Spectral characterization of iterating lossy mappings

Joan Daemen[1,2]

[1] STMicroelectronics
[2] Radboud University

Abstract. In this paper we study what happens to sets when we iteratively apply lossy (round) mappings to them. We describe the information loss as imbalances of parities of intermediate distributions and show that their evolution is governed by the correlation matrices of the mappings. At the macroscopic level we show that iterating lossy mappings results in an increase of a quantity we call *total imbalance*. We quantify the increase in total imbalance as a function of the number of iterations and of round mapping characteristics. At the microscopic level we show that the imbalance of a parity located in some round, dubbed *final*, is the sum of distinct terms. Each of these terms consists of the imbalance of a parity located at the output of a round, multiplied by the sum of the correlation contributions of all linear trails between that parity and the final parity. We illustrate our theory with experimental data. The developed theory can be applied whenever lossy mappings are repeatedly applied to a state. This is the case in many modes of block ciphers and permutations for, e.g., iterated hashing or self-synchronizing stream encryption. The main reason why we have developed it however, is for applying it to study the security implications of using non-uniform threshold schemes as countermeasure against differential power and electromagnetic analysis.

Keywords: iterative lossy mappings, correlation matrices, non-uniformity

1 Introduction

Differential power analysis (DPA) is a class of statistical attacks allowing to extract the key out of cipher implementations exploiting dependence of the power consumption on the data being processed. As a countermeasure to be used in hardware implementations, so-called threshold schemes have been proposed [11,12]. These schemes are a special case of masking schemes, where the sensitive intermediate variables are represented by a number c of shares and the represented value, dubbed *native*, is the (bitwise) sum of those shares. A threshold scheme is designed such that any combinatorial circuit in the implementation takes as input at most $c - 1$ shares. If the sharing is uniform, i.e., if the missing share is uniformly distributed, the power consumption of such a combinatorial circuit is independent of the native value for the same reason that the one-time pad is provably secure. From this it is easy to prove that a threshold scheme is provably secure against first-order DPA as long as the shares are uniform.

We have proposed to apply a 3-share threshold schemes to the Keccak-f permutation [3] to be used for keyed modes of Keccak or Keccak-f itself [2]. However, the threshold sharing we proposed for the non-linear layer is not *uniform*. Concretely, our shared implementation of the non-linear step χ is not invertible and it seems no invertible 3-share threshold scheme exists for χ. This implies that if we start with a uniformly shared state, it is no longer uniform after an iteration. We have proposed different fixes for this problem[4]. In fact, the loss of uniformity can be compensated by some extra circuitry and injecting 4 random bits per round. However, some of us felt that this may be unnecessary. To better understand this, we thought it would be good to take a closer look at this loss of uniformity. The result of these investigations lead to some theory that is not specific for the threshold sharing setting and insights specific for threshold sharing. This paper reports on the former.

Although non-uniformity threshold schemes is the trigger for this work, it can be applied to other settings. For example Merkle-Damgård based or sponge-based hashing, self-synchronizing stream ciphers or ciphers with a non-invertible state-updating function. An example of the latter is the sponge function Gluon[1]. Gluon was already investigated in [14], that can be considered prior art to this work. As opposed to [14] that concentrates on macroscopic aspects, we start from the spectral domain and make extensive use of correlation matrices to derive macroscopic metrics for non-uniformity in a second stage.

1.1 Overview

Section 2 explains how distributions over $\mathrm{GF}(2)^n$ can be fully characterized by the imbalances of their parities. The array of imbalances is called the imbalance spectrum and the link between the probability distribution and the imbalance spectrum is the Walsh-Hadamard transformation. We derive how to compute the spectrum of the product of independent distributions and of a projected distribution.

Section 3 recalls correlation matrices of Boolean mappings and linear trails in iterative mappings. It provides expressions for the occurrence of imbalances in (iterative) Boolean mappings and iterative mappings and their propagation through them. These expressions are the basis for the remainder of the paper.

Section 4 defines macroscopic non-uniformity metrics for distributions and mappings: the total imbalance (contribution) and collision probability. It shows that under independence assumptions, iteratively applying lossy mappings to a variable accumulates the imbalance contributions of the lossy mappings in the total imbalance of the variable.

In Section 5 we characterize the distributions, spectra and total imbalance that result when sampling $\mathrm{GF}(2)^n$ both for the cases with and without replacement and the corresponding distributions of random mappings.

In Section 6 we show that for some classes of mappings, i.e., lossy round functions, it is easy to determine their so-called *collision profile* that fully determines their total imbalance. We illustrate this with an example.

Finally, in Section 7 we provide some experimental evidence that the independence assumptions of Section 4 are reasonable.

1.2 Conventions and notation

We consider distributions over domains of type $\mathrm{GF}(2)^n$, i.e., sets of n-bit vectors. We denote them by a capital, e.g., X. For a given n-bit value x, we denote $\Pr(X = x)$, the probability that $X = x$, by $X(x)$.

We use the Kronecker delta function with a slightly different notation than usual for clarity: $\delta(x = y)$. This function takes two arguments x and y and is 1 if $x = y$ and zero otherwise.

If x is an n-bit vector and y is an m-bit vector $x||y$ denotes the $n+m$-bit vector with first n components those of x and m last components those of y.

For quantities a and b, we use $a \ggg b$ to indicate that a is much larger than b. When using addition and summation, the kind of addition (in $\mathrm{GF}(2)^n$, in $\mathbb{R}$, ...) performed is implicitly determined by the type of summands.

We use vectors and matrices and their products. The vectors are supposed to be column vectors and the transpose operation applied to a vector or matrix switches rows and columns. The transpose of vector v is denoted as v^{T} and the transpose of matrix M is denoted as M^{T}. We denote the $n \times n$ unity matrix by $\mathbf{I_n}$. The component of a vector v with index i is denoted as v_i and the element in a matrix M in row with index r and column with index c is denoted as $M_{r,c}$.

2 Distributions and their (imbalance) spectrum

In this section we show how distributions can be characterized in the spectral domain by means of imbalances in certain parities. Large imbalances can give rise to cryptanalytic or side-channel attacks.

2.1 Parities, imbalances and spectrum

Definition 1. *A distribution X over* $\mathrm{GF}(2)^n$ *is uniform if* $X(x) = 2^{-n}$ *for all* $x \in \mathrm{GF}(2)^n$.

We can describe distributions over $\mathrm{GF}(2)^n$ with *imbalances* over *parities* that are defined by n-dimensional binary vectors called *masks*.

Definition 2. *The parity of a vector x defined by a mask v is the linear function $v^{\mathrm{T}}x$ from* $\mathrm{GF}(2)^n$ *to* $\mathrm{GF}(2)$ *given by*

$$v^{\mathrm{T}}x = \sum_i v_i x_i \ ,$$

where the summation corresponds to the addition in $\mathrm{GF}(2)$.

Definition 3. *The imbalance $\widetilde{X}(v)$ of a mask v for a distribution X is given by*

$$\widetilde{X}(v) = \sum_x X(x)(-1)^{v^{\mathrm{T}}x} \ . \tag{1}$$

Imbalances range between -1 (parity is always 1) and $+1$ (parity is always 0). If it is zero we say it is *balanced*.

Filling in $v = 0$ in Equation (1) yields $\widetilde{X}(0) = 1$. Naturally, $\widetilde{X}(0)$ is the imbalance of the constant function zero and so equal to 1. This leads us to the following definition.

Definition 4. *The spectrum of a distribution with $\widetilde{X}(0)$ omitted is the* reduced *spectrum and denoted by $\widehat{X}$.*

Let $\mathcal{L}$ be a mapping from the space of binary vectors to the space of real-valued vector that transforms a binary vector of dimension n to a real-valued vector of dimension 2^n. $\mathcal{L}$ is defined by

$$\mathcal{L} : \mathrm{GF}(2)^n \to \mathbb{R}^{2^n} : a \mapsto \mathcal{L}(x) \Leftrightarrow \forall u \in \mathrm{GF}(2)^n : \mathcal{L}(x)_u = (-1)^{u^{\mathrm{T}}x} \ . \tag{2}$$

Since $\mathcal{L}(x \oplus y) = \mathcal{L}(x) \cdot \mathcal{L}(y)$, $\mathcal{L}$ is a group homomorphism from $\langle \mathrm{GF}(2)^n, + \rangle$ to $\langle (\mathbb{R} \backslash \{0\})^{2^n}, \cdot \rangle$, where '$\cdot$' denotes the component-wise product.

$\mathcal{L}(x)$ contains the 2^n parities of an n-bit vector x. Equivalently, it contains the parities of the distribution X over $\mathrm{GF}(2)^n$ that has probability 1 in x and zero elsewhere: $\Pr(X = x) = \delta(x = a)$. We can express the spectrum of a distribution X in terms of $\mathcal{L}$:

$$\widetilde{X} = \sum_x X(x)\mathcal{L}(x) \ . \tag{3}$$

2.2 The Walsh-Hadamard transform

From Equation (1), it is clear that the vector $\widetilde{X}$ of values $\widetilde{X}(v)$ for all v can be obtained by applying the Walsh-Hadamard transform [8] to X. This transform is a linear transformation operating on a vector space $\mathbb{R}^{2^n}$ that can be modelled as multiplication by a square matrix $\mathcal{W}$ with 2^n rows and columns. The rows and columns are not indexed by integers but rather by n-bit binary vectors and the element in row v and column x is given by $(-1)^{v^{\mathrm{T}}x} = \mathcal{L}(x)_v$. So we have $\widetilde{X} = \mathcal{W} \times X$. Clearly, $v^{\mathrm{T}}x = x^{\mathrm{T}}v$ so $\mathcal{W}$ is symmetric: $\mathcal{W}^{\mathrm{T}} = \mathcal{W}$.

We can define an inner product $\langle A, B\rangle$ with A and B vectors in $\mathbb{R}^{2^n}$. Assuming A and B are column arrays containing coordinates with respect to an orthonormal basis, this inner products is given by $\langle A, B\rangle = A^{\mathrm{T}}B = \sum_i A_i B_i$. Two vectors A and B are orthogonal if their inner product is zero.

A transformation $\mathcal{M}$ is said to be orthogonal if for all vectors A and B it holds that $\langle \mathcal{M}A, \mathcal{M}B\rangle = \langle A, B\rangle$. It is easy to see that this is the case if the columns of $\mathcal{M}$ form an orthonormal basis, i.e., if we denote two columns of $\mathcal{M}$ by M_i and M_j, we have ${M_i}^{\mathrm{T}}M_j = \delta(i = j)$. This can be expressed more compactly as $\mathcal{M}^{\mathrm{T}}\mathcal{M} = \mathbf{I_{2^n}}$.

The Walsh-Hadamard transform can be decomposed in an orthogonal transformation and an expansion by $2^{n/2}$. We have $\mathcal{W} = 2^{n/2}\check{\mathcal{W}}$ and $\check{\mathcal{W}}\check{\mathcal{W}}^{\mathrm{T}} = \mathbf{I_{2^n}}$. The inverse of $\mathcal{W}$ is therefore given by $\mathcal{W}^{-1} = 2^{-n}\mathcal{W}^{\mathrm{T}} = 2^{-n}\mathcal{W}$. It follows that we can reconstruct a distribution X from its spectrum $\widetilde{X}$ in the following way:

$$X(x) = 2^{-n}\sum_v \widetilde{X}(v)(-1)^{v^{\mathrm{T}}x} \ , \tag{4}$$

or equivalently

$$X = 2^{-n}\sum_v \widetilde{X}(v)\mathcal{L}(v) \ .$$

2.3 Product of independent distributions

Let X be a distribution of a 2^n-bit string x and Y a distribution of a 2^m bit string y, with X and Y independent and let z be the joint distribution of x and y. Then the distribution Z of z is given by:

$$Z(z = (x, y)) = X(x)Y(y) \ .$$

For the imbalances this implies the following:

$$\widetilde{Z}(v = (v_x, v_y)) = \widetilde{X}(v_x)\widetilde{Y}(v_y) \ .$$

This can be generalized to the concatenation of s string with independent distributions. Let $x = (x_{(0)}, x_{(1)}, \ldots, x_{(s-1)})$ and $v = (v_{(0)}, v_{(1)}, \ldots, v_{(s-1)})$. We have:

$$\widetilde{X}(v) = \prod_i \widetilde{X_{(i)}}(v_{(i)}) \ . \tag{5}$$

Note that in the product on the right hand side of Equation (5), only factors with $v_{(i)} \neq 0$ can be different from 1. We call these *active* component masks. Moreover, for $\widetilde{X}(v)$ to be non-zero, all terms in the product on the right hand side shall be different from zero. In words, for $v^{\mathrm{T}}x$ to be imbalanced, all parities ${v_{(i)}}^{\mathrm{T}}x_{(i)}$ must be imbalanced. This implies that $\widetilde{X}(v) = 0$ as soon as there is a single parity ${v_{(i)}}^{\mathrm{T}}x_{(i)}$ that is balanced.

2.4 Projection of a distribution

Consider now the distribution of a subset of the bits of a string x. We denote this by the term *projection.* We consider the projection reducing x to its first k bits denoted by $x_{(u)}$ and denote the last $n-k$ bits by $x_{(\overline{u})}$. We have

$$X_{(u)}(x_{(u)}) = \sum_{x_{(\overline{u})}} X(x_{(u)}||x_{(\overline{u})}) \ ,$$

and for the spectrum:

$$\widetilde{X_{(u)}}(v_{(u)}) = \widetilde{X}(v_{(u)}||0) \ .$$

So the spectrum of the projection of X is just a truncation of the spectrum of X. This can be generalized by defining $x_{(u)} = Zx$ with Z a binary *projection matrix* with k rows and n columns:

$$X_{(u)}(x_{(u)}) = \sum_{x} \delta(x_{(u)} = Zx)X(x) \ ,$$

and for the spectrum:

$$\widetilde{X_{(u)}}(v_{(u)}) = \widetilde{X}(Z^{\mathrm{T}}v_{(u)}) \ . \tag{6}$$

It may be the case that for a non-uniform distribution X, the projection is uniform. This is in fact the case if the spectrum is zero for all masks v that can be formed as $Z^{\mathrm{T}}v_{(u)}$. So global non-uniformity and local uniformity are not mutually exclusive.

3 Lossy mappings and their impact on local imbalance

In this section we show how mappings from $\mathrm{GF}(2)^n$ to $\mathrm{GF}(2)^m$ transform the spectrum of variables.

3.1 Correlation matrices and linear trails

The correlation between two Boolean functions with domain $\mathrm{GF}(2)^n$ can be expressed by a *correlation coefficient* that ranges between -1 and 1:

Definition 5. *The correlation coefficient* $\mathrm{C}(g(x), h(x))$ *associated with a pair of Boolean functions* $g(x)$ *and* $h(x)$ *is given by*

$$\mathrm{C}(g(x), h(x)) = 2 \cdot \Pr(g(x) = h(x)) - 1 \ ,$$

or equivalently

$$\mathrm{C}(g(x), h(x)) = \sum_{x} (-1)^{g(x)+h(x)} \ .$$

The structure of input-output correlations of a Boolean mapping $f(x)$ form an equivalent representation in the spectral domain. In particular, this contains the correlations between Boolean functions $u^{\mathrm{T}}f(x)$ on the one hand and $v^{\mathrm{T}}x$ on the other. This structure is the *correlation matrix*[5].

The correlation between an input mask v and an output mask u of a Boolean mapping is defined as:

$$\mathrm{C}(u^{\mathrm{T}}f(x), v^{\mathrm{T}}x) = \sum_{x} (-1)^{u^{\mathrm{T}}f(x)+v^{\mathrm{T}}x} \ .$$

Definition 6 ([5]). *The correlation matrix* C^f *of an* n*-bit to* m*-bit mapping* f *is a* $2^n \times 2^m$ *matrix with element* $C^f_{u,w}$ *in row* u *and column* w *equal to* $\mathrm{C}(u^{\mathrm{T}}f(x), w^{\mathrm{T}}x)$.

Row u of a correlation matrix can be interpreted as

$$(-1)^{u^{\mathrm{T}} f(x)} = \sum_w C^f_{u,w} (-1)^{w^{\mathrm{T}} x} \ .$$

This expresses an output parity with respect to the basis of input parities.

A correlation matrix C^f defines a linear map with domain $\mathbb{R}^{2^n}$ and range $\mathbb{R}^{2^m}$. Clearly, we have

$$\mathcal{L}(f(x)) = C^f \mathcal{L}(x) \ .$$

In words, applying a Boolean function f to a Boolean vector x and multiplying the corresponding vector $\mathcal{L}(x)$ with the correlation matrix C^f are just different representations of the same operation. This is illustrated in Fig. 1.

$$\begin{tikzcd}
x \arrow[r, "f"] \arrow[d, Leftrightarrow, "\mathcal{L}"] & y = f(x) \arrow[d, Leftrightarrow, "\mathcal{L}"] \\
\mathcal{L}(x) \text{ with } \mathcal{L}(x)_v = (-1)^{x^{\mathrm{T}} v} \arrow[r, "C^f"] & \mathcal{L}(y) = C^{(f)} \mathcal{L}(x)
\end{tikzcd}$$

Fig. 1. The equivalence of a Boolean mapping and its correlation matrix.

Let F be a Boolean mapping that is the composition of a number of Boolean mappings f_i:

$$F = f_r \circ \ldots \circ f_2 \circ f_1 \ .$$

We call the mappings f_i round mappings.

The correlation matrix of F is the product of the correlation matrices of the round mappings f_i. We have

$$C^F = C^{f_r} \times \ldots \times C^{f_2} \times C^{f_1} \ .$$

An r-round *linear trail* Q [5], denoted by

$$Q = (q_0, q_1, q_2, \ldots q_r) \ ,$$

consists of the chaining of r successive correlations of the type $\mathrm{C}({q_i}^{\mathrm{T}} f_i(x), {q_{i-1}}^{\mathrm{T}} x)$. To this linear trail corresponds a *correlation contribution coefficient* C_Q ranging between -1 and $+1$ defined as:

$$C_Q = \prod_i C^{f_i}_{q_i, q_{i-1}} \ .$$

From this we can derive following lemma.

Lemma 1 ([5]). *The correlation between $u^{\mathrm{T}} F(x)$ and $w^{\mathrm{T}} x$ is the sum of the correlation contribution coefficients of all r-round linear trails Q with initial selection vector w and terminal selection vector u.*

$$C(u^{\mathrm{T}} F(x), w^{\mathrm{T}} x) = \sum_{q_0 = w, q_r = u} C_Q \ .$$

3.2 Propagation of imbalance through a mapping

Let f be a Boolean mapping from $\mathrm{GF}(2)^n$ to $\mathrm{GF}(2)^m$ and X is a distribution over $\mathrm{GF}(2)^n$, the domain of this mapping. Then the distribution Y of $y = f(x)$ is given by:

$$\Pr(Y = y) = \sum_x \delta(f(x) = y) \Pr(X = x) \ . \tag{7}$$

Given an input x with a given spectrum $\widetilde{X}$, we can compute the spectrum $\widetilde{Y}$ of $y = f(x)$ by applying the inverse Walsh-Hamadard transform to get X, apply Equation (7) to X to get Y and then apply the Walsh-Hadamard transform again to get $\widetilde{Y}$. However, we can also do it in a single step using the correlation matrix.

Lemma 2. *Given a Boolean mapping f and the spectrum $\widetilde{X}$ of its input x, the spectrum $\widetilde{Y}$ of its output $y = f(x)$ is given by*

$$\widetilde{Y} = C^f \times \widetilde{X} \ .$$

Proof. The spectrum of Y can be written as:

$$\widetilde{Y} = \sum_y \Pr(Y = y)\mathcal{L}(y) \ .$$

For the probabilities of Y we have:

$$\Pr(Y = y) = \sum_x \Pr(X = x)\delta(y = f(x)) \ .$$

Filling this in yields:

$$\widetilde{Y} = \sum_y \left(\sum_x \Pr(X = x)\delta(y = f(x)) \right) \mathcal{L}(y) \ .$$

Re-ordering and re-grouping this gives:

$$\begin{aligned}
\widetilde{Y} &= \sum_x \Pr(X = x) \left(\sum_y \delta(y = f(x))\mathcal{L}(y) \right) \\
&= \sum_x \Pr(X = x)\mathcal{L}(f(x)) \\
&= \sum_x \Pr(X = x)C^f\mathcal{L}(x) \\
&= C^f \sum_x \Pr(X = x)\mathcal{L}(x) \\
&= C^f\widetilde{X} \ .
\end{aligned}$$

□

In a correlation matrix, row 0 contains correlations where the output mask is all-zero. It immediately follows that in the correlation matrix of any mapping, all elements in row 0 are zero, except the element in column 0, that contains the correlation between two constant functions both equal to zero and is hence one. Column 0 contains correlations of output parities with input parities with zero input mask. An input x that is uniformly distributed has a spectrum that is all-zero for all non-zero masks and 1 in the zero mask. So column 0 contains the spectrum $\widetilde{Y}$ of $y = f(x)$ given a uniformly distributed x.

Analogous to the reduced spectrum of a distribution, we can now define the reduced correlation matrix C^{*f} of a mapping f as C^f with row 0 and column 0 removed. This technique was also used by Jérémy Parriaux [13]. For an n-bit to m-bit mapping, C^{*f} has $2^m - 1$ rows and $2^n - 1$ columns. Moreover, we denote the first column of the correlation matrix, with the element in row 0 removed, by I^f and call it the *imbalance vector* of f. It is simply the reduced spectrum $\widehat{Y}$ of $y = f(x)$ with x uniformly distributed. Note that for a balanced mapping the imbalance vector I^f is all-zero. We have:

$$\begin{bmatrix} 1 \\ \widehat{Y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ I^f & C^{*(f)} \end{bmatrix} \times \begin{bmatrix} 1 \\ \widehat{X} \end{bmatrix} .$$

We can now re-formulate Lemma 2 in terms of reduced spectra, correlation matrix and imbalance vector:

Lemma 3. *Given a Boolean mapping f and an input x with reduced spectrum $\widehat{X}$, the reduced spectrum $\widehat{Y}$ of $y = f(x)$ is given by*

$$\widehat{Y} = I^f + C^{*f} \times \widehat{X} .$$

In other words, the reduced spectrum of $y = f(x)$ consists of the sum of two terms. The first term is the imbalance vector of f and independent of x and the second term is the reduced spectrum of x multiplied by the reduced correlation matrix of f.

3.3 Propagation of imbalance through iterative mappings

Applying Lemma 3 to an iterative mapping $F = f_r \circ \ldots \circ f_2 \circ f_1$ yields following expression:

$$\widehat{Y} = \sum_{1 \leq i \leq r} \left(\left(\prod_{i < j \leq r} C^{*f_j} \right) \times I^{f_i} \right) + \left(\prod_{1 \leq j \leq r} C^{*f_j} \right) \times \widehat{X} . \tag{8}$$

When in Equation (8) considering the imbalance of an individual mask in $\widetilde{Y}$, we can express it using linear trails Q by applying Lemma 1 to the products of the round mapping correlation matrices:

$$\widehat{Y}(u) = \sum_{1 \leq i \leq r} \sum_{w} \left(\sum_{Q \text{ with } q_i = w \text{ and } q_r = u} C_Q \right) \times I^{f_i}[w] + \sum_{w} \left(\sum_{Q \text{ with } q_0 = w \text{ and } q_r = u} C_Q \right) \times \widehat{X}(w) . \tag{9}$$

So from Equation (9) it follows that the imbalance of a mask u equals the sum of the products of the non-zero components $I^{f_i}[w]$ of the imbalance vectors of all previous rounds, each one multiplied by the sum of the correlation contributions of the linear trails from w to u. Note that the effect of the imbalance vector of the last round, $I^{f_r}[w]$ is immediate: $\widetilde{Y}(u) = I^{f_r}[u] +$ other terms. The contribution of components of the imbalance vector of the penultimate round, $I^{f_{r-1}}[w]$ is diluted by the multiplication of correlations over f_{r-1}. In particular, a component $I^{f_{r-1}}[w]$ contributes $C^{f_{r-1}}_{u,w} I^{f_{r-1}}[w]$. Note that contributions can be constructive or destructive as the imbalances and correlations are signed. The contribution of components of earlier rounds becomes more and more diluted as the distance to the final round grows. They are multiplied by the correlation contribution of linear trails and typically cryptographic round functions are designed to not exhibit multiple-round linear trails with high correlation contribution. Equation (9) is useful when studying the possible loss of security due to non-uniformity of threshold scheme anti-DPA mechanisms [4].

4 Lossy mappings and their impact on macroscopic imbalance

In this section we define macroscopic non-uniformity metrics for distributions and study their evolution through iterative mappings. We repeatedly apply transformations f_i from a fixed set of transformations with known imbalance contribution. This is similar to but different from studying the cycle structure of a single transformation f. In the latter case iteration leads to cycles while in the case of different transformations no such cycles appear.

4.1 Collision probability and total imbalance

The *norm* of a vector A is defined as $\sqrt{\langle A, A\rangle}$. It turns out that a useful measure for the non-uniformity of a distribution X is the square of its norm, when seen as a vector, i.e., $||X||^2 = \langle X, X\rangle$. This quantity coincides with the *collision probability* of X, defined as:

Definition 7. *The collision probability* $\Pr_{coll}(X)$ *of a distribution* X *is the probability that two elements independently chosen according to the distribution* X *are the same. It is given by:*

$$\Pr_{coll}(X) = \sum_x X(x)^2 = ||X||^2 \ .$$

The negative of the binary logarithm of the collision probability is the so-called *collision entropy* [15]. It can be shown that the collision entropy forms a lower bound for the more familiar Shannon entropy by Jensen's inequality [16].

As the Walsh-Hadamard transform is the composition of an orthogonal transformation and a scaling, we have $||\widetilde{X}||^2 = 2^n ||X||^2$, or equivalently:

$$\sum_v \widetilde{X}(v)^2 = 2^n \Pr_{\text{coll}}(X) \ . \tag{10}$$

In other words, the sum of the squared imbalances over all masks for a given distribution X is fully determined by its collision probability.

The squared norm of the reduced spectrum is the sum of the non-trivial squared imbalances and it plays a central role in our analysis.

Definition 8. *The total imbalance* ϕ_X *of a distribution* X *is the squared norm of its reduced spectrum:*

$$\phi_X = ||\widehat{X}||^2 = \sum_{u \neq 0} \widetilde{X}(u)^2 \ .$$

Clearly, the total imbalance is fully determined by the collision probability through Equation (10):

$$\phi_X = 2^n \Pr_{\text{coll}}(X) - 1 \ . \tag{11}$$

The collision probability and total imbalance reach a minimum with a uniform distribution. A uniform distribution over $\text{GF}(2)^n$ has collision probability 2^{-n} and total imbalance 0. Uniformity of a distribution can be expressed alternatively as *having an all-zero reduced spectrum.*

The collision probability and total imbalance reach a maximum when the distribution is only non-zero for a single element in the domain. In that case the collision probability equals 1 and the total imbalance equals $2^n - 1$

For the collision probability of the product of independent distributions, it is trivial to prove following lemma.

Lemma 4. *The collision probability of a distribution that is the product of a number of independent distributions is the product of those of the component distributions*

$$\Pr_{coll}(X) = \prod_i \Pr_{coll}(X_{(i)}) \ .$$

4.2 Collision probability and imbalance contribution

We define the collision probability for a mapping f analogous to that of a distribution. It is the collision probability of the distribution Y of $y = f(x)$ if x has the uniform distribution.

Definition 9. *The collision probability* $\Pr_{coll}(f)$ *of a mapping* f *is the probability that* $f(x) = f(x')$ *holds for two randomly and uniformly chosen inputs* x *and* x'.

Similarly we can define the imbalance contribution in terms of its collision probability.

Definition 10. *The imbalance contribution* ϕ_f *of a mapping* f *is its collision probability multiplied by* 2^m*, minus 1:*

$$\phi_f = 2^m \Pr_{coll}(f) - 1 \ .$$

Clearly, the imbalance contribution of a mapping f is simply the squared norm of its imbalance vector I^f.

We can now define *balancedness* of a mapping f.

Definition 11. *A mapping* f *is balanced if it transforms an input with a uniform distribution into an output with uniform distribution. Equivalently, a mapping is balanced if its imbalance contribution is zero, or equivalently, its imbalance vector is zero.*

Given two transformations f and g operating on domains $\mathrm{GF}(2)^m$ and $\mathrm{GF}(2)^k$ respectively, their Cartesian product $h = f \times g$ operates on $\mathrm{GF}(2)^{m+k}$ and is defined as $h(x, y) = (f(x), g(y))$. Transformation h simply consists of the parallel application of f and g.

The collision probability of $h = f \times g$ is simply the product of those of f and g.

Lemma 5. *If* $h = f \times g$ *then* $\Pr_{coll}(h) = \Pr_{coll}(f) \Pr_{coll}(g)$.

Proof. Consider $x = (x_{(f)}, x_{(g)})$ and $y = (y_{(f)}, y_{(g)})$. We have $h(x) = h(y)$ iff $f(x_{(f)}) = f(y_{(f)})$ and $g(x_{(g)}) = g(y_{(g)})$. It follows immediately that the probability of a collision in h is the product of the collision probabilities in f and g. □

The following corollary is useful for computing the collision probability of S-box layers.

Corollary 1. *If* h *is the parallel application of a number of mappings* f_i*, then* $\Pr_{coll}(h) = \prod_i \Pr_{coll}(f_i)$.

For imbalance contributions this translates to:

$$\phi_f = \prod_i (\phi_{f_i} + 1) - 1 \ .$$

The properties of the serial composition of two transformations $h = g \circ f$ depends on the specific way f and g interact and in general not easy to determine exactly. In the special case that one of f and g is a permutation, the composed transformation simply inherits the collision probability and imbalance contribution of the other one.

4.3 Total imbalance evolution through a lossy mapping

From Lemma 3, we see that the reduced spectrum after f consists of the sum of the imbalance vector I^f and the spectrum before f multiplied by the reduced correlation matrix of f. Making some independence assumptions allows us to say something about the expected total imbalance after f.

First, we quantify the effect of the multiplication with C^{*f} on the (squared) norm of a vector. It is well known that a permutation f has an orthogonal correlation matrix [5] and for that case multiplication by the correlation matrix, or its reduced version, does not impact the norm. The mappings we are interested in are not invertible and have some imbalance contribution. We now show that multiplication by C^{*f} tends to multiply the norm with $1 - \frac{\phi_f}{2^n-1}$. We will denote this by c_f.

Lemma 6. *The expected value over the space of all possible input vectors X with $||X||^2 = 1$ of $||C^{*f} \times X||^2$ is exactly $1 - \frac{\phi_f}{2^n-1} = c_f$.*

Proof. For readability we will denote C^{*f} by C in this proof and use $E_{\text{condition}(X)}(f(X))$ to express the expected value of $f(X)$ chosen uniformly with only restriction that X satisfies the mentioned condition. Let $Y = C^{*f} \times X$. We have $||Y||^2 = ||CX||^2 = (CX)^{\mathrm{T}}CX = X^{\mathrm{T}}C^{\mathrm{T}}CX$. Let UDV be the singular value decomposition of C[10]. Here U and V are orthonormal matrices and D a diagonal matrix with on the diagonal the singular values d_i of C. Then we have $||Y||^2 = VX^{\mathrm{T}}D^{\mathrm{T}}U^{\mathrm{T}}UDVX = VX^{\mathrm{T}}D^2VX$ and hence $E_{||X||^2=1}(||Y||^2) = E_{||X||^2=1}(VX^{\mathrm{T}}D^2VX)$. If we denote VX by X', X' has the same norm as X as V is an orthonormal matrix. We now have (with x_i denoting the components of X':

$$E_{||X'||^2=1}(X'^{\mathrm{T}}D^2X') = E_{\sum_i x_i^2=1}(x_i^2 d_i^2) = \frac{\sum_i d_i^2}{2^n-1} \ .$$

So c_f equals the average of the squared singular values of the reduced correlation matrix C^{*f}.

The sum of the squared singular values of a matrix equals the sum of squared elements of that matrix [10]. So $\sum_i d_i^2 = \sum_{u\neq 0, w\neq 0} C^2_{u,w}$. As the only non-zero element in the first row of any correlation matrix is the element in column zero, we have $\sum_i d_i^2 = \sum_{u,w} C^2_{u,w} - \sum_u C^2_{u,0}$. Each row in a correlation matrix has norm 1, so this becomes $\sum_i D_i^2 = 2^n - 1 - \phi_f$. It follows that $c_f = 1 - \frac{\phi_f}{2^n-1}$. □

It follows that the term $C^{*f} \times \widehat{X}$ has an expected imbalance contribution $c_f\phi_X$. Second, we assume that I^f is independent of $C^{*f} \times \widehat{X}$. We think this is a reasonable assumption as they have different origins. In that case the squared norm of the sum of the two vectors is the sum of the squared norms of the vectors. We have:

$$\phi_Y \approx \phi_f + c_f\phi_X \ .$$

4.4 Total imbalance evolution in iterative mappings

If we make the same independence assumptions for Equation (8) we obtain:

$$\phi_Y = \sum_{1\leq i\leq r}\left(\left(\prod_{i<j\leq r} c_{f_j}\right) \times \phi_{f_i}\right) + \left(\prod_{1\leq j\leq r} c_{f_j}\right) \times \phi_X \ . \tag{12}$$

In typical use cases we have $r \lll 2^n/\phi_{f_j}$ implying $r \lll (1-c_{f_j})$ and hence $\prod_j c_{f_j} \approx 1$. This allows simplifying Equation (12) to:

$$\phi_Y \approx \sum_{1\leq i\leq r} \phi_{f_i} + \phi_X \ .$$

The expected total imbalance of Y is simply the sum of the imbalance contributions of the round mappings f_i plus the total imbalance of X. In other words, the total imbalance increases linearly with the number of rounds by simply accumulating their imbalance contributions. Similarly, the collision probability increases linearly and hence the collision entropy decreases logarithmically with the number of rounds.

Assuming all f_i have the same imbalance contribution ϕ_f, Equation (12) simplifies to:

$$\phi_Y = \frac{1-{c_f}^r}{1-c_f}\phi_f + {c_f}^r\phi_X \ .$$

If the mappings f_i are not invertible we have $c_{f_i} < 1$ and for r going to infinity this expression becomes

$$\phi_Y = \frac{\phi_f}{1-c_f} = 2^n - 1 \ .$$

This corresponds with Y having a peak distribution equal to 1 in a single value and zero elsewhere.

5 Sampling noise and random mappings

In many applications one samples from a set. Even if the sampling is done according to a uniform distribution, the resulting sets will exhibit imbalance and have non-zero total imbalance (unless every element from the domain happens to be sampled exactly one time). In this section we characterize the distributions that result from random sampling of $\mathrm{GF}(2)^n$, in a way similar to [6]. We consider two types of sampling: with and without replacement. It turns out that a random mapping can be modeled as a sampling. An injective random mapping corresponds to sampling without replacement and in absence of an injectivity requirement it corresponds to sampling with replacement.

5.1 Sampling with replacement and random transformations

In sampling with replacement, we take z independent samples from $\mathrm{GF}(2)^n$. Let U be the multi-set containing the z samples. It is well known that if $z \ggg 1$, the number of times a given value x occurs in U, its *cardinality*, has a Poisson distribution with $\lambda = z2^{-n}$ [7]. Hence, the components of $X(x)$ are distributed according to a Poisson distribution scaled by a factor z^{-1}:

$$\Pr\left(X(x) = \frac{i}{z}\right) = \frac{z^i 2^{-ni}}{i!} e^{-z2^{-n}} \ .$$

We can compute the distribution of the imbalance of a non-zero parity v using the expression $\widetilde{X}(v) = z^{-1}\sum_{x\in U}(-1)^{v^{\mathrm{T}}x}$. The imbalance is given by $1-2p/z$ with p the number of elements x in U with parity 1 in v. Each element of U is independent and the probabilities of this parity being 1 or -1 are both 1/2. It follows that the number p has a binomial distribution with mean $z/2$ and variance $z/4$. So for non-zero v, $\widetilde{X}(v)$, has a distribution with mean 0 and variance z^{-1}. If $z \ggg 1$, this distribution has a normal shape.

The expected collision probability is $z^{-1}+(1-z^{-1})2^{-n}$. The term z^{-1} is the probability of taking the same instance among the samples and the second term is the complement of that probability multiplied by the probability that two independent samples collide. Applying Equation (11) yields an expected total imbalance equal to $(2^n - 1)z^{-1}$.

The set of images of a random mapping from $\mathrm{GF}(2)^n$ to $\mathrm{GF}(2)^m$ simply coincides with that of a random sample with replacement of 2^n elements out of 2^m and hence the expected collision probability is $2^{-n} + (1 - 2^{-n})2^{-m}$ and the expected imbalance contribution $(2^m - 1)2^{-n}$. For a random transformation we have $n = m$ and this becomes $2^{-n+1} - 2^{-2n}$ and $1 - 2^{-n}$ respectively. Remarkably, a random transformation has an imbalance contribution close to 1.

When applying Lemma 4 we see that parallel composition of mappings with an imbalance contribution lower than that of a random transformation may result in a mapping with imbalance contribution higher than that of a random transformation. For example, parallel application of d S-boxes with imbalance contribution 1 results in an S-box layer with imbalance contribution $2^d - 1$.

The effect of projection on total imbalance depends on the shape of the spectrum. Assuming that the imbalances have a (near) flat distribution, projection from n to k bits reduces the total imbalance by dividing it by a factor $(2^n - 1)/(2^k - 1) \approx 2^{n-k}$.

5.2 Sampling without replacement and random injective mappings

In sampling without replacement, the sample set U contains z different elements from $\mathrm{GF}(2)^n$, with $z \leq 2^n$. It follows that $X(x)$ has a two-valued distribution with value 0 in $2^n - z$ elements and z^{-1} in z elements. The collision probability equals $\Pr_{\mathrm{coll}}(X) = z^{-1}$ and the total imbalance is $z^{-1}2^n - 1$. Note that if the size of the sample and the domain are equal, i.e. $z = 2^n$, we have a uniform distribution and the total imbalance becomes zero.

We can compute the distribution of the imbalance of a non-zero parity v using the expression $\widetilde{X}(v) = z^{-1} \sum_{x \in U} (-1)^{v^{\mathrm{T}}x}$. The imbalance is given by $1 - 2p/z$ with p the number of elements x in U with parity 1 in v. The number p has the probability distribution of p successes in z draws from a set of 2^n without replacement, where the total number of successes in the set is 2^{n-1}. This is given by the hypergeometric distribution [7]:

$$\Pr(p = i) = \frac{\binom{2^{n-1}}{i}\binom{2^{n-1}}{z-i}}{\binom{2^n}{z}} .$$

This distribution has mean $z/2$ and variance $(1 - z2^{-n})\frac{z}{4}$. It follows that for non-zero v, $\widetilde{X}(v)$ has a distribution with mean 0 and variance $(1 - z2^{-n})z^{-1}$. If $z \ggg 1$, this distribution has a normal shape.

The collision probability is equal to z^{-1}: one over the size of the sample. The total imbalance hence equals $z^{-1}2^n - 1$.

The collision probability of an injective mapping (implying $m \geq n$) coincides with that of a sample without replacement. The size of the sample is given by $z = 2^{-n}$, so we have $\Pr_{\mathrm{coll}}(f) = 2^{-n}$ and $\phi_f = 0$. An injective mapping with $n = m$ is a permutation and it has total imbalance 0 and collision probability 2^{-n}.

5.3 Summary of this section

We summarize the results of this section in Table 1.

6 Imbalance contribution of mappings with known collision profile

In this section we deal with mappings where the non-uniformity can be quantitatively characterized by a so-called *collision profile*. It turns out that this fully determines the collision probability and

	with replacement	**without replacement**
$X(x)$	scaled Poisson $\Pr(X = \frac{i}{z}) = \frac{\lambda^i}{i!}e^{-\lambda}$ with $\lambda = \frac{z}{2^n}$	two-valued $\Pr(X = 0) = 1 - \frac{z}{2^n}$ $\Pr(X = \frac{1}{z}) = \frac{z}{2^n}$
$\widetilde{X}(v)$	very close to normal mean: 0 variance: z^{-1}	very close to normal mean: 0 variance: $(1 - z2^{-n})z^{-1}$
$\Pr_{\text{coll}}(X)$	mean: $z^{-1} + (1 - z^{-1})2^{-n}$	equals z^{-1}
ϕ_X	mean: $z^{-1}(2^n - 1)$	equals $z^{-1}2^n - 1$
ϕ_X if $z = 2^n$	mean: $1 - 2^{-n}$	equals 0

Table 1. Statistical characteristics of samples with size z.

imbalance contribution. We also provide some experimental evidence of the theoretically predicted evolution of the total imbalance.

6.1 Collision profile and implications

Definition 12. *The* collision partition *of a mapping f is the one defined by $f(x) = f(y)$. In other words, two elements x and y of the domain are in the same subset if and only if $f(x) = f(y)$. We call the subsets of the partition* collision sets *and a collision set with i elements an i*-collision*.*

Based on the collision partition of a transformation f we can define its collision profile.

Definition 13. *The collision profile of a transformation f is the list $(C_f[1], C_f[2], \ldots)$ where $C_f[i]$ denotes the number of i-collisions in f.*

Clearly, the total number of inputs in i-collisions is $iC_f[i]$ and so it follows that $\sum_i iC_f[i] = 2^n$.

The collision probability of a mapping f is determined by its collision profile.

Lemma 7. *The collision probability of an n-bit to m-bit mapping f with known collision profile is given by:*

$$\Pr_{coll}(f) = \frac{1}{2^{2n}} \sum_i i^2 C_f[i] \ .$$

Proof. The probability equals the number of cases (x, y) leading to a collision divided by the total number of cases:

$$\Pr_{\text{coll}}(f) = \frac{1}{2^{2n}} \sum_{x,y} \delta(f(x) = f(y)) \ .$$

In other words:

$$\Pr_{\text{coll}}(f) = \frac{1}{2^{2n}} \sum_{x,y} \delta(x \text{ and } y \text{ are in the same collision set}) \ .$$

The number of colliding pairs (x, y) in an i-collision set is i^2, hence:

$$\Pr_{\text{coll}}(f) = \frac{1}{2^{2n}} \sum_i i^2 C_f[i] \ .$$

□

The value of the imbalance contribution follows from this:

Corollary 2. *The imbalance contribution of an n-bit to m-bit mapping f with known collision profile is given by:*

$$\phi_f = \frac{1}{2^{2n-m}} \sum_i i^2 C_f[i] - 1 \ .$$

6.2 Example: a round function with lossy S-boxes

Assume we have a round function consisting of a lossy nonlinear S-box layer N and a linear layer L and we wish to determine its total imbalance. First, thanks to the invertibility of the linear layer, the total imbalance of the round function is the total imbalance of the lossy S-box layer. Second, for an S-box of reasonable width, it is easy to determine the collision profile and hence its collision probability. This allows determining the collision probability of the non-linear layer. Assume we have m identical S-boxes of width n. Then the collision probability of the nonlinear layer is $\Pr_{\text{coll}}(N) = \Pr_{\text{coll}}(S)^n$. Translated to total imbalances this gives: $\phi_N = 2^{nm}\Pr_{\text{coll}}(S)^m - 1$.

Let now $\Pr_{\text{coll}}(S)$ be 2^{-a}: the S-box reduces the set of 2^m inputs to a set with the same collision probability as a set of 2^a elements. Then we have $\phi_N = 2^{n(m-a)} - 1$ and $c_N = 1 - \frac{2^{n(m-a)}-1}{2^{nm}-1}$. If $n(m-a) \ggg 1$, we have these expressions simplify to $\phi_N \approx 2^{n(m-a)}$ and $c_N \approx 1 - 2^{-na}$. Assume we have a block cipher with a block size nm of 128 bits and a 4-bit S-box with $\Pr_{\text{coll}}(S) = 2^{-3}$. Then we have $\phi_N \approx 2^{32}$ and $c_N \approx 1 - 2^{-96} \approx 1$. The total imbalance after r rounds is simply $2^{32}r$ implying a collision probability of $2^{-96}r$. This lower bounds the collision entropy, and hence also the Shannon entropy, to $96 - \log_2(r)$.

7 Experiments

We did a number of experiments to check the validity of our independence assumptions. More particularly, we randomly constructed transformations f with domains of size 2^e with e ranging from 22 to 27 and for each of them we tracked the total imbalance when applying randomized versions of f to it iteratively. We did this by tracking the image set as the number of rounds increases. We initialize the image set to the full domain and randomize the application of f by bitwise addition with a constant that is randomly generated for each i but equal for all elements in the image set.

Initially each element in the image set has probability 2^{-e} and the total imbalance is zero. If the first iteration of f maps w elements to some element, this element has probability $w2^{-e}$. In our experiments we keep track of these probabilities and compute from them the total imbalance.

We studied two types of pseudorandomly generated transformations. Those in the first category were generated without side conditions. Those in the second category satisfy specific collision profiles: only 2^{-f} of the images are possible and each image has 2^f pre-images. We composed these of a random permutation followed by a simple transformation satisfying the collision profile, followed by a (independently generated) random permutation. The random permutations were generated with the Fisher-Yates shuffle [9].

Figure 2 illustrates the outcome of our experiments. The continuous lines represent the values taken by simply multiplying the imbalance contribution of the transformations by the number of iterations for the random transformation (imbalance contribution 1) and the one that maps 64 values to a single one (imbalance contribution 63). The figure shows that the experimentally measured total imbalances follows these linear profiles quite closely.

8 Conclusions and acknowledgments

In this paper we have provided a formalism to describe non-uniformity in the spectral domain using imbalances. The occurrence and propagation of these imbalances can be described by correlation matrices and linear trails. We have introduced macroscopic metrics for non-uniformity in the form of total imbalance. When iteratively applying lossy mappings to a variable, its total imbalance increases linearly with the number of rounds and its entropy decreases logarithmically. The tools we

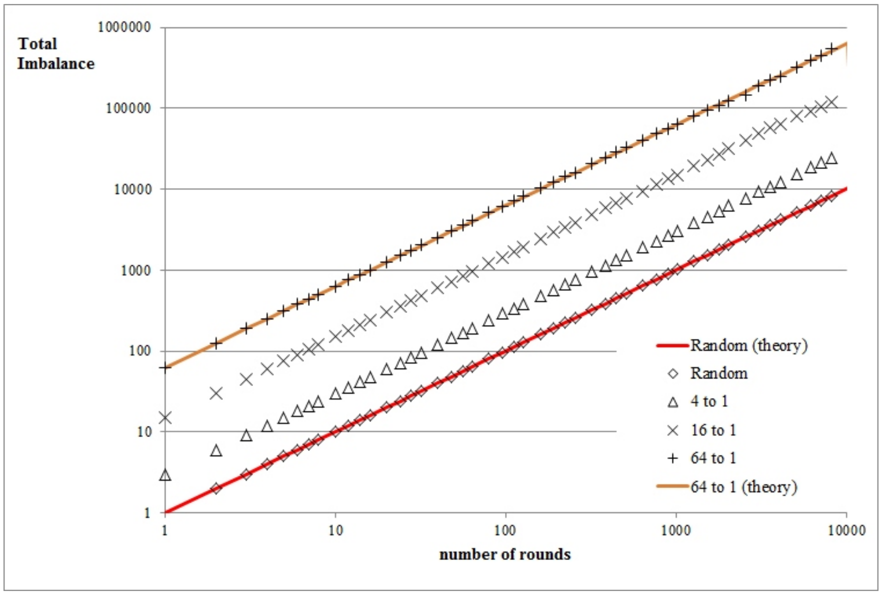

Fig. 2. Evolution of total imbalance for different transformations.

provide in this paper are helpful when studying non-invertible cryptographic modes and primitives, including non-uniform threshold schemes.

8.1 Acknowledgements

I thank Gilles Van Assche, Guido Bertoni, Svetla Nikova, Ventzi Nikov and Begül Bilgin for useful comments.

References

1. Thierry P. Berger, Joffrey D'Hayer, Kevin Marquet, Marine Minier, and Gaël Thomas, *The GLUON family: A lightweight hash function family based on fcsrs*, Progress in Cryptology - AFRICACRYPT 2012 - 5th International Conference on Cryptology in Africa, Ifrance, Morocco, July 10-12, 2012. Proceedings (Aikaterini Mitrokotsa and Serge Vaudenay, eds.), Lecture Notes in Computer Science, vol. 7374, Springer, 2012, pp. 306–323.
2. G. Bertoni, J. Daemen, M. Peeters, and G. Van Assche, *Building power analysis resistant implementations of* Keccak, Second SHA-3 candidate conference, August 2010.
3. ______, *The* Keccak *reference*, January 2011, `http://keccak.noekeon.org/`.
4. Begül Bilgin, Joan Daemen, Ventzislav Nikov, Svetla Nikova, Vincent Rijmen, and Gilles Van Assche, *Efficient and first-order DPA resistant implementations of keccak*, Smart Card Research and Advanced Applications - 12th International Conference, CARDIS 2013, Berlin, Germany, November 27-29, 2013. Revised Selected Papers (Aurélien Francillon and Pankaj Rohatgi, eds.), Lecture Notes in Computer Science, vol. 8419, Springer, 2013, pp. 187–199.
5. J. Daemen, *Cipher and hash function design strategies based on linear and differential cryptanalysis, PhD thesis*, K.U.Leuven, 1995.
6. J. Daemen and V. Rijmen, *Probability distributions of correlation and differentials in block ciphers*, Journal of Mathematical Cryptology **1** (2007), no. 3, 221–242.
7. W. Feller, *Introduction to probability theory and its applications, vol. 1*, Wiley & Sons, 1968.
8. S. Golomb, *Shift register sequence*, Holden-Day, 1967.
9. D. E. Knuth, *The art of computer programming, vol. 2, third edition*, Addison-Wesley Publishing Company, 1998.
10. D. Lay, S. Lay, and J. McDonald, *Linear algebra and its applications*, 5 ed., Pearson, 2016.

11. S. Nikova, V. Rijmen, and M. Schläffer, *Secure hardware implementation of nonlinear functions in the presence of glitches*, ICISC (P. J. Lee and J. H. Cheon, eds.), Lecture Notes in Computer Science, vol. 5461, Springer, 2008, pp. 218–234.
12. ———, *Secure hardware implementation of nonlinear functions in the presence of glitches*, J. Cryptology **24** (2011), no. 2, 292–321.
13. J. Parriaux, P. Guillot, and G. Millerioux, *Towards a spectral approach for the design of self-synchronizing stream ciphers*, Cryptography and Communications **3** (2011), no. 4, 259–274.
14. Léo Perrin and Dmitry Khovratovich, *Collision spectrum, entropy loss, t-sponges, and cryptanalysis of GLUON-64*, Fast Software Encryption - 21st International Workshop, FSE 2014, London, UK, March 3-5, 2014. Revised Selected Papers (Carlos Cid and Christian Rechberger, eds.), Lecture Notes in Computer Science, vol. 8540, Springer, 2014, pp. 82–103.
15. A. Rényi, *On measures of information and entropy*, Proceedings of the fourth Berkeley Symposium on Mathematics, 1960, p. 547–561.
16. E. Weisstein, *Jensen's inequality from mathworld – a wolfram web resource*, `http://mathworld.wolfram.com/JensensInequality.html`.