrate in the following example why the monomial basis (though extensively used in the side-channel literature [11,14,15]) is not appropriate for estimating the deterministic part (that is, the function f in Eq. (1)) of the leakage model.

Example 9. Let a leakage function $f : \mathbb{F}_2^n \to \mathbb{R}$, which simply consists in $f(t) = t_0 t_1$. In the understanding of the state-of-the-art, this function models the sole interaction of bits 0 and 1 of bitvector $t = (t_i)_{0 \leq i \leq n-1}$.

We show that the blind application of the above correlation method (Lemma 8) does not allow to recover easily the fact that f consists in the interaction between bits 0 and 1. In fact, letting $u \in \mathbb{F}_2^n$, the correlation between the monomial basis vector ϕ_u and leakage X (Eq. (11)) equals

---

[1] The term *correlation* is used here in the sense of *scalar product* between two data series. This shall not be confused with the *Pearson correlation coefficient* used, for instance, in the *Correlation Power Analysis* [2].

$$a_u = 2^n \mathbb{E}(X\phi_u(T \oplus k)) \tag{12}$$

$$= \sum_{t \in \mathbb{F}_2^n} t_0 t_1 \ \phi_u(t) \qquad \text{(by the change of variable } t \leftarrow t \oplus k)$$

$$= \sum_{t \in \mathbb{F}_2^n} t_0 t_1 \prod_{i/u_i=1} t_i = \sum_{t \in \mathbb{F}_2^n} \prod_{i \in \{0,1\} \cup \{i/u_i=1\}} t_i = 2^{n-2-\sum_{i=2}^{n-1} u_i}$$

$$= \begin{cases} 2^{n-2} & \text{for } u = (0,0,0,\dots,0), (1,0,0,\dots,0), (0,1,0,\dots,0), (1,1,0,\dots,0); \\ 2^{n-3} & \text{for all } u \text{ such that } \sum_{i=2}^{n-1} u_i = 1, e.g., u = (0,0,0,\dots,0,1), \\ & (1,0,0,\dots,0,1), (0,1,0,\dots,0,1), (1,1,0,\dots,0,1), \text{ etc.} \\ \vdots \\ 2 & \text{for } u \text{ such that } \sum_{i=2}^{n-1} u_i = n-3, \text{and} \\ 1 & \text{for } u \text{ such that } \sum_{i=2}^{n-1} u_i = n-2. \end{cases} \tag{13}$$

While the value of a_u is indeed largest for $u = (1,1,0,\dots,0)$ as expected, this maximum value $(=2^{n-2})$ is also reached by $u = (1,0,0,\dots,0)$ and $=(0,1,0,\dots,0)$, which represent single bits. Moreover, there are non-zero terms (albeit smaller) for coefficients a_u such that $w_H(u) > 2$.

Therefore, the covariance method is clearly ill-fitted to characterize that particular leakage function f. The reason for this failure is of course that Lemma 8 is applied in this (counter-)example using the monomial basis $(\phi_u)_{u \in \mathbb{F}_2^n}$, which is not orthonormal.

In summary, we face the problem that the leakage model f cannot be characterized using the *covariance* tool in the monomial basis. This explains why, from Sect. 3 onwards, we investigate a suitable basis, which should have both properties of: (1) being orthonormal (for easy application of the covariance method of Lemma 8) and (2) being interpretable in terms of bits interaction. This will allow to select which vectors of the basis to keep when performing an attack.

References

1. Banerjee, S., Roy, A.: Linear Algebra and Matrix Analysis for Statistics. Texts in Statistical Science, 1st edn. Chapman and Hall/CRC, Hoboken (2014). ISBN 978-1420095388
2. Brier, E., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004). doi:10.1007/978-3-540-28632-5_2
3. Bruneau, N., Danger, J.-L., Guilley, S., Heuser, A., Teglia, Y.: Boosting higher-order correlation attacks by dimensionality reduction. In: Chakraborty, R.S., Matyas, V., Schaumont, P. (eds.) SPACE 2014. LNCS, vol. 8804, pp. 183–200. Springer, Cham (2014). doi:10.1007/978-3-319-12060-7_13
4. Carlet, C.: Boolean functions for cryptography and error correcting codes. In: Crama, Y., Hammer, P. (eds.) Chapter of the Monography Boolean Models and Methods in Mathematics, Computer Science, and Engineering, pp. 257–397. Cambridge University Press (2010)

5. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: Kaliski, B.S., Koç, K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003). doi:10.1007/3-540-36400-5_3

6. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. stochastic methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006). doi:10.1007/11894063_2

7. Heuser, A., Kasper, M., Schindler, W., Stöttinger, M.: How a symmetry metric assists side-channel evaluation - a novel model verification method for power analysis. In: Proceedings of the 14th Euromicro Conference on Digital System Design (DSD 2011), Washington, DC, pp. 674–681. IEEE Computer Society (2011)

8. Heuser, A., Kasper, M., Schindler, W., Stöttinger, M.: A new difference method for side-channel analysis with high-dimensional leakage models. In: Dunkelman, O. (ed.) CT-RSA 2012. LNCS, vol. 7178, pp. 365–382. Springer, Heidelberg (2012). doi:10.1007/978-3-642-27954-6_23

9. Heuser, A., Rioul, O., Guilley, S.: Good is not good enough. In: Batina, L., Robshaw, M. (eds.) CHES 2014. LNCS, vol. 8731, pp. 55–74. Springer, Heidelberg (2014). doi:10.1007/978-3-662-44709-3_4

10. Heuser, A., Schindler, W., Stöttinger, M.: Revealing side-channel issues of complex circuits by enhanced leakage models. In: Rosenstiel, W., Thiele, L. (eds.) DATE, pp. 1179–1184. IEEE (2012)

11. Kasper, M., Schindler, W., Stöttinger, M.: A stochastic method for security evaluation of cryptographic FPGA implementations. In: Bian, J., Zhou, Q., Athanas, P., Ha, Y., Zhao, K. (eds.) FPT, pp. 146–153. IEEE (2010)

12. Lomné, V., Prouff, E., Roche, T.: Behind the scene of side channel attacks. In: Sako, K., Sarkar, P. (eds.) ASIACRYPT 2013. LNCS, vol. 8269, pp. 506–525. Springer, Heidelberg (2013). doi:10.1007/978-3-642-42033-7_26

13. Renauld, M., Standaert, F.-X., Veyrat-Charvillon, N., Kamel, D., Flandre, D.: A formal study of power variability issues and side-channel attacks for nanoscale devices. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 109–128. Springer, Heidelberg (2011). doi:10.1007/978-3-642-20465-4_8

14. Schindler, W.: On the optimization of side-channel attacks by advanced stochastic methods. In: Vaudenay, S. (ed.) PKC 2005. LNCS, vol. 3386, pp. 85–103. Springer, Heidelberg (2005). doi:10.1007/978-3-540-30580-4_7

15. Schindler, W., Lemke, K., Paar, C.: A stochastic model for differential side channel cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005). doi:10.1007/11545262_3

16. Standaert, F.-X., Koeune, F., Schindler, W.: How to compare profiled side-channel attacks? In: Abdalla, M., Pointcheval, D., Fouque, P.-A., Vergnaud, D. (eds.) ACNS 2009. LNCS, vol. 5536, pp. 485–498. Springer, Heidelberg (2009). doi:10.1007/978-3-642-01957-9_30

17. Zhao, H., Zhou, Y., Standaert, F.-X., Zhang, H.: Systematic construction and comprehensive evaluation of kolmogorov-smirnov test based side-channel distinguishers. In: Deng, R.H., Feng, T. (eds.) ISPEC 2013. LNCS, vol. 7863, pp. 336–352. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38033-4_24