Università degli studi di Pavia
Dipartimento di Matematica
Corso di Laurea Magistrale in Matematica

UNIVERSITÀ
DI PAVIA

Sui gruppi di automorfismo di alcuni anelli radicali con applicazione alla crittografia

On the automorphism group of certain radical rings with application to cryptography

Tesi di Laurea Magistrale in Matematica

Relatore (Supervisor):
Chiar.mo Prof. Alberto Canonaco
Correlatore (Co-Supervisor):
Chiar.mo Prof. Marco Calderini

Tesi di Laurea di:
Riccardo Invernizzi
Matricola 492174

Anno Accademico 2020-2021

Introduction

Differential cryptanalysis [3] is probably one of the most efficient and more used techniques to attack modern ciphers. It consists in trying to find recurrent relationship between a change in the plain message and the consequent change in the ciphertext (the encrypted message). If one is able to find a relationship that occurs with high enough frequency, it can unveil important information about the secret key used in the encryption.
Since the message space is generally a vector space over $\mathbb{F}_2$, changes in the message can be seen as differences with respect to the XOR, the usual sum over $(\mathbb{F}_2)^n$. In this work we analyze the results that we can obtain by applying differential cryptanalysis with respect to an alternative sum defined on the message space. This general idea has been widely discussed by many authors (e.g. [1], [2]). However more recently Calderini et al. ([7],[5],[6]) suggested a new approach, based on the use of groups isomorphic to the translation group. Each of these groups allows us to define a sum which induce on the message space the structure of vector space. They also show strong connections with important algebraic structures such as radical rings ([10]), and this fact can be exploited ([7], [12]) to give them a convenient representation and a fast way to compute them. In [12] these particular sums are used to mount an attack on a block cipher that can be considered secure with respect to classical differential cryptanalysis, but may be broken with this new technique. Of course, the structure and the behavior of block ciphers is well known when dealing with the traditional XOR. We need to better investigate and understand it in terms of the newly introduced operation in order to mount a successful attack.
The aim of this work is to extend the results obtained in [7] and [12] to some cases they did not cover. In this way we provide a better understanding of bounds and conditions given in those articles, and in some cases we also suggest a further generalization of the results obtained. Moreover, we open a path for the use of attacks like the one presented in [12] in new settings, potentially leading to new interesting results.
In Chapter 1, we present basic concepts and definitions that we will use through all the work. In Chapter 2, we give a short but sufficient (at least for our purposes) introduction to cryptosystems, and especially block ciphers. We explain what it means for a cryptosystem to be secure, and how this security can be threatened. In Chapter 3, we focus on a specific component of a block cipher, namely the S-box, which in a certain sense contains most of the security of the cipher itself. We present some well known results on optimal ways to design this component. In Chapter 4, we outline the differential attack in our setting. We

split the interaction of our new operation with a block cipher in the interaction with each component of it. The first one (namely, the key addition) gives us a small disadvantage with respect to classical differential cryptanalysis; however, we cannot act on it. The second one (the diffusion layer, which is given by an invertible matrix) if not properly controlled will probably result in a total failure of our whole attack. However, a diffusion layer which is linear also with respect to our alternative sum gives us no disadvantage at all. Finally, the last component, the confusion layer, can give us a great advantage. It is then clear that our performance depends on the last two components. In Chapter 5, the heart of our work, we extensively analyze the automorphism group of the radical ring connected to our new sum. It turns out that this group contains almost all the information regarding the interaction of our sum with the diffusion layer. We firstly present some general results and then the most important ones, which are due to [7] and [12]. Finally, we show how we are able to extend them in new settings. These results are useful for trying to exploit attacks based on alternative sums which have not been considered in [7] and [12]. In Chapter 6, the last one, we present the interaction of our new sum with the confusion layer (more precisely with the optimal S-boxes presented in Chapter 3) and how we are able to outperform the classic XOR-based analysis.

Contents

Chapter 1

Preliminary definitions

1.1 Alternative sums

In this work $V = (\mathbb{F}_2)^n$ (with $n \geq 2$) will be a binary vector space. For a vector $v \in V$, v^j is the j-th component of v. The canonical basis will be denoted by $\{e_i\}$ with $e_i^j = 1$ if and only if $i = j$, otherwise $e_i^j = 0$. $\mathrm{Sym}(V)$ is the group of all the permutations on V. $\mathrm{GL}(V) \subseteq \mathrm{Sym}(V)$ is the general linear group on V, i.e. the group of all the linear permutations on V, while

$$\mathrm{T} := \{\sigma_a | a \in V, \sigma_a : x \mapsto x + a\} \subseteq \mathrm{Sym}(V)$$

is the group of all the translations on V. Finally, we define $\mathrm{AGL}(V) := \mathrm{GL}(V) \ltimes \mathrm{T}(V)$, or the normalizer of T in $\mathrm{Sym}(V)$, as the affine general linear group on V. A more detailed presentation of these facts, together with some proofs omitted here for brevity, can be found in [7] and [12]. Theorem 1.6 is due to [17].

Notation 1.1. We will use postfix notation for function evaluation, i.e. if $g \in \mathrm{Sym}(V)$ and $v \in V$ we write vg to mean $g(v)$.

Definition 1.2. *An elementary abelian group is an abelian group in which every nontrivial element has order a prime p.*

Definition 1.3. *The action of a group G on a set V is said to be transitive if for all $x, y \in V$ there exist $g \in G$ such that $y = xg$.*

Definition 1.4. *A permutation group G acting transitively on a set V is said to be regular if for all $v \in V$, $G_v := \{g \in G | vg = v\}$ (the stabilizer of G at v) is trivial.*

Remark 1.5. *If G is a regular group acting on a finite set V, then for all $x, y \in V$ there exists $g \in G$ such that $xg = y$.*

Theorem 1.6. *Let $\mathcal{T} < \mathrm{Sym}(V)$ an elementary abelian regular subgroup. There exists $g \in \mathrm{Sym}(V)$ such that $\mathcal{T} = \mathrm{T}^g = g^{-1}\,\mathrm{T}\,g$.*

Remark 1.7. *In this setting, it is possible to represent $\mathcal{T} = \{\tau_a | a \in V\}$ where τ_a is the unique map in $\mathcal{T}$ sending 0 to a.*

Definition 1.8. *Let $\mathcal{T} < \mathrm{Sym}(V)$ be an elementary abelian regular subgroup. We can define an additive law $\circ$ on V by letting for each a, b in V*

$$a \circ b := a\tau_b,$$

where τ_b is the unique element of $\mathcal{T}$ sending 0 to b.

Proposition 1.9. *Let $\mathcal{T} < \mathrm{Sym}(V)$ be an elementary abelian regular subgroup, and $\circ$ be the operation related to it as in Definition 1.8. Then, $(V, \circ)$ is a vector space over $\mathbb{F}_2$, with associated translation group $\mathrm{T}_\circ = \mathcal{T}$. Moreover, $(V, \circ) \cong (V, +)$.*

Proof. First of all, let us prove that $(V, \circ)$ is an abelian group. By definition,

$$a \circ (b \circ c) = 0\tau_a(\tau_b\tau_c) = 0(\tau_a\tau_b)\tau_c.$$

0 is clearly the neutral element. Moreover, since $\mathcal{T}$ is elementary $a\tau_a = 0$ for each a. Finally, since $\mathcal{T}$ is an abelian group, for each $a, b \in V$ we have $\tau_a\tau_b = \tau_b\tau_a$. As a consequence it holds

$$a \circ b = 0\tau_a\tau_b = 0\tau_b\tau_a = b \circ a.$$

This also proves that $(V, \circ)$ is a vector space over $\mathbb{F}_2$, and since $|V| < \infty$, $(V, \circ)$ and $(V, +)$ are isomorphic vector spaces. ∎

Definition 1.10. *Let $\mathcal{T} < \mathrm{Sym}(V)$ be an elementary abelian regular subgroup, $\circ$ the operation defined by it and $\mathrm{T}_\circ$ the associated translation group. We denote by $\mathrm{AGL}(V, \circ) = \mathrm{AGL}(V)^g$ the normalizer of $\mathrm{T}_\circ$ and by $\mathrm{GL}(V, \circ)$ the stabilizer of $\{0\}$ in $\mathrm{AGL}(V, \circ)$. For sake of clarity, we will sometimes denote T as T_+ , $\mathrm{AGL}(V)$ by $\mathrm{AGL}(V, +)$ and $\mathrm{GL}(V)$ by $\mathrm{GL}(V, +)$.*

Definition 1.11. *Given an operation $\circ$ as above, a vector $k \in V$ is called a weak key if for each $x \in V$ it holds $x + k = x \circ k$. The set*

$$W_\circ := \{k | k \in V,\ k \text{ is a weak key}\}$$

is called the weak-keys space, and is a subspsace of both $(V, +)$ and $(V, \circ)$. Hereafter $\dim(W_\circ)$ will be denoted by d.

Definition 1.12. *Given an operation $\circ$ as above, we can introduce a dot product on V defined for each $a, b \in V$ by*

$$a \cdot b := a + b + a \circ b.$$

The set of elements that can be expressed as dot prouct is deonte by

$$U_\circ := \{x \cdot y | x, y \in V\}$$

and is called set of errors.

Definition 1.13. *Given an operation $\circ$ as above, we define*

$$H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ).$$

A matrix $\lambda \in (\mathbb{F}_2)^{n\times n}$ is said to be compatible with $\circ$ if $\lambda \in H_\circ$.

1.2 Boolean Functions

We now give a brief introduction to Boolean functions, a kind of function we will use extensively in the next chapters. A more detailed presentation of these facts, together with some proofs omitted here for brevity, can be found in [11].

Definition 1.14. *A Boolean function in n variables is $f : \mathbb{F}_2^n \to \mathbb{F}_2$. We also define $\mathcal{B}_n = \{f : \mathbb{F}_2^n \to \mathbb{F}_2\}$ the set of all Boolean functions.*

Remark 1.15. $|\mathcal{B}_n| = |\mathbb{F}_2|^{|\mathbb{F}_2^n|} = 2^{2^n}$.

Definition 1.16. *Given $\mathbb{F}_2^n = \{v_0, ..., v_{2^n-1}\}$, we define the evaluation map $\mathrm{ev} : \mathcal{B}_n \to \mathbb{F}_2^{2^n}$ given by*

$$f \mapsto (f(v_0), ..., f(v_{2^n-1})).$$

The vector $\overline{f} = \mathrm{ev}(f)$ is called the truth table of f.

Definition 1.17. *Let $[n] := \{1, ..., n\}$, and $S \subseteq [n]$. The monomial*

$$x^S = \prod_{i\in S} x_i \in \mathbb{F}_2[x_1, ..., x_n]$$

is called square-free. A polynomial $p \in \mathbb{F}_2[x_1, \ldots, x_n]$ is called square-free if it is composed only by square-free monomials. By definition $x^\emptyset = 1$.

Remark 1.18. Square-free polynomials can also be seen as elements of

$$\mathbb{F}_2[x_1, ..., x_n] \Big/ (x_i^2 + x_i).$$

Definition 1.19. *We denote by R_n the space of square-free polynomials in n variables. Notice that this is a vector space with basis*

$$L = \left\{x^I : I \subseteq [n]\right\}.$$

Theorem 1.20 (Algebraic Normal Form)**.** *Every Boolean function can be uniquely represented as a square-free polynomial*

$$f(x_1, ..., x_n) = \sum_{I\subseteq[n]} a_I x^I,\ a_I \in \mathbb{F}_2.$$

This representation is called Algebraic Normal Form (ANF).

Proof. Let $p \in \mathbb{F}_2[x_1, ..., x_n]$ square-free. We can map

$$p \mapsto f_p \in \mathcal{B}_n$$

given by $f_p(v) = p(v)$. Notice that $|R_n| = |\mathcal{B}_n| = 2^{2^n}$. Now we want to show that for two square-free polynomials $p_1 \neq p_2$ we have $f_{p_1} \neq f_{p_2}$. We have

$$p_1 = \sum_{I \subseteq [n]} a_I x^I$$

and

$$p_2 = \sum_{I \subseteq [n]} b_I x^I.$$

Let us assume $f_{p_1} = f_{p_2}$. We get $f_{p_1} - f_{p_2} = 0$ or equivalently $p_1(v) - p_2(v) = 0$ for each $v \in \mathbb{F}_2^n$. We call the difference $F := p_1 - p_2$. We know that f is also square-free, and so we can write

$$F = \sum_{I \subseteq [n]} c_I x^I,$$

with $c_I = a_I - b_I$. But $F(v) = 0$ for all v, hence $F(0) = c_\emptyset = 0$. From the coefficient of $\emptyset$ we easily get all the degree one coefficients, since $F(e_i) = c_\emptyset + c_{\{i\}} = 0$, and so on. We have proven that $F = 0$ and consequently $p_1 = p_2$. ∎

Definition 1.21. *We define the support of a vector v as*

$$\operatorname{supp}(v) := \{i : v_i \neq 0\}.$$

Proposition 1.22. *Let $f \in \mathcal{B}_n$. The coefficients a_I defined in Theorem 1.20 for f can be computed as*

$$a_I = \sum_{\substack{v \in \mathbb{F}_2^n \\ \operatorname{supp}(v) \subseteq I}} f(v).$$

Proof. Let

$$b_I = \sum_{\substack{v \in \mathbb{F}_2^n \\ supp(v) \subseteq I}} f(v)$$

and

$$g(x) = \sum_{I \subseteq [n]} b_I x^I$$

be a square-free polynomial. By Theorem 1.20 it is enough to prove that $g(w) = f(w)$ for each $w \in \mathbb{F}_2^n$. We have

$$g(w) = \sum_{I \subseteq supp(w)} \sum_{\substack{v \in \mathbb{F}_2^n \\ supp(v) \subseteq I}} f(v) = \sum_{v \in \mathbb{F}_2^n} f(v) \sum_{\substack{I \subseteq [n] \\ supp(v) \subseteq I \subseteq supp(w)}} 1$$

If $\operatorname{supp}(v) \not\subseteq \operatorname{supp}(w)$ the last sum is 0. Otherwise

$$|\{I \subseteq [n] : supp(v) \subseteq I \subseteq (supp(w)\}| = 2^{|supp(v)| - |supp(w)|}$$

and the one and only possibility for the sum not to vanish is $|supp(v)| - |supp(w)| = 0$ or $v = w$. This implies $g(w) = f(w)$ for each $w \in \mathbb{F}_2^n$. ∎

Example 1.23. Let us fix $n = 3$ and let $f : \mathbb{F}_2^3 \to \mathbb{F}_2$ be a Boolean function with truth table

x_1	x_2	x_3	hex	f(x)
0	0	0	0_x	0
0	0	1	1_x	0
0	1	0	2_x	0
0	1	1	3_x	1
1	0	0	4_x	1
1	0	1	5_x	1
1	1	0	6_x	0
1	1	1	7_x	1

where each vector is associated with its hexadecimal representation, most significant bit first. We have $a_\emptyset = f(0_x) = 0$, since 0_x is the only vector with empty support. For higher degrees, we have for example $a_1 = f(0_x) + f(1_x) = 1$, $a_{12} = f(000) + f(100) + f(010) + f(110) = 1$ and so on. We obtain the ANF for f as

$$f(x) = x_1 + x_1x_2 + x_2x_3.$$

Definition 1.24. *Let $f \in \mathcal{B}_n$. We define the weight of f as*

$$w(f) = |\{v \in \mathbb{F}_2^n : f(v) = 1\}|,$$

$w(f)$ can also be seen as the Hamming weight for the vector $\overline{f}$.

Definition 1.25. *A function $f \in \mathcal{B}_n$ is called balanced if $w(f) = 2^{n-1}$. Notice that in this case we have the same number of 0 and 1 as outcome.*

Definition 1.26. *Let $f \in \mathcal{B}_n$ with algebraic normal form*

$$f(x) = \sum_{I \subseteq [n]} a_I x^I.$$

We define the algebraic degree of f as

$$\deg_A(f) = \max\{|I| : a_I \neq 0\}.$$

If $\deg_A(f) \leq 1$ we will call f an affine Boolean function, and we can write $f(x) = a_0 + a_1x_1 + ... + a_nx_n$. The set of all affine functions in $\mathcal{B}_n$ is denoted by A_n.

Remark 1.27. For $I \subseteq [n]$, it holds $w(x^I) = 2^{n-|I|}$. In fact we have

$$x^I(v) = \begin{cases} 0 \text{ if } I \not\subseteq supp(v) \\ 1 \text{ if } I \subseteq supp(v). \end{cases}$$

We also notice that $w(x^I)$ is odd $\iff I = [n]$.

Proposition 1.28. *$w(f)$ is odd $\iff \deg_A(f) = n$.*

Proof. First of all, notice that

$$w(f+g) = w(f) + w(g) - 2\left|\{v : f(v) = g(v) = 1\}\right|,$$

since on those elements the sum vanishes. Hence if $w(f)$ and $w(g)$ have the same parity $w(g+f)$ will be even, otherwise it will be odd. We know that $f \in \mathcal{B}_n$ is a sum of monomial of the form x^I. Moreover, thanks to Observation 1.27, if $\deg_A(f) < n$ the sum is over elements with even weight and than also f will have even weight. Otherwise we may write

$$f(x) = x_1 \cdots x_n + g(x)$$

with $\deg_A(g) < n$, and since g has even weight and $x_1 \cdots x_n$ has odd weight $w(f)$ will be odd. ∎

Definition 1.29. *Let $f, g \in \mathcal{B}_n$. We define the distance between f and g as*

$$d(f,g) := |\{v : f(v) \neq g(v)\}|.$$

Remark 1.30. $d(f,g) = w(f+g)$.

Definition 1.31. *Let $A_n = \{g : \deg_A(g) \leq 1\}$ be the set of all affine functions. The non-linearity of a Boolean function $f \in \mathcal{B}_n$ is given by*

$$\mathrm{NL}(f) := d(f, A_n) = \min_{g \in A_n} d(f,g),$$

i.e. the distance of f from the closest affine function.

For some applications, better explained in the next chapters, we are interested in Boolean functions with high nonlinearity.

Definition 1.32. *We define the Fourier transform of f in a as*

$$\mathcal{F}_f(a) = \sum_{x \in \mathbb{F}_2^n} f(x)(-1)^{a \cdot x}.$$

Moreover, we define the Walsh transform of f in a as

$$\mathcal{W}_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+a\cdot x} = \mathcal{F}_{(-1)^{f(\cdot)}}(a).$$

Remark 1.33. It holds

$$\mathcal{W}_f(a) = 2^n \delta_0(a) - 2\mathcal{F}_f(a)$$

where

$$\delta_0(a) = \begin{cases} 1 \text{ if } a = 0 \\ 0 \text{ otherwise} \end{cases}$$

Proof.

$$\mathcal{W}_f(a) + 2\mathcal{F}_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+a\cdot x} + 2 \sum_{x \in \mathbb{F}_2^n} f(x)(-1)^{a\cdot x}. \tag{1.1}$$

Let now $a = 0$. Then

$$(1.1) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)} + 2 \sum_{x \in \mathbb{F}_2^n} f(x) =$$
$$= |\{x : f(x) = 0\}| - |\{x : f(x) = 1\}| + 2|\{x : f(x) = 1\}| =$$
$$= |\{x : f(x) = 0\}| + |\{x : f(x) = 1\}| = 2^n,$$

because we are counting over all $\mathbb{F}_2$. This proves the equality if when $a = 0$. If $a \neq 0$ instead we have

$$(1.1) = \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot x} [2f(x) + (-1)^{f(x)}] = \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot x} = \tag{1.2}$$
$$= |\{x : a \cdot x = 0\}| - |\{x : a \cdot x = 1\}| = 0,$$

since we can se $a \cdot x$ as a balanced function and hence the two sets have the same cardinality. ∎

Remark 1.34. As a consequence we have

$$w(f) = \mathcal{F}_f(0) = 2^{n-1} - \frac{1}{2}\mathcal{W}_f(0) = d(f, 0).$$

Moreover, if we denote by $l_a(x) = a \cdot x$, it holds

$$d(f, l_a) = w(f + l_a) = 2^{n-1} - \frac{1}{2}\mathcal{W}_{f+l_a}(0) =$$
$$2^{n-1} - \frac{1}{2} \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+l_a} = 2^{n-1} - \frac{1}{2}\mathcal{W}_f(a).$$

Finally, $d(f, l_a + 1) = 2^{n-1} + \frac{1}{2}\mathcal{W}_f(a)$.

Proposition 1.35. *It holds* $NL(f) = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |\mathcal{W}_f(a)|$.

Proof. It follows directly from Remark 1.34. ∎

Lemma 1.36 (Parseval)**.** *We have*

$$\sum_{a \in \mathbb{F}_2^n} \mathcal{W}_f^2(a) = 2^{2n}.$$

Proof.

$$\sum_{a \in \mathbb{F}_2^n} \left(\sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+a \cdot x} \right)^2 = \sum_{a \in \mathbb{F}_2^n} \sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x)+f(y)+a \cdot (x+y)} =$$
$$\sum_{x,y \in \mathbb{F}_2^n} (-1)^{f(x)+f(y)} \cdot \sum_{a \in \mathbb{F}_2^n} (-1)^{a \cdot (x+y)}.$$

Now if $x \neq y$ like above $l_{x+y}(v) = (x + y) \cdot v$ is a balanced function and this

implies

$$\sum_{a\in\mathbb{F}_2^n}(-1)^{a\cdot(x+y)}=0.$$

If instead $x+y=0$ we have

$$\sum_{a\in\mathbb{F}_2^n}(-1)^{a\cdot(x+y)}=2^n.$$

But then

$$\sum_{x,y\in\mathbb{F}_2^n}(-1)^{f(x)+f(y)}\cdot\sum_{a\in\mathbb{F}_2^n}(-1)^{a\cdot(x+y)}=\sum_{x\in\mathbb{F}_2^n}2^n=2^{2n}.$$

■

Proposition 1.37. *It holds*

$$NL(f)\leq 2^{n-1}-2^{\frac{n}{2}-1}.$$

Proof. From Lemma 1.36 we obtain that

$$\max_{a\in\mathbb{F}_2^n}|\mathcal{W}_f(a)|\geq 2^{\frac{n}{2}}$$

and hence, thanks to Proposition 1.35

$$NL(f)=2^{n-1}-\frac{1}{2}\max_{a\in\mathbb{F}_2^n}|\mathcal{W}_f(a)|=2^{n-1}-2^{\frac{n}{2}-1}.$$

■

Definition 1.38. *If* $\mathrm{NL}(f)=2^{n-1}-2^{\frac{n}{2}-1}$ *then* f *is called bent.*

Remark 1.39. If f is bent then

$$\{\mathcal{W}_f(a):a\in\mathbb{F}_2^n\}=\left\{\pm 2^{\frac{n}{2}}\right\}.$$

Since $\mathcal{W}_f(a)\in\mathbb{Z}$ must be an integer bent functions only exists for even values of n.

Example 1.40. Let $n=2$, $f(x)=x_1x_2$ is bent. We have $\mathrm{NL}(f)\geq 1$ since $\mathrm{NL}(f)=0$ would imply f affine but we know that $\deg_A(f)=2$ since $f(11)=1$. Moreover $d(f,0)=1=NL(f)$. But we have

$$2^{n-1}-2^{\frac{n}{2}-1}=2^{2-1}-\frac{1}{2}2^1=1$$

and hence f is bent.

Definition 1.41. *Let* $f\in\mathcal{B}_n$ *bent. We define the dual boolean function of* f *to be the function* $\widetilde{f}$ *such that*

$$\frac{\mathcal{W}_f(a)}{2^{\frac{n}{2}}}=(-1)^{\widetilde{f}(a)}.$$

Notice that this is well defined since for f bent the left product is always ± 1.

Proposition 1.42. *If f is bent then also $\widetilde{f}$ is bent. Moreover $\widetilde{\widetilde{f}} = f$.*

Proof.

$$\mathcal{W}_{\widetilde{f}}(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{\widetilde{f}(x)+a\cdot x} =$$

$$\sum_{x \in \mathbb{F}_2^n} (-1)^{a\cdot x} \frac{\mathcal{W}_f(x)}{2^{n/2}} = \frac{1}{2^{n/2}} \sum_{x \in \mathbb{F}_2^n} (-1)^{a\cdot x} \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)+x\cdot y} = \tag{1.3}$$

$$= \frac{1}{2^{n/2}} \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)} \sum_{x \in \mathbb{F}_2^n} (-1)^{(a+y)\cdot x}.$$

Like above, in the second summation we get 2^n if $a + y = 0$ and 0 otherwise. From that follows

$$(1.3) = \frac{2^n}{2^{n/2}} (-1)^{f(a)} = (-1)^{f(a)} 2^{n/2},$$

and hence $\widetilde{f}$ is bent. It is then clear from the definition that $\widetilde{\widetilde{f}} = f$. ∎

Definition 1.43. *The derivative of f with direction a is*

$$\mathrm{D}_a f(x) = f(x + a) + f(x).$$

Proposition 1.44. *$f \in \mathcal{B}_n$ is bent $\iff$ $\forall a \neq 0$, $\mathrm{D}_a f$ is balanced.*

Proposition 1.45. *Let $\varphi : \mathbb{F}_2^n \to \mathbb{R}$. Such a function is called pseudo-boolean. Let us consider its Fourier transform*

$$\mathcal{F}_\varphi(a) = \sum_{x \in \mathbb{F}_2^n} \varphi(x)(-1)^{a\cdot x}.$$

Let $E < \mathbb{F}_2^n$ a subspace, with $\dim(E) = K$. Then

$$\sum_{x \in E} \mathcal{F}_\varphi(x) = 2^k \sum_{x \in E^\perp} \varphi(x).$$

Proof. Given l a linear boolean function such that $l\big|_E \neq 0$, then $l\big|_E$ is balanced. Let us define $H_l = \{v \in \mathbb{F}_2^n : l(v) = 0\}$ an hyperplane. It holds $\dim(H_l) = n-1$. Since $l\big|_E \neq 0$, we have $\dim(H_l \cap E) = k - 1$. Then it holds

$$\sum_{x \in E} \mathcal{F}_\varphi(x) = \sum_{x \in E} \sum_{y \in \mathbb{F}_2^n} \varphi(y)(-1)^{y\cdot x} = \sum_{y \in \mathbb{F}_2^n} \varphi(y) \sum_{x \in E} (-1)^{l_y(x)}.$$

If $y \in E^\perp$ then $l_y(x) = 0$ for all $x \in E$, otherwise $l_y\big|_E$ is balanced. The sum is then $2^k \sum_{y \in E^\perp} \varphi(y)$ as desired. ∎

Theorem 1.46. *Let $f \in \mathcal{B}_n$ bent. Then $\deg_A(f) \leq \frac{n}{2}$.*

Proof. $f(x) = \sum_{I \subseteq [n]} a_I x^I$, where

$$a_I = \sum_{\substack{x \in \mathbb{F}_2^n \\ supp(x) \subseteq I}} f(x).$$

We can see f as a pseudo-boolean function. We now want to show that

$$\sum_{\substack{x \in \mathbb{F}_2^n \\ \mathrm{supp}(x) \subseteq I}} f(x) = 2^{|I|-1} - 2^{\frac{n}{2}-1} + 2^{|I|-\frac{n}{2}} \sum_{\substack{x \in \mathbb{F}_2^n \\ \mathrm{supp}(x) \subseteq I^C}} \widetilde{f}(x). \tag{1.4}$$

Let us define $\varphi(x) = 1 - 2f(x) = (-1)^{f(x)}$. Thanks to Proposition 1.45

$$\sum_{\substack{x \in \mathbb{F}_2^n \\ \mathrm{supp}(x) \subseteq I^C}} \mathcal{F}_\varphi(x) = 2^{|I^C|} \sum_{\mathrm{supp}(x) \subseteq I} . \tag{1.5}$$

Moreover by definition

$$\begin{gathered} \mathcal{F}_\varphi(x) = \sum_{y \in \mathbb{F}_2^n} (-1)^{f(x)+x \cdot y} = \\ \mathcal{W}_f(x) = 2^{n/2}(-1)^{\widetilde{f}(x)} = 2^{n/2}(1 - 2\widetilde{f}(x)). \end{gathered} \tag{1.6}$$

Rewriting (1.5) we obtain

$$\begin{gathered} \sum_{\mathrm{supp}(x) \subseteq I} \varphi(x) = 2^{\frac{n}{2}-|I^C|} \sum_{\mathrm{supp}(x) \subseteq I^C} (1 - 2\widetilde{f}(x)) \Rightarrow \\ \sum_{\mathrm{supp}(x) \subseteq I} \varphi(x) = 2^{|I|} - 2 \sum_{\mathrm{supp}(x) \subseteq I} f(x) = 2^{\frac{n}{2}} - 2^{\frac{n}{2}-|I^C|+1} \sum_{\mathrm{supp}(x) \subseteq I^C} \widetilde{f}(x) \Rightarrow \\ 2^{|I|-1} - 2^{\frac{n}{2}-1} + 2^{|I|-\frac{n}{2}} \sum_{\mathrm{supp}(x) \subseteq I^C} \widetilde{f}(x) = \sum_{\mathrm{supp}(x) \subseteq I} f(x) \end{gathered}$$

Taking the last sum *mod* 2 we obtain the coefficient a_I. Let us suppose that $|I| > n/2$. Then the sum in (1.4) is even, and that implies $a_I = 0$. We than have $\deg_A(f) \leq \frac{n}{2}$. ∎

Definition 1.47. *We say that $f, g \in \mathcal{B}_n$ are affine equivalent (we denote it by $f \sim_A g$) if there exists $L : \mathbb{F}_2^n \to \mathbb{F}_2^n$ linear permutation, $b \in \mathbb{F}_2^n$ and $\epsilon \in \mathbb{F}_2$ such that*

$$g(x) = f(L(x) + b) + \epsilon.$$

This is clearly an equivalence relation.

Proposition 1.48. *Let $f \sim_A g \in \mathcal{B}_n$. Then,*

1. *$\{|\mathcal{W}_f(a)|\} = \{|\mathcal{W}_g(a)|\}$; in particular $\mathrm{NL}(f) = \mathrm{NL}(g)$ and hence f is bent if and only if g is;*
2. *$\deg_A(f) = \deg_A(g)$;*

3. if $\epsilon = 0$ then $w(f) = w(g)$, otherwise $w(f) = 2^n - w(g)$.

Proof. (1) Let $g(x) = f(L(x) + b) + \epsilon$. We have

$$\begin{aligned} &\mathcal{W}_g(a) = \\ &\sum_{x \in \mathbb{F}_2^n} (-1)^{f(L(x)+b)+\epsilon+a\cdot x} \overset{(x \mapsto L^{-1}(x)+L^{-1}(b))}{=} \\ &\sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+\epsilon+a\cdot L^{-1}(x)+a\cdot L^{-1}(b)} = \\ &(-1)^{\epsilon+a\cdot L^{-1}(b)} \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+a'\cdot x} \end{aligned} \tag{1.7}$$

where $a' = L^{-1^*}(a)$, and L^{-1^*} is the adjunct operator of L^{-1}, such that

$$x \cdot L^{-1}(y) = L^{-1^*}(x) \cdot y$$

$\forall x, y$. It follows

$$(1.7) = (-1)^{\epsilon+a\cdot L^{-1}(b)} \mathcal{W}_f(a'),$$

and, since both inverse and adjunct are bijective we obtain $\{|\mathcal{W}_f(a)|\} = \{|\mathcal{W}_g(a)|\}$. (3) is clear from definition, while the proof of (2) is out of the scope of this work. ∎

1.3 Vectorial Boolean Functions

Definition 1.49. *$F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is called an (n, m)-vectorial Boolean function.*

Through this section, if not stated otherwise, we will assume $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$.

Remark 1.50. We can also write $F = (f_1, ..., f_m)$ where the $f_i : \mathbb{F}_2^n \to \mathbb{F}_2$ are called coordinate functions.

Definition 1.51. *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. For all $v \in \mathbb{F}_2^m \setminus \{0\}$ we can define the Boolean function*

$$v \cdot F(x) = \sum_{i=1}^{m} v_i f_i(x).$$

$v \cdot F$ is called component of F.

Definition 1.52 (Algebraic Normal Form)**.** *Like we did for the scalar case we can define the algebraic normal form for F as*

$$F(x) = \sum_{I \subseteq [n]} a_I x^I,$$

with $a_I = (a_I^{(1)}, ..., a_I^{(m)}) \in \mathbb{F}_2^m$. In this notation

$$F_i(x) = \sum_{I \subseteq [n]} a_I^{(i)} x^I.$$

Remark 1.53. Again like in the scalar case, it holds

$$a_I = \sum_{\substack{x \in \mathbb{F}_2^n \\ \mathrm{supp}(x) \subseteq I}} F(x).$$

Definition 1.54. *Let*

$$F = \sum_{I \subseteq [n]} a_I x^I$$

be the algebraic normal form of F. We define the algebraic degree of F as $\deg_A(F) = \max_i \deg_A(f_i)$.

Remark 1.55. Notice that if F is an (n,n)-function we can see it as $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$.

Theorem 1.56. *Let $q \in \mathbb{N}$ and $F : \mathbb{F}_q \to \mathbb{F}_q$. Then there exists a unique $P \in \mathbb{F}_q[X]$ such that $\deg(P) < q$ and $F(v) = P(v)$ on $\mathbb{F}_q$. In particular, this means that any $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ can be represented as a polynomial of degree at most $2^n - 1$.*

Proof. Let $\mathbb{F}_q = \{0, 1 = \alpha^{q-1}, \alpha, ..., \alpha^{q-2}\} = \{a_0, ..., a_{q-1}\}$ where α is a primitive element. We have $F(0) = F(a_0) = b_0$ and $F(a_i) = b_i$. Then, by Lagrange interpolation, we obtain the polynomial

$$P(x) = \sum_{i=0}^{q-1} b_i \prod_{\substack{j=0 \\ j \neq i}}^{q-i} \frac{x - a_j}{a_i - a_j}.$$

■

Proposition 1.57. *Let $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$. We may write F as*

$$F(x) = \sum_{i=0}^{2^n - 1} a_i x^i.$$

Then it holds $\deg_A(F) = \max\{w(i) : a_i \neq 0\}$, where $w(i)$ is the Hamming weight of the binary vector associated with the integer i.

Definition 1.58. *$F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ is said to be balanced if for all $u \in \mathbb{F}_2^m$ it holds $|F^{-1}(u)| = 2^{n-m}$.*

Proposition 1.59. *F is balanced $\iff$ $v \cdot F$ is balanced $\forall v \in \mathbb{F}_2^m, v \neq 0$.*

Corollary 1.60. *$F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is balanced $\iff$ it is bijective (or equivalently a permutation).*

Definition 1.61. *We define the non-linearity of a vectorial Boolean function F as*

$$\mathrm{NL}(F) = \min_{v \in \mathbb{F}_2^n \setminus \{0\}} \mathrm{NL}(v \cdot F).$$

Definition 1.62. *The Walsh coefficient of F computed in $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^m$ is given by*

$$\mathcal{W}_F(a,b) = \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot x + b \cdot F(x)}.$$

Remark 1.63. We also have $\mathcal{W}_F(a,b) = \mathcal{W}_{b \cdot F}(a)$.

Proposition 1.64. *Like for the scalar case it holds*

$$\mathrm{NL}(F) = 2^{n-1} - \max_{\substack{a \in F_2^n \\ 0 \neq b \in \mathbb{F}_2^m}} \frac{|\mathcal{W}_f(a,b)|}{2} \leq 2^{n-1} - 2^{\frac{n}{2}-1}.$$

Definition 1.65. *F is called bent if the above bound is attained, i.e.*

$$\mathrm{NL}(F) = 2^{n-1} - 2^{\frac{n}{2}-1}.$$

Definition 1.66. *We can define again the derivative of F with direction a as*

$$D_a F(X) := F(x+a) + F(x)$$

Remark 1.67. It holds $D_a(v \cdot F)(x) = v \cdot D_a F$.

Proposition 1.68. *F is bent $\iff$ $v \cdot F$ is bent $\forall v \in \mathbb{F}_2^m, v \neq 0$.*

Corollary 1.69. *Like in the scalar case F is bent $\iff$ $D_a F(X)$ is balanced $\forall a \neq 0$.*

Proof. It holds

$$D_a(v \cdot F)(x) = v \cdot F(x+a) + v \cdot F(x) = v \cdot (F(x+a) + F(x)) = v \cdot D_a.F.$$

The result the follows from the scalar case applied to each component. ∎

Theorem 1.70. *F bent $\Rightarrow m \leq \frac{n}{2}$.*

Proof. If F is bent then $v \cdot F$ is bent for all $v \in \mathbb{F}_2^m$, $v \neq 0$. $v \cdot F$ bent implies that also $\widetilde{v \cdot F}$ is bent. Then by definition we have

$$\mathcal{W}_F(y,v) = \mathcal{W}_{v \cdot F}(y) = 2^{n/2} (-1)^{\widetilde{v \cdot F}(y)}.$$

Let us observe that for all (n,m)-function it holds, for fixed u,

$$\frac{1}{2^m} \sum_{\substack{x \in \mathbb{F}_2^n \\ v \in \mathbb{F}_2^m}} (-1)^{v \cdot (F(x)+u)} = \frac{1}{2^m} \left(\sum_{\substack{x \in \mathbb{F}_2^n \\ F(x)=u}} \sum_{v \in \mathbb{F}_2^m} (-1)^{v \cdot 0} + \sum_{\substack{x \in \mathbb{F}_2^n \\ F(x) \neq u}} (-1)^{v \cdot (F(x)+u)} \right).$$

The last sum is 0 since the exponent is balanced, while the first is $|F^{-1}(u)| 2^m$

and so in the end we get $|F^{-1}(u)|$. We then have

$$|F^{-1}(u)| = 2^{-m} \sum_{v\in\mathbb{F}_2^m} (-1)^{v\cdot u} \sum_{x\in\mathbb{F}_2^n} (-1)^{v\cdot F(x)} = 2^{-m} \sum_{v\in\mathbb{F}_2^m} (-1)^{v\cdot u} \mathcal{W}_f(0,v) =$$
$$= 2^{\frac{n}{2}-m} \sum_{0\neq v\in\mathbb{F}_2^m} (-1)^{v\cdot u+\widetilde{v\in F}(0)} + 2^{n-m}$$

since when $v \neq 0$ $\mathcal{W}_f(0,v) = 2^{n/2}(-1)^{\widetilde{v\cdot F(0)}}$, and the last term takes into account the case $v = 0$. Then let $N = \sum_{0\neq v\in\mathbb{F}_2^m} (-1)^{v\cdot u+\widetilde{v\cdot F}(0)}$. N is odd since we are summing ± 1 $2^n - 1$ times. Since $|F^{-1}(u)|$ must be an integer then $2^{\frac{n}{2}-m}$ is an integer and this implies $\frac{n}{2} \geq m$. ∎

Theorem 1.71 (Sidenilkov-Chabaud-Vadenay Bound)**.** *Let $m \geq n-1$. Then,*

$$\mathrm{NL}(F) \leq 2^{n-1} - \frac{1}{2}\sqrt{3\cdot 2^n - 2 - 2\frac{(2^n-1)(2^{n-1}-1)}{2^m-1}}.$$

In particular, if $n = m$ it holds

$$\mathrm{NL}(F) \leq 2^{n-1} - 2^{\frac{n-1}{2}}.$$

Proof. Notice that if we have $k_1, ..., k_r \geq 0$ it holds $\max_i k_i \geq \frac{\sum k_i^2}{\sum k_i}$.

$$NL(F) = 2^{n-1} - \frac{1}{2} \max_{\substack{a\in\mathbb{F}_2^n \\ 0\neq b\in\mathbb{F}_2^m}} |\mathcal{W}_F(a,b)|,$$

and this implies

$$\max_{\substack{a\in\mathbb{F}_2^n \\ 0\neq b\in\mathbb{F}_2^m}} |\mathcal{W}_F^2(a,b)| \geq \frac{\sum \mathcal{W}_F^4(a,b)}{\sum \mathcal{W}_F^2(a,b)}.$$

Thanks to Lemma 1.36 we notice that $\forall b \in \mathbb{F}_2^m$ it holds

$$\sum_{a\in\mathbb{F}_2^n} \mathcal{W}_F^2(a,b) = 2^{2n} \tag{1.8}$$

since if $b \neq 0$ then $\sum_{a\in\mathbb{F}_2^n} \mathcal{W}_F^2(a,b) = \sum_a \mathcal{W}_{b\cdot F}^2(a) = 2^{2n}$ and the case $b = 0$ can be

directly computed. Moreover

$$\begin{aligned}
&\sum_{a,b} \mathcal{W}_F^4(a,b) = \\
&= \sum_{a,b} \left(\sum_{x \in \mathbb{F}_2^n} (-1)^{b \cdot F(x) + a \cdot x} \right) \left(\sum_{y \in \mathbb{F}_2^n} (-1)^{b \cdot F(y) + a \cdot y} \right) \cdot \\
&\cdot \left(\sum_{z \in \mathbb{F}_2^n} (-1)^{b \cdot F(z) + a \cdot z} \right) \left(\sum_{t \in \mathbb{F}_2^n} (-1)^{b \cdot F(t) + a \cdot t} \right) = \\
&= \sum_{x,y,z,t \in \mathbb{F}_2^n} \left(\sum_{b \in \mathbb{F}_2^m} (-1)^{b \cdot (F(x)+F(y)+F(z)+F(t))} \right) \left(\sum_{a \in \mathbb{F}_2^n} (-1)^{a \cdot (x+y+z+t)} \right).
\end{aligned} \tag{1.9}$$

If $F(x)+F(y)+F(z)+F(t) \neq 0$ the second sum is 0, otherwise 2^m. The same holds for the third one with $x+y+z+t$. We then have

$$\begin{aligned}
(1.9) &= 2^{n+m} |\{(x,y,z,t) \in \mathbb{F}_2^{4n} : x+y+z+t = 0, \\
&\qquad F(x)+F(y)+F(z)+F(t) = 0\}| = \\
&= 2^{n+m} \, |\{(x,y,z) : F(x)+F(y)+F(z)+F(x+y+z) = 0\}| \geq \\
&\geq 2^{n+m} \, |\{(x,y,z) : x = y \circ x = z \circ y = z\}| = \\
&2^{n+m} 3 \, |\{(x,x,y) : x,y \in \mathbb{F}_2^n\}| - 2 \, |\{(x,x,x) : x \in \mathbb{F}_2^n\}| = \\
&= 2^{n+m}(2^{2n}3 - 2^n 2).
\end{aligned}$$

But then

$$\begin{aligned}
\max_{\substack{a \in \mathbb{F}_2^n \\ b \in \mathbb{F}_2^m}} \mathcal{W}_F^2(a,b) &\geq \frac{2^{n+m}(2^{2n}3 - 2^n 2) - 2^{4n}}{(2^m - 1)2^{2n}} = \\
&= 2^n 3 - 2 - 2\frac{(2^n - 1)(2^{n-1} - 1)}{2^m - 1},
\end{aligned}$$

where 2^{4n} comes from the case $b = 0$, and the denominator from (1.8). ∎

Definition 1.72. *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. If the SCV bound is attained, i.e.*

$$\mathrm{NL}(F) = 2^{n-1} - 2^{\frac{n-1}{2}},$$

F is called Almost Bent (AB).

Remark 1.73. There extist AB functions for odd values of n.

Proposition 1.74. *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ an almost bent function. Then it holds*

$$\deg_A(F) \leq \frac{n+1}{2}.$$

Proof. Let us assume that there exist $b \neq 0$ such that $d = \deg_A(b \cdot F) > \frac{n+1}{2}$. Then in the algebraic normal form of $b \cdot F$ we will have a monomial x^I where $|I| = d$. We may suppose, up to index permutation, that $I = \{1, ..., d\}$, and

$x^I = x_1 \cdots x_d$. Let us call $E = \{u \in \mathbb{F}_2^n : u_i = 0, i \in I\}$. Clearly $\dim(E) = n-d$. Then

$$b \cdot F\big|_{E^\perp} = \sum_{S \subseteq I} a_S x^S + x^I$$

where the a_S are the coefficients of the algebraic normal form of $b \cdot F$. But $E^\perp \cong \mathbb{F}_2^d$, and hence

$$w\big|_{E^\perp}(b \cdot F\big|_{E^\perp}) = 2h + 1$$

is odd. This implies

$$\left|\left\{x \in E^\perp : b \cdot F(x) = 0\right\}\right| = 2^d - (2h+1).$$

But then

$$\sum_{a \in E} \mathcal{W}_{bF}(a) = 2^{n-d} \sum_{a \in E^\perp} (-1)^{bF(a)} \tag{1.10}$$

thanks to Proposition 1.45 applied to the pseudo-boolean function $(-1)^{bF}$. It follows

$$(1.10) = 2^{n-d}(2^d - (2h+1) - (2h+1)) = 2^{n-d}2(2^{d-1} - 2h - 1).$$

The content of the last parenthesis is odd, and hence $2^{n-d+1} | \sum_{a \in E} \mathcal{W}_F(a,b)$ and $2^{n-d+2} \nmid \sum_{a \in E} \mathcal{W}_F(a,b)$. But all the nonzero Walsh coefficients must be equal to $2^{\frac{n+1}{2}}$, and then for all $k \leq \frac{n+1}{2}$ it holds $2^k | \mathcal{W}_F(a,b)$. But now $d > \frac{n+1}{2}$ implies $n - d + 2 \leq \frac{n+1}{2}$, which is absurd for what we just observed, and hence $d \leq \frac{n+1}{2}$. ∎

Definition 1.75. *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. We define*

$$\delta_F(a,b) = |\{x : \mathrm{D}_a F(x) = b\}|$$

and the differential uniformity of F

$$\delta(F) := \max_{\substack{a \neq 0 \\ b \in \mathbb{F}_2^m}} \delta_F(a,b)$$

for $0 \neq a \in \mathbb{F}_2^n$, $b \in \mathbb{F}_2^m$. F is called δ-uniform (or differentially δ-uniform) where $\delta = \delta(F)$.

Remark 1.76. In general it holds $2^{n-m} \leq \delta_F$, since

$$2^n = \sum_{b \in \mathrm{Im}(\mathrm{D}_a F)} \left|\mathrm{D}_a F^{-1}(b)\right| \leq \sum_{b \in \mathrm{Im}(\mathrm{D}_a F)} \delta(F) \leq \delta(F) 2^m.$$

For $n = m$ we obtain $\delta(F) \geq 1$. However, this bound cannot be attained, since if x is a solution for $F(x+a) + F(x) = b$ also $x + a$ does. So for $m = n$ it holds $\delta(F) \geq 2$.

Definition 1.77. *If $\delta(F) = 2$ we call $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ an Almost Perfect Nonlinear (APN) function.*

Remark 1.78. Notice that for linear and affine F it holds

$$F(x+a)+F(x)=F(x)+F(a)+F(x)=F(a).$$

which implies $\delta(F)=2^n$.

Definition 1.79. *Two functions $F,G:\mathbb{F}_2^n\to\mathbb{F}_2^n$ are said to be affine equivalent if there exist $A_1,A_2:\mathbb{F}_2^n\to\mathbb{F}_2^n$ affine permutations such that*

$$G(x)=A_2\circ F\circ A_1(x).$$

We denote this by $F\sim_A G$, just like in the scalar case.

Proposition 1.80. *For $F\sim_A G$ we have*

1. $\deg_A(F)=\deg_A(G)$;
2. $\mathrm{NL}(F)=\mathrm{NL}(G)$, *and in particular F is almost bent $\iff$ G is;*
3. $\delta(F)=\delta(G)$, *and in particular F is APN $\iff$ G is.*

Proposition 1.81. *Let $F:\mathbb{F}_2^n\to\mathbb{F}_2^n$ be an APN function. Then, it holds* $\mathrm{NL}(F)>0$.

Proof. We may suppose, without loss of generality, $F(0)=0$ since affine equivalences does not change non linearity. Suppose by absurd that $NL(F)=0$. Then, there exists $v\neq 0$ such that vF is linear. Up to a basis change, let $v=e_1$. We then have $F=(f_1,...,f_n)$ with f_1 linear, and $D_aF(x)=(f_1(a),y)$ with $f_1(a)$ constant. This implies

$$\operatorname{Im} D_aF=\left\{(f_1(a),y):y\in\mathbb{F}_2^{n-1}\right\}.$$

Let us now consider $F'=(f_2,...,f_n)$. We know that $\operatorname{Im} D_aF'=\mathbb{F}_2^{n-1}$, and since each vector in the image of D_aF is in correspondence with either 2 or 0 elements of the preimage (since F is APN) we have $\left|(D_aF')^{-1}(b)\right|=2$ for all $b\in\mathbb{F}_2^{n-1}$. This implies that D_aF' is balanced for all $a\neq 0$ and hence F' is bent. But this is absurd, since bent functions goes from $\mathbb{F}_2^n\to\mathbb{F}_2^m$ with $m\leq\frac{n}{2}$. ∎

Chapter 2

Cryptosystems

2.1 Security of a cryptosystem

Cryptography is the discipline which studies and designs ciphers that enable two parties to communicate in the presence of an eavesdropper that can monitor all communication between them. In the private key setting, the one we will consider in this work, the ciphers rely on the assumption that the two parties are able to share in advance a secret key, unknown to the eavesdropper. Then one party can encrypt the message, also called plaintext, using the secret key, thus obtaining an obfuscated message called ciphertext that is transmitted to the receiver. The receiver uses the same key to decrypt the ciphertext and obtain the message. On the other hand, the public key setting is used to study communication that cannot rely on a previously shared secret key between the parts, and hence must be protected in a different ways. Public key encryption is often used in order to share, through an insecure channel, the secret private key. We will now provide some formal definitions.

Definition 2.1. *A cryptosystem is a quintuple* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *of finite sets such that for all* $k \in \mathcal{K}$ *there exist* $c_k \in \mathcal{E}, c_k : \mathcal{P} \to \mathcal{C}$ *and* $d_k \in \mathcal{D}, d_k : \mathcal{C} \to \mathcal{P}$ *such that* $d_k \circ c_k(p) = p$ *for every* $p \in \mathcal{P}$.
$\mathcal{E}$ *is the set of encryption functions,* $\mathcal{D}$ *the set of decryption functions,* $\mathcal{P}$ *is the set of plaintexts,* $\mathcal{C}$ *the set of ciphertexts, and* $\mathcal{K}$ *the key space.*

In the design of a cryptosystem it is assumed that the only information not known by the eavesdropper is the secret key used by the two parties: a key, if discovered, is easier to change than designing a whole new cryptosystem. This is known as Kerckhoffs' principle. Informally, a cryptographic system is called computationally secure if the best possible algorithm for breaking it requires N operations, where N is large enough to be infeasible in reasonable time. However, under this definition no actual system can be proved secure, since we never know whether there is a better algorithm than the ones known. Hence, in practice we say a system is computationally secure if the best known algorithm for breaking it requires an unreasonably large amount of computational resources. Of course in this setting a cryptosystem can only be considered secure against an adversary whose computational resources are bounded. A system is said to

be unconditionally secure (or information-theoretically secure) when we place no limit on the computational power of the adversary.
Let us put a distribution probability on $\mathcal{P}, \mathcal{C}$ and $\mathcal{K}$. We will denote all of them by $\mathbb{P}$ with an abuse of notation. Denoting by X a random variable for plaintext, Y for ciphertext and K for the key, we assume $\mathbb{P}(X = x) > 0$ for all $x \in \mathcal{P}$ and $\mathbb{P}(K = k) > 0$ for all $k \in \mathcal{K}$.

Definition 2.2 (Perfect secrecy)**.** *A cryptosystem is said to have perfect secrecy if* $\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x)$.

Lemma 2.3. *Let* $\mathbb{P}(Y = y) > 0$*, if* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *has perfect secrecy, then for all* $x \in \mathcal{P}$ *and for all* $y \in \mathcal{C}$ *there exists* $k \in \mathcal{K}$ *such that* $c_k(x) = y$*. Moreover,* $|\mathcal{K}| \geq |\mathcal{P}|$.

Proof. Since the cryptosystem has perfect secrecy, it holds $\mathbb{P}(x|y) = \mathbb{P}(x)$ and then for Bayes Theorem $\mathbb{P}(y|x) = \mathbb{P}(y)$. Let us fix $x \in \mathcal{P}$ and $y \in \mathcal{C}$. Then there exists $k \in K$ such that $e_k(x) = y$, otherwise for Bayes Theorem again we would have $\mathbb{P}(y) = \mathbb{P}(y|x) = 0$ and this is a contradiction.
For the second part, let us fix $x \in \mathcal{P}$. For all $y \in \mathcal{C}$ there exists $k \in \mathcal{K}$ such that $e_k(x) = y$. But then $\mathcal{C} = \{e_k(x) : k \in \mathcal{K}\}$ and this implies $|\mathcal{K}| \geq |\mathcal{C}|$. But e_k is injective, since it has left inverse d_k, and hence $|\mathcal{C}| \geq |\mathcal{P}|$. ∎

Theorem 2.4 (Shannon)**.** *Let* $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$*. A cryptosystem* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *has perfect secrecy if and only if*

1. $\forall k \in \mathcal{K}$, $\mathbb{P}(k) = \frac{1}{|K|}$ *(i.e. all keys have the same probability)*
2. $\forall (x, y) \in \mathcal{P} \times \mathcal{C}$ $\exists! k \in \mathcal{K}$ *such that* $c_k(x) = y$.

Proof. ($\Rightarrow$) Let us start by proving point (2). Thanks to Lemma 2.3 we know that there exists such a k. Its uniqueness follows from the fact that $|\mathcal{K}| = |\mathcal{P}| = |\mathcal{C}|$. We can now move to condition (1). Let us fix $x \in \mathcal{P}$ and $y \in \mathcal{C}$. From condition (2) there exists a unique $k_{x,y}$ such that $e_k(x) = y$. Since y is fixed we have $\mathbb{P}(y) = \mathbb{P}(y|x) = \mathbb{P}(k_{x,y})$, for the uniqueness of k. If we consider $x \neq x'$ then there will exist two unique elements $k \neq k'$ such that $e_k(x) = e_{k'}(x')$. But then $\mathbb{P}(k) = \mathbb{P}(y) = \mathbb{P}(k')$ and hence $\mathbb{P}(k) = \frac{1}{|K|}$.
($\Leftarrow$) Let us fix $y \in \mathcal{C}$. Then for all $x \in \mathcal{P}$ there exists a unique k such that $e_k(x) = y$. But then we may write $\mathcal{P} = \{d_k(y) : k \in \mathcal{K}\}$, $\mathcal{K} = \{k : y \in e_k(\mathcal{P})\}$ and

$$\mathbb{P}(Y = y) = \sum_{k \in \mathcal{K}} \mathbb{P}(K = k)\mathbb{P}(d_k(y)) = \sum_{k \in \mathcal{K}} \frac{1}{|\mathcal{K}|}\mathbb{P}(d_k(y)) =$$
$$= \frac{1}{|\mathcal{K}|} \sum_{x \in \mathcal{P}} \mathbb{P}(x) = \frac{1}{|\mathcal{K}|}.$$

But then from condition (2) we have that for all $x \in \mathcal{P}$, $y \in \mathcal{C}$ it holds $\mathbb{P}(y|x) = \mathbb{P}(k) = \frac{1}{|\mathcal{K}|} = \mathbb{P}(y)$. Bayes theorem than implies $\mathbb{P}(x|y) = \mathbb{P}(x)$ and this concludes the proof. ∎

Example 2.5 (One Time Pad)**.** Let $m \geq 1$, $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbb{F}_2^m$. We can define for each $k \in \mathcal{K}$ $c_k(x) = x \oplus k$ and $d_k(y) = y \oplus k$ (where $\oplus$ denotes the

usual bitwise XOR). This cryptosystem is called One Time Pad. If the keys are equally probable, the OTP is able to achieve theoretical perfect secrecy. However, it has many problems in practice. First of all, we are forced to use keys that are as long as the message; key distribution hence can become a serious problem. Moreover, the same key cannot be used twice. Let's see how. Suppose that we have two parts, Alice and Bob, who want to communicate each other, and an eavesdropper Eve who tries to understand what they are saying. If Eve is able to obtain a pair (m, c) of plaintext and ciphertext then she can easily compute the key, since $k = m \oplus c$. If the key is reused, it is enough for Eve to perform this attack (called known-plaintext attack) once to have the key for all the messages. Even if Eve can only have access to different ciphertext c_1 and c_2 encrypted with the same key she can determine some partial information on the messages m_1 and m_2 since

$$c_1 \oplus c_2 = (m_1 \oplus k) \oplus (m_2 \oplus k) = m_1 \oplus m_2.$$

2.2 Block Ciphers

One of the most common symmetric cryptosystems that are used in modern cryptography are block ciphers. Block ciphers are so called because they act on block of fixed length. The message, generally represented by a binary string, is first split into substrings $m_1, ..., m_l$ each of length n. Each block, seen as an element of $(\mathbb{F}_2)^n$, is then encrypted by a function φ_k which depends on a secret key k. Using the above notation, we can give a formal definition of a block cipher.

Definition 2.6. *An algebraic block cipher is a cryptosystem* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *such that*

- $\mathcal{P} := \mathcal{C} := (\mathbb{F}_2)^n$*;*
- $\mathcal{K} = (\mathbb{F}_2)^l$*;*
- $\mathcal{E} := \{\varphi_k | k \in \mathcal{K}\} \subseteq \mathrm{Sym}((\mathbb{F}_2)^n)$*;*
- $\mathcal{D} := \left\{\varphi_k^{-1} | \varphi_k \in \mathcal{E}\right\}$

The space of plaintext and ciphertext is usually denoted by V*. We may also define a function*

$$\varphi : \mathcal{K} \times V \to V$$

such that $\varphi(k, x) = \varphi_k(x)$*; the definition block cipher is sometimes referred to this function.*

Definition 2.7. *A block cipher* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ *is an iterated block cipher if each* φ_k *is given by the composition of* $\varphi_{\psi(k,0)}, ..., \varphi_{\psi(k,r)}$*, with* r *fixed, where*

- $r \geq 1$*; each* $\varphi_{\psi(k,i)}$ *is called round or round function and hence* $r + 1$ *is the number of rounds*

- $\psi : (\mathbb{F}_2)^l \times \{0, ..., r\} \to (\mathbb{F}_2)^n$ *is called the key scheduling function;* k *is called master key while* $\psi(k, h)$ *is the* h*-th round key;*

- $\forall h \in \{0, ..., r\}$ *it holds* $\varphi_{\psi(k,h)} \in \mathrm{Sym}((\mathbb{F}_2)^n)$

From now on, with block cipher we will always mean iterated block cipher.

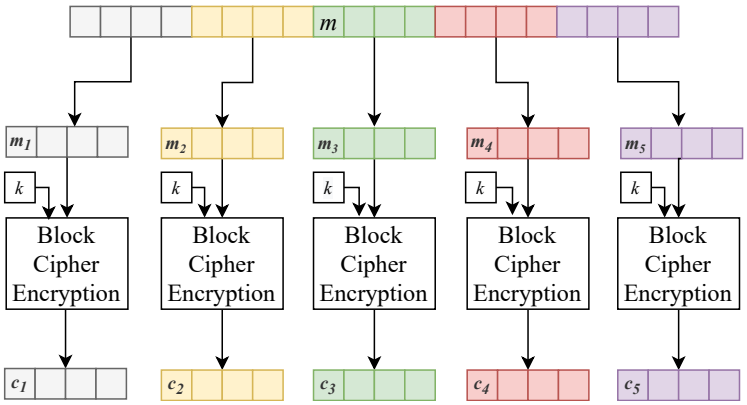

Figure 2.1: Generic scheme of a Block Cipher

A block cipher must behave like a random permutation. The space of possible permutations on n-bit strings is $2^n!$; it is then infeasible to represent them all. The real challenge when designing a cipher is construct a set of permutations with a concise description (namely, a short key) that behaves like a random permutation. In particular, just as evaluating a random permutation at two inputs that differ in only a single bit should yield two (almost) independent outputs, so too changing one bit of the input to a block cipher should yield an (almost) independent result. This implies that a one-bit change in the input should in some way affect every bit of the output.
In addition to his work on perfect secrecy, Shannon ([26]) also introduced a basic paradigm to achieve this goal. It is called the confusion-diffusion paradigm. Confusion means that each bit of the ciphertext should depend on several parts of the key. This property helps to hide the relationship between the ciphertext and the key. Diffusion means that if we change a single bit of the plaintext, then about half of the bits in the ciphertext should change. Similarly, if we change one bit of the ciphertext, then about half of the plaintext bits should change. If a ciphertext is designed properly following these two principles, it should be able to produce the desired effect, i.e. every bit of the output is affected by a small change in the input. This is also known as avalanche effect.

2.2.1 Substitution-Permutation Networks and the AES

A Substitution-Permutation Network (SPN) can be viewed as a direct implementation of the confusion-diffusion paradigm. Its peculiarity is that round functions are not chosen from the set of all possible permutations on some

domain. Instead, the $i-th$ round function can be written as

$$\varphi^i = \gamma\lambda\sigma_{k_i}$$

where

- $\gamma \in \mathrm{Sym}(V)$ (called confusion layer) is a nonlinear transformation which acts in parallel on smaller blocks of the message, made of j bits each (j is usually 4 or 8); if m is the message

  $$(m_1, ..., m_n)\gamma = ((m_1, ..., m_j)\gamma', ..., (m_{n-j+1}, ..., m_n)\gamma').$$

  The map $\gamma' \in \mathrm{Sym}(\mathbb{F}_2^j)$ is often called S-box.

- $\lambda \in \mathrm{Sym}(V)$ (called diffusion layer or mixing layer) is a linear map acting on the whole message.

- $\sigma_{k_i} : V \to V, x \mapsto x \oplus k_i$ represents the key addition, where $\oplus$ is the usual bitwise XOR. Here k_i stands for $\psi(k, i)$ for brevity.

As hinted by names, the aim of γ is to introduce confusion, while λ is responsible for diffusion. An easy heuristic way to introduce avalanche effect is to ensure that

- S-boxes are designed so that changing a single bit in the input of an S-box changes at least two bits in its output.

- The mixing layer is designed so that the output bits of any S-box is used to activate multiple S-boxes in the next round (an S-box is called active if it takes in input a nonzero element).

- Sufficiently many rounds are used.

Anyway, the choice of S-boxes must be really careful in order to avoid exposition to attacks. Moreover, if we want the avalanche effect to apply also to the inverse cipher, we may need to increase further the number of rounds.

The most used SPN is the so called Advanced Encryption Standard, or AES [16]. It is a cryptosystem selected by the United States National Institute of Standards and Technology (NIST) after a four year competition in which the best cryptographers and cryptanalysts from all over the world submitted a total of 15 different algorithms [24]. Each team's candidate cipher was intensively analyzed by members of NIST, the public, and especially the other teams. In October 2000, NIST announced that the winning algorithm was Rijndael (from the name of its designers, Vincent Rijmen and Joan Daemen). The process of selecting AES was ingenious because any group that submitted an algorithm, and was therefore interested in having its algorithm adopted, had strong motivation to find attacks on the other submissions. In this way, the world's best cryptanalysts focused their attention on finding even the slightest weaknesses in the candidate ciphers submitted to the competition. After only a few years each candidate algorithm was already subjected to intensive study,

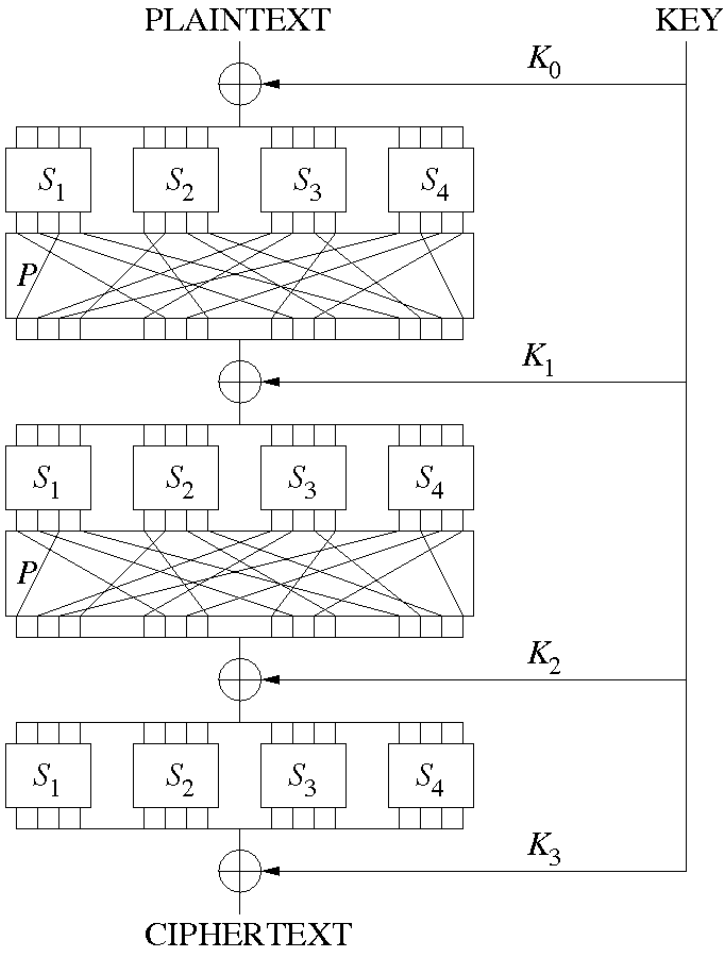

Figure 2.2: Example of Substitution Permutation Network

thus increasing our confidence in the security of the winning algorithm.
Let us briefly outline AES functioning, to give a good example of a block cipher. It is based on an array of bytes, called stated, which is seen as composed by blocks of 4. The state is initially set equal to the input to the cipher, which is 128 bits (or 16 bytes). The following operations are then applied to the state during each round:

1. **Add Round Key**: in every round, a 128-bit subkey is derived from the key schedule algorithm and is XORed to the state;

2. **Sub Bytes**: each byte of the state array is replaced by another byte according to a single fixed lookup table (the S-box) which is a bijection over $(\mathbb{F}_2)^8$;

3. **Shift Rows**: the state is now considered as a 4 by 4 square, and the bytes of each row are cyclically shifted as follows: the first row is untouched, the second one is shifted one place to the left, the third one is shifted two places to the left and the fourth one three places to the left;

4. **Mix Columns**: always looking at the state as a square, an invertible linear transformation is applied to the four bytes of each column; this transformation has the property that if two inputs differ in $b > 0$ bytes, then applying the transformation yields two outputs differing in at least $5 - b$ bytes.

In the final round, **Mix Columns** is replaced with **Add Round Key** because the last round would be otherwise invertible and hence useless.
To date, there are no practical cryptanalytic attacks that are significantly better than an exhaustive search for the key.

2.3 Attacks overview

Cryptanalysis is the science that studies cryptosystems, trying to identify weaknesses that could be exploited in an attack. The aim of the eavesdropper is to be able to retrieve either a part or the entire plaintext from a given ciphertext, some information about the key or even the key itself. Attack scenarios are generally classified depending on the information available to the eavesdropper. When talking about private key cryptosystems, and especially block ciphers, the cipher design is fixed and publicly available thanks to the aforementioned Kerckhoffs' principle while the key is assumed to be secret. What is then left to decide is the access to plaintext, ciphertext or pairs of related plaintext and ciphertext. We can classify the most common attacks as follows, basing on this discriminant.

- **Ciphertext Only**: the attacker can only observe some ciphertext, without the associated plaintext. This information is generally easy to obtain; however, a large sample of ciphertexts is needed to be able to decrypt the message. The second attack we presented on the One Time Pad is an example of ciphertext only attack.

- **Known Plaintext**: the attacker knows a certain amount of plaintext-ciphertext pairs. In general it is assumed that available plaintext are collected from a random sample. However, it is possible that they come from a non random distribution, showing redundancy that can be helpful in the decryption process. Sometimes the whole plaintext for a specific ciphertext is not available, but the attacker have some information about it. This happened for example in the famous Enigma decryption during World War II, when Allies made extensive use of information or guesses about occurrence of specific terms in German messages. The principal example of known plaintext attack is linear cryptanalysis.

- **Chosen Plaintext**: the attacker can (possibly adaptively) ask for the ciphertexts of arbitrary plaintext messages. This is formalized by allowing the adversary to interact with an encryption oracle, viewed as a black box. Notable examples are differential cryptanalysis, where the attacker aims to recover the key or part of it, but also the well known byte-at-a-time AES-ECB decryption, that allows the adversary to obtain the full plaintext without knowing anything about the key.

- **Chosen Ciphertext**: the attacker can choose one (or some) ciphertext and obtain the correspondent plaintext. Of course if the attacker is able to obtain the decryption of any ciphertext at any moment the system is broken and there is no need of attack. Chosen ciphertext attack are hence generally based on the availability of a limited number of (adaptive or not) queries to the decryption oracle. From here comes the notion of lunchtime attack, which refers to the idea that a user's computer, with the ability to decrypt, is available to the attacker while the user is out to lunch. Moreover, if the target is a specific ciphertext, the attacker is not allowed to decrypt it.

There are other kind of attacks, not treated here, that do not rely on theoretical weaknesses of the cipher but instead exploit their poor implementation. The main possible outcomes of the attacks described above are classified according to the type of information recovered during them. Such outcomes are ordered from the least favorable for the attacker to the most one.

- **Distinguishing Attack**: the attacker is able to distinguish the encrypted data from random data. Shannon's principles of diffusion and confusion are particularly helpful in avoiding this kind of vulnerability (as well as the other ones).
- **Partial Key Recovery**: the attacker is able to get some information about the key, such as some bits of the key or other relations among them.
- **Global deduction**: the attacker finds a functionally equivalent algorithm for encryption and/or decryption with a fixed key k without requiring the knowledge of the key itself.
- **Key recovery (total break)**: The attacker is able to recover the secret key k.

Even if block ciphers are often relatively complicated, and hence difficult to analyze, it is often surprisingly easy to find attacks on most constructions. The two most common attacks among the ones cited above are linear and differential cryptanalysis. Every modern block cipher is designed with resistance against these attacks well in mind. We will now give a brief introduction to linear cryptanalysis, while differential cryptanalysis will be presented in greater detail in the next section, since it will play a very important role in the rest of this work.
Linear cryptanalysis [23] was developed my Matsui in the early 1990s. The basic idea is to search linear relationships between the input and output that occurs with higher probability than would be expected for a random permutation. For linear relationship we mean an expression like

$$x_{i_1} \oplus \cdots \oplus x_{i_l} \oplus y_{j_1} \oplus \cdots \oplus y_{j_m},$$

where (x, y) is a plaintext-ciphertext pair. If we let x run through all the possible plaintext, and map it to y via a random permutation, we expect an expression

like that to be zero about half of the times. For this reason, assuming uniform x and k, we define the linear bias of a set of bit positions $i_1, ..., i_l, j_1, ..., j_m$ as

$$\epsilon := \left| \mathbb{P}[x_{i_1} \oplus \cdots \oplus x_{i_l} \oplus y_{j_1} \oplus \cdots \oplus y_{j_m} = 0] - \frac{1}{2} \right|.$$

Matsui showed how a large enough bias in a cipher can be used to find the secret key. As noted above, one important feature of this attack is that it is a known plaintext attack. This is very significant, since for example an encrypted file can provide a huge amount of known plaintext. Matsui showed that DES ([27], the standard encryption algorithm then replaced by AES) can be broken with just 2^{43} known plaintext-ciphertext pairs.

2.4 Differential cryptanalysis

Differential cryptanalysis is generally attributed to Eli Biham and Adi Shamir [3], who discovered it in the late 1980s. However, apparently IBM and the National Security Agency of the United States were well aware of this kind of technique ([13]) more than ten years before, and decided not to reveal it for safety and political reasons. It exploits the fact that some input differences may propagate with unusually high or low probability during the encryption process, leading to a non-uniform distribution of the output differences. For this reason, differential cryptanalysis is usually a chosen plaintext attack.
Let us fix a vectorial Boolean function f on V.

Definition 2.8. *A differential over f is a pair (δ_I, δ_O) of elements of V with associated differential probability*

$$p_{(\delta_I,\delta_O),f} := \mathbb{P}\left[xf + (x+\delta_I)f = \delta_O\right] = \mathbb{P}\left[\mathrm{D}_{\delta_I} f(x) = \delta_O\right]$$

for x uniformly distributed in V.

It represents the probability that, given two vectors whose difference is δ_I, the difference after applying f becomes δ_O.

Remark 2.9. As already shown in Remark 1.78, if f is linear, we get $xf + (x+\delta_I)f = \delta_I f$ for every x and so the only possible differential is $(\delta_I, \delta_I f)$ with probability 1. Similarly, if f is a translation it holds $xf + (x+\delta_I)f = \delta_I$ and the only possible differential is (δ_I, δ_I).

Definition 2.10 (Difference Distribution Table)**.** *The Difference Distribution Table, or DDT, of a function f is defined as the integer table with entries*

$$DDT_f[\delta_I, \delta_O] = 2^n p_{(\delta_I,\delta_O),f} = \#\left\{x : xf + (x+\delta_I)f = \delta_O\right\}$$

for each possible differential $(\delta_I, \delta_O) \in V^2$.

Remark 2.11 (Differential uniformity)**.** *The differential uniformity of f is defined as*

$$\delta(f) := \max_{\delta_I \neq 0} DDT_f[\delta_I, \delta_O]$$

and the function f is said to be differentially δ-uniform if $\delta(f) = \delta$.

Notice that this is exactly Definition 1.75, with a slightly changed notation. It is clear from the definition that a low differential uniformity is a desirable property for a function in order to make a differential attack harder. From Remark 2.9 it follows that $\delta(f) = 2^n$ if f is a translation or a linear function, while we have already shown in Remark 1.76 that for a generic Boolean function f it holds $\delta(f) \geq 2$. We recall that if $\delta(f) = 2$, f is said Almost Perfect Nonlinear (APN).
Given a key space $\mathcal{K}$ and $1 \leq s \leq r$ (where r is the number of rounds), we denote by $\varphi_k^{(s)}$ the composition of the first s round function, with keys $k = k_1, ..., k_s$ generated through the key schedule algorithm with master key k.

Definition 2.12. *An s-round differential is a pair (δ_I, δ_O) whose corresponding expected probability is*

$$p_{(\delta_I,\delta_O)} := \mathbb{P}_{x,k}\left[x\varphi_k^{(s)} + (x + \delta_I)\varphi_k^{(s)} = \delta_O\right]$$

where x and k are uniformly distributed respectively on V and K.

Definition 2.13. *Given $1 \leq s \leq r$, an s-round differential trail for a differential (δ_I, δ_O) is an $(s+1)$-tuple $(\beta_0, ..., \beta_s)$ of intermediate differences at each round such that $\beta_0 = \delta_I$ and $\beta_s = \delta_O$.*

The probability of a given s-round differential (δ_I, δ_O) is obtained as the sum of the probabilities of its differential trails. Notice that for each differential trail, only the confusion layer requires a probabilistic analysis since the diffusion layer is linear and the key addition layer is a fixed translation. It can be shown (see for example [19] or [28]) that the number of pairs of known plaintext required for the attack to recover the key (or part of it) is proportional to the inverse of probability of the trail.
Modern ciphers (like AES) are designed in order to be resistant against this kind of attack. However, for its flexibility and effectiveness, the differential attack is still widely study in order to find more general attacks that can break even those ciphers considered secure against it. As an example, the idea behind this work is to perform differential attack studying differentials with respect to a new sum $\circ$ instead of the usual XOR. Of course such an approach can present some complications; for example diffusion layer and key addition are no more affine, and must be hence taken into consideration during the analysis.

Chapter 3

Classification of 4-bit permutations

S-boxes play a fundamental role for the security of nearly all modern block ciphers. As already observed, in Substitution-Permutation Networks S-boxes form the only non-linear part of the cipher. Therefore, S-boxes have to be chosen carefully to make the cipher resistant against all kinds of attacks. In real life application, small S-boxes are generally preferred, since they are much more efficient to implement in hardware. For this reason, many modern block ciphers use 4 or 8-bit S-boxes; for example, the aforementioned AES cipher relies on 8-bit S-boxes.
Even from the brief introduction we gave it is quite evident that given a Boolean S-box f, a high nonlinearity $\mathrm{NL}(f)$ and a low differential uniformity $\delta(f)$ are crucial properties in order to make f resistant against linear and differential cryptanalysis respectively. A more detailed presentation of this fact is given in [22]. Starting from here, we may affirm that some S-boxes are better than others. However, since there exists a total of $2^n!$ permutations on $\mathbb{F}_2^n$, an exhaustive search in order to determine the best S-boxes is infeasible even for not so large values of n. This is the case, for example, of the AES S-box, called the Inverse function. It is defined by identifying $\mathbb{F}_2^n$ with $\mathbb{F}_{2^n}$ as

$$I : x \mapsto x^{2^n-2}.$$

It can be shown ([21]) that the Inverse is AB (and thus APN) for odd n, and differentially 4-uniform for even n. No permutation is known with better resistance against linear and differential cryptanalysis for $n = 8$. However, it is still not clear if these values are optimal.
The situation is a bit different for $n = 4$. First of all, it is known that there are no APN permutations on $\mathbb{F}_2^4$ ([22]), i.e. no possible S-boxes with $\delta(f) = 2$. As already observed, since $f(x)+f(x+a) = f(x+a)+f(x+a+a)$ the differential uniformity must always be even. We then conclude that the minimal differential uniformity is 4. Moreover, as shown in [25] it must always hold $\max\{\mathcal{W}(f)\} \geq 8$ which implies $\mathrm{NL}(f) \leq 4$. Thanks to these observations, we can now give the following definition:

Definition 3.1. *Let $S : \mathbb{F}_2^4 \to \mathbb{F}_2^4$ be an S-box. If*

1. *S is a bijection,*
2. *$\mathrm{NL}(S) = 4$*
3. *and $\delta(S) = 4$*

we call S an optimal S-box.

Moreover, thanks to Proposition 1.80, both nonlinearity and differential uniformity are invariant under affine equivalence. This means that instead of checking all the $16! \sim 2^{44}$ possible permutation one can restrict to 302 equivalence classes. In [22] these classes are analyzed, and 16 of them are found to be made of optimal permutations. Notably one of those (the one named G_3) is affine equivalent to the Inverse function in dimension 4. The next table gives the hexadecimal table for a representative from each class:

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
G_0	0	1	2	D	4	7	F	6	8	B	C	9	3	E	A	5
G_1	0	1	2	D	4	7	F	6	8	B	E	3	5	9	A	C
G_2	0	1	2	D	4	7	F	6	8	B	E	3	A	C	5	9
G_3	0	1	2	D	4	7	F	6	8	C	5	3	A	E	B	9
G_4	0	1	2	D	4	7	F	6	8	C	9	B	A	E	5	3
G_5	0	1	2	D	4	7	F	6	8	C	B	9	A	E	3	5
G_6	0	1	2	D	4	7	F	6	8	C	B	9	A	E	5	3
G_7	0	1	2	D	4	7	F	6	8	C	E	B	A	9	3	5
G_8	0	1	2	D	4	7	F	6	8	E	9	5	A	B	3	C
G_9	0	1	2	D	4	7	F	6	8	E	B	3	5	9	A	C
G_{10}	0	1	2	D	4	7	F	6	8	E	B	5	A	9	3	C
G_{11}	0	1	2	D	4	7	F	6	8	E	B	A	5	9	C	3
G_{12}	0	1	2	D	4	7	F	6	8	E	B	A	9	9	C	5
G_{13}	0	1	2	D	4	7	F	6	8	E	C	9	5	B	A	3
G_{14}	0	1	2	D	4	7	F	6	8	E	C	B	3	9	5	A
G_{15}	0	1	2	D	4	7	F	6	8	E	C	B	9	3	A	5

With this reduction to 16 equivalence classes it is now easy to study additional criteria. As an example, in [22] we find a detailed analysis of algebraic degree of these equivalence classes. Recall that algebraic degree for a vectorial Boolean function F is defined as the highest algebraic degree of its components, i.e.

$$\max_{b \in V} \deg_A(b \cdot S).$$

As already observed, the set $\{b \cdot S : b \in V\}$ is invariant under affine equivalence, and clearly so is $\deg_A(F)$. High algebraic degree is often used as a criterion for good S-boxes. It is known that any 4-bit bijection must have degree smaller than 3. It is then interesting to observe that all the 16 equivalence classes of optimal S-boxes have degree exactly 3. Moreover, for 8 of them $\deg_A(b \cdot S) = 3$ for all $b \in V$. Again, one example of such an S-box is the Inverse function.

For each optimal class, the number of vectors b for which the algebraic degree of $b \cdot S$ is 2 or 3 are reported below:

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
$\deg_A(b \cdot S) = 2$	3	3	3	0	0	0	0	0
$\deg_A(b \cdot S) = 2$	12	12	12	15	15	15	15	15

	G_8	G_9	G_{10}	G_{11}	G_{12}	G_{13}	G_{14}	G_{15}
$\deg_A(b \cdot S) = 2$	3	1	1	0	0	0	1	1
$\deg_A(b \cdot S) = 2$	12	14	14	15	15	15	14	14

Another interesting topic is the resistance against the algebraic attack, introduced by Courtois and Pieprzyk ([15]). It is still not completely clear which conditions exactly enable this kind of attack. However, the main criterion to successfully mount an algebraic attack is the number of linear independent low degree equations that are fulfilled by the input and output values of the S-box. It can be shown that these optimal S-boxes are optimal also with respect to algebraic attacks, in the sense that each of them fulfills 21 quadratic equations, which is the minimum number for permutations in dimension 4.
In the last chapter, we will analyze these optimal permutations testing their resistance with respect to an alternative sum.

Chapter 4

Differential Cryptanalysis revised

We are now ready to introduce more extensively the idea of an alternative sum, where it comes from and what are its possible applications. We have already defined block ciphers, and especially Substitution-Permutation Networks, as ciphers generally consisting in iterated application of encryption functions in the form of

$$\varphi = \rho\sigma_k$$

where σ_k is the key addition, while ρ depends on the design of the cipher and is assumed to be fixed and publicly available. The security of the encryption process, i.e. the inability of a non-authorized party to recover the message, strongly relies on the way the function ρ is designed. We already suggested how attacks like linear and differential cryptanalysis can threat the security of a block cipher and presented some general (empirical and theoretical) rules to follow in order to reduce exposition to these attacks. A deeper study on the properties that a generic cipher function ρ must satisfy to be considered secure is out of the scope of this work. From now on, following the setting of [7] and [12], we may assume as a minimum and crucial requirement that $\rho \notin \mathrm{AGL}(V)$. This guarantees that $\langle \rho, \mathrm{T} \rangle$ (where T is the translation group with respect to the usual XOR sum), the well studied group of the round functions introduced in [14], is not the affine group $\mathrm{AGL}(V)$. Although it is rather easy to satisfy this requirement, it is much harder to prove that $\langle \rho, \mathrm{T} \rangle$ is not contained in any conjugate of $\mathrm{AGL}(V)$ in $\mathrm{Sym}(V)$. If this is the case, i.e. if there exists $g \in \mathrm{Sym}(V)$ such that $\langle \rho, \mathrm{T} \rangle < \mathrm{AGL}(V)^g$ then there exists an operation $\circ$ such that

$$\langle \rho, \mathrm{T} \rangle < \mathrm{AGL}(V, \circ),$$

which means that each encryption function is affine with respect to the operation $\circ$, a serious threat for the security of the cipher. Some example of attacks that can be performed in this case are shown in [12] and [9]. For this reason, from now on we will restrict our attention to the investigation of alternative operations $\circ$ on V such that $\mathrm{T} < \mathrm{AGL}(V, \circ)$. Moreover, we will always assume $\mathrm{T}_\circ < \mathrm{AGL}(V)$, since it guarantees faster computation. Those hypotesis are also

central in [7], from which we will recover many important results.
Our goal is to show that an SPN which can be considered secure against classic attacks (and especially differential cryptanalysis with respect to the usual sum) might be broken by means of differential cryptanalysis carried out with this newly introduced operation. For this purpose, many of the definitions given in Section 2.4 can be rephrased in this setting just replacing + by $\circ$.

Definition 4.1. *A $\circ$-differential over f is a pair (δ_I, δ_O) of elements of V with associated differential probability*

$$p^{\circ}_{(\delta_I,\delta_O),f} := \mathbb{P}\left[xf \circ (x \circ \delta_I)f = \delta_O\right]$$

for x uniformly distributed in V.

Definition 4.2. *We may then define a new difference distribution table as*

$$DDT^{\circ}_f[\delta_I, \delta_O] = 2^n p_{(\delta_I,\delta_O),f} = \#\left\{x : xf \circ (x \circ \delta_I)f = \delta_O\right\}$$

Definition 4.3. *Finally, the $\circ$-differential uniformity of a Boolean function f is defined as*

$$\delta^{\circ}(f) := \max_{\delta_I \neq 0} DDT^{\circ}_f[\delta_I, \delta_O].$$

Our aim is to detect ciphers and operations for which

$$p_{(\delta_I,\delta_O),f} := \mathbb{P}_{x,k}\left[x\varphi_k^{(s)} \circ (x \circ \delta_I)\varphi_k^{(s)} = \delta_O\right]$$

is higher than the standard probability $p_{(\delta_I,\delta_O)}$ for some differential (δ_I, δ_O). In order to do that, we need to investigate how our new operation deals with the different components of the cipher, whose behavior is well known only in terms of the standard XOR. Doing so we will also explain the ratio that lies behind some of the definitions given in the first chapter of this work, since they will be necessary to present and better understand those interactions.

4.1 Interaction with the key-addition layer

As shown in Remark 2.9, the classical differential attack can rely on the property that each + difference is maintained the same after the key is XORed. This is not the case when considering $\circ$-differences. Let's consider two input with difference Δ, denoted by x and $x \circ \Delta$. After the key addition, the difference becomes

$$(x + k) \circ ((x \circ \Delta) + k) =: \Delta^{\circ}.$$

It is easy to see that $\Delta^{\circ} = \Delta$ for each $x, k \in V$ if and only if $+ = \circ$. However, for $k \in W_{\circ}$ we can replace + with $\circ$ and the previous equation holds, which means that the key addition layer behaves as a tranlation also with respect to the $\circ$ difference. This shows the importance of the weak key space $W_{\circ}$ when studying a new sum, and also explains where the name weak key comes from.
Of course in the general case we cannot assume $k \in W_{\circ}$. This makes a further investigation on Δ° necessary in order to better understand how differences

propagates through the key addition layer. First of all, notice that we already assumed $T_+ < \mathrm{AGL}(V, \circ)$. This makes the key addition a key-dependent affine transformation with respect to $\circ$. Furhter details and explainations are given in the next chapter, however it can be proven (see Theorem 5.15) that the error committed when considering $\circ$ difference instead of $+$ differences lives in the error set $U_\circ$, which also happens to be a subset of $W_\circ$. Moreover (Proposition 5.18) Δ° is independent from the state x, and it holds

$$\Delta^\circ = \Delta + k \cdot \Delta.$$

By definition, $k \cdot \Delta \in U_\circ$, so the number of possible output differences is $|U_\circ|$ (one being Δ itself). Hence another important factor when trying to control the effect of the key addition layer is the dimension of $U_\circ$. Starting from the previous equation, we may introduce a new table, called Key Distribution Table, that will help us in the study of possible errors committed.

Definition 4.4. *The Key Distribution Table (KDT) of an operation $\circ$ is the integer table defined by*

$$KDT^\circ[\delta_I, \delta_O] := \#\left\{k \in V | \delta_I + k \cdot \delta_I = \delta_O\right\}.$$

Notice that if the input difference $\delta_I \in W_\circ$, no matter the key considered, the output difference after the key addition layer is δ_I with probability one. This is because, as previously stated

$$(x + k) \circ ((x \circ \delta_I) + k) = \delta_I + k \cdot \delta_I = \delta_I$$

since $\delta_I \in W_\circ$ implies

$$k \cdot \delta_I = k + \delta_i + k \circ \delta_I = k + \delta_i + k + \delta_i = 0.$$

If $\delta_I \notin W_\circ$, the output difference equals $\delta_I + k \cdot \delta_I$. The error may be zero, leading to the output difference δ_I (this is always the case e.g. when $k \in W_\circ$), or may be different from zero, leading to $\delta_I + u$ for some $u \in U_\circ$. A more detailed analysis of KDT, which will be covered in the next section, will suggest other important conditions that our operation $\circ$ should satisfy in order to make key addition layer analysis successful.

4.2 Interaction with the confusion layer

While in classical differential cryptanalysis differential probabilities are only induced by the confusion layer, in the previous section we illustrated that, with new operations, probabilities are also added by the key-addition layer. For the probability of a $\circ$-differential to be larger than the probability of a $+$-differential, we should either have trails with larger probabilities and/or more trails. The first goal can only be achieved if the values in the DDT of the S-box computed with respect to $\circ$ are larger than those in the classical DDT computed with respect to the XOR. Let us give an example where this is true.

Example 4.5 ([12])**.** Let $n = 3$ and $d = 1$. As we will see, there exists only one possible alternative sum on $V = (\mathbb{F}_2)^3$ satisfying our constraints, with defining matrix (see Theorem 5.13)

$$\Theta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We will denote it by $\diamond$. Details on the notation and on how sums with this operation are actually computed are given in the next chapter. For now, we just present the results obtained as an example. Let us define an S-box $\gamma : (\mathbb{F}_2)^3 \to (\mathbb{F}_2)^3$ by

x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
$x\gamma$	0_x	6_x	2_x	1_x	5_x	7_x	4_x	3_x

Here each vector is interpreted as a binary number, most significant bit first, and then represented using the hexadecimal notation. By computing the DDT of γ with respect to the classic XOR, we obtain the following result:

$+$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	8	·	·	·	·	·	·	·
1_x	·	·	2	2	·	·	2	2
2_x	·	2	2	·	2	·	·	2
3_x	·	2	·	2	2	·	2	·
4_x	·	2	2	·	·	2	2	·
5_x	·	2	·	2	·	2	·	2
6_x	·	·	·	·	2	2	2	2
7_x	·	·	2	2	2	2	·	·

As we can see, γ is APN with respect to classic $+$ operation. However, if we compute the DDT with respect to our new $\diamond$ sum, we obtain

$\diamond$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	8	·	·	·	·	·	·	·
1_x	·	·	·	4	·	·	4	·
2_x	·	·	4	·	·	·	·	4
3_x	·	4	·	·	·	4	·	·
4_x	·	4	·	·	·	4	·	·
5_x	·	·	4	·	·	·	·	4
6_x	·	·	·	·	8	·	·	·
7_x	·	·	·	4	·	·	4	·

We can see that γ is differentially 8-uniform with respect to $\diamond$; if the difference between two input is 6_x the difference between the outputs will be 4_x, no matter what the inputs are. This is a clear weakness that can easily be exploited by an attacker using this new sum instead of the classic one.

This example may seem an isolated coincidence. However, the last chapter of this work will be devoted to show how it is possible to obtain many results like this on optimal 4-bit S-boxes (in the sense of Definition 3.1).

4.3 Interaction with the diffution layer

The last fact we have to consider, when trying to transport differential cryptanalysis to our setting, is how $\circ$-differences propagate through the diffusion layer, which is, in our model, a $+$-linear map. The role of the diffusion layer, in the sense of keeping the cipher safe from differential attack, is to spread the differences as fast and far as possible in the block, i.e. to quickly activate as many S-boxes as possible. However, it does not have direct role in terms of differential probability when differentials are computed with respect to the XOR, since it is a XOR-linear map, and consequently (thanks to Remark 2.9 again) each $+$-differential is deterministic with respect to the diffusion layer. On the other hand though, in the case of $\circ$-differentials, an attacker willing to predict the output difference of the diffusion layer λ, given an input difference Δ, must determine the distribution of the elements of the kind of

$$x\lambda \circ (x \circ \Delta)\lambda \tag{4.1}$$

with λ which in general is not linear with respect to $\circ$. This results in a huge $\circ$-non-linear map with 2^n inputs, which will make further analysis non trivial. Moreover, unlike what we manage to do in the case of the key addition layer, it can be shown ([12]) that Equation 4.1 is not independent from the state x. It is then clear that a successful attack with respect to an alternative operation $\circ$ may rely on the linearity of the diffusion layer with respect also to $\circ$. For this reason, the structure of the subgroup $H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$ is one of the key features to take into consideration when studying a new sum. A lot of work has been done to better understand it in [12] and [7]. In the next chapter we will at first present the main results obtained in those two papers, and then show how we tried to extend them to slightly different settings and which results we have achieved.

Chapter 5

Analysis of $H_\circ$

As just observed, a successful differential attack on a Substitution-Permutation Network with an alternative sum $\circ$ must rely on the fact that the mixing layer λ is linear also with respect to $\circ$, i.e. $\lambda \in H_\circ$. For this reason, the analysis of $H_\circ$ plays a very important role in determining the success possibility of such an attack. This chapter is hence the heart of our work. In Section 5.1 some results on a generic $\circ$ operation and the related groups are presented. Those results, summarized from [12] and [7], are the starting point for the next sections. In Section 5.2 we show the properties (especially Theorem 5.28 and Theorem 5.30) that hold if we force the dimension of $W_\circ$ to be $n-2$, where n is the dimension of V. In the last two sections, we try to reproduce those results in different settings: in Section 5.3 we fix $d = n-3$, while in Section 5.4 we study a smaller sum that acts in parallel.

5.1 General results

We are now ready to summarize the main results from [12] and [7] for a generic alternative sum.

Definition 5.1. *A Jacobson radical ring is a ring $(V, +, \cdot)$ such that $(V, \diamond)$ is a group, where the operation $\diamond$ is defined as $a \diamond b = a + b + a \cdot b$ for each $a, b \in V$.*

Theorem 5.2 ([10])**.** *Let $\mathbb{K}$ be any field, and $(V, +)$ be a vector space of any dimension over $\mathbb{K}$. There is a one to one correspondence between*

1. *abelian regular subgroups of $\mathrm{AGL}(V, +)$;*
2. *commutative, associative $\mathbb{K}$-algebra structures $(V, +, \cdot)$ that one can impose on the vector space structure $(V, +)$ such that the resulting ring is radical.*

In this corrispondence, isomorphism classes of $\mathbb{K}$-algebras correspond to conjugacy classes of abelian regular subgroups of $\mathrm{AGL}(V, +)$, where the conjucation is under the action of $\mathrm{GL}(V, +)$.

Remark 5.3. The correspondence mentioned in the previous result may be written explicitly, proceedings as follows. Let $\mathcal{T} < \mathrm{AGL}(V)$. Thanks to Remark 1.7, it can be written in the form

$$\mathcal{T} = \{\tau_a | a \in V\}.$$

For each $a \in V$ there exists $M_{a,\mathcal{T}} \in \mathrm{GL}(V,+)$ and $\sigma_b \in T_+$ for some $b \in V$ such that

$$\tau_a = M_{a,\mathcal{T}}\sigma_b.$$

Since $\mathcal{T}$ is fixed for now, in order to keep the notation lighter $M_{a,\mathcal{T}}$ will be simply denoted by M_a. For any $a \in V$, let us define the map $\delta_a = M_a - \mathbb{1}_V$. Then, the operation $\cdot$ defined by $x \cdot a = x\delta_a$ is such that the structure $(V,+,\cdot)$ is a commutative $\mathbb{K}$-algebra and the resulting ring is radical. Moreover, since by definition $0\tau_a = a$ we have $a = 0\tau_a = 0M_a\sigma_b = b$, which implies $\tau_a = M_a\sigma_a$ for each $a \in V$. Denoting by $\circ$ the operation induced by $\mathcal{T}$, we finally have $\mathrm{T}_\circ = \mathcal{T}$ as before. Computing explicitly δ_a we get

$$x \cdot a = xM_a - x = x \circ a + x + a$$

which is the dot product we gave in Definition 1.12. Notice that it is distributive over $+$, and that $x \circ a = x + a + x \cdot a$ is the operation that makes $(V,+,\cdot)$ a radical ring. We will now focus on the case $\mathbb{K} = \mathbb{F}_2$, even if many of this results remains true also in a more general setting.

Proposition 5.4 ([12])**.** *It also holds* $\mathrm{Aut}(V,+,\cdot) = H_\circ$.

Definition 5.5. *In the above setting we define*

$$\Omega(\mathrm{T}_\circ) = \Omega_\circ = \{M_a | a \in V\} < \mathrm{GL}(V).$$

Theorem 5.6 ([8])**.** *Let* $d = \dim(W_\circ)$. *Then* $0 < d \leq n-2$.

Now we want to prove a characterization due to [7] that we will extensively use in the rest of the work. We will need the following two results.

Theorem 5.7 ([7])**.** *Let* $\mathcal{T} < \mathrm{AGL}(V)$ *be an elementary abelian regular group, with associated operation* $\circ$. *Let* $d = \dim(W_\circ)$ *and let* $m = n - d$. *Then there exists* $g \in \mathrm{GL}(V)$ *such that* $\Omega(\mathrm{T}_\circ{}^g) < \mathcal{U}(V)$, *where* $\mathcal{U}(V)$ *is the group of upper triangular linear maps on* V, *and* $W_\circ^g = \mathrm{span}\{e_{m+1}, ..., e_n\}$.

Lemma 5.8 ([10])**.** *Let* $\mathcal{T} < \mathrm{AGL}(V)$ *be abelian and regular. Then for each* $\sigma_x \in T_+$ *and* $\tau_y \in T_\circ$ *we have*

$$[\sigma_x, \tau_y] = \sigma_{x \cdot y},$$

where $\cdot$ *is the product of the* $\mathbb{F}_2$*-algebra related to* $\mathcal{T}$ *as in Theorem 5.2, and* $[\sigma_x, \tau_y] := \sigma_x^{-1}\tau_y^{-1}\sigma_x\tau_y$.

Remark 5.9. Notice that in our setting Lemma 5.8 implies that T_+ normalises $\mathcal{T} < \mathrm{AGL}(V)$ if and only if $\sigma_{x \cdot y} \in \mathcal{T}$ for all $x, y \in V$. Indeed, if for all $\sigma_x \in T_+$

we have $\mathcal{T}^{\sigma_x} = \mathcal{T}$, then

$$\sigma_{x\cdot y} = \sigma_x^{-1}\tau_y^{-1}\sigma_x\tau_y \in \mathcal{T}.$$

Conversely, if $\sigma_{x\cdot y} \in \mathcal{T}$ for each $x, y \in V$ then

$$\sigma_x^{-1}\tau_y^{-1}\sigma_x = \sigma_{x\cdot y}\tau_y^{-1} \in \mathcal{T}.$$

Finally, we notice that the condition $\sigma_{x\cdot y} \in \mathcal{T}$ for all $x, y \in V$ is equivalent to $x \cdot y \cdot z = 0$ for all $x, y, z \in V$.

We are finally ready to prove this useful characterization.

Theorem 5.10. *Let $\mathcal{T} < \mathrm{AGL}(V, +)$ abelian regular and let $\circ$ the operation induced on V. Let $d = \dim(W_\circ)$, $m = n - d$ and let us assume $W_\circ = \mathrm{span}\{e_{m+1}, ..., e_n\}$. Then, $T_+ < AGL(V, \circ)$ if and only if for all $M_y \in \Omega_\circ$ there exists a matrix $\Sigma_y \in (\mathbb{F}_2)^{m\times d}$ such that*

$$M_y = \begin{pmatrix} \mathbb{1}_m & \Sigma_y \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix}.$$

Proof. By Theorem 5.7, we know that there exists another group operation $\diamond$ on V such that the corresponding translation group is conjugated, by an element of $\mathrm{GL}(V)$, to $\mathrm{T}_\circ$ and satisfies $W_\diamond = W_\circ$ and $\Omega(T_\diamond) = \{\overline{M}_a | a \in V\} < \mathcal{U}(V)$. Let now $y \in V$, $A_y \in (\mathbb{F}_2)^{m\times m}$ an upper triangular matrix and $\Sigma_y \in (\mathbb{F}_2)^{m\times d}$ such that

$$\overline{M_y} = \begin{pmatrix} A_y & \Sigma_y \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix}.$$

Notice that the lower structure of the matrix is due to the property $e_i \in W_\diamond$ for each $m+1 \le i \le n$, i.e. $y \diamond e_i = e_i\overline{M}_y + y = y + e_i$ for each $m+1 \le i \le n$. Recall that, thanks to Lemma 5.8, $T_+ < \mathrm{AGL}(V, \diamond)$ if and only if for all $x, y \in V$ it holds $x \cdot y \in W_\diamond$. But this is true if and only if $x\overline{M}_y - x \in W_\diamond$ for all $x, y \in V$. Considering $x \in \mathrm{Span}\{e_1, ..., e_m\}$, we have that $x\overline{M}_y - x \in W_\diamond$ if and only if $A_y = \mathbb{1}_m$.
In order to conclude, we need to prove that each conjugate $T_\circ = T_\diamond^g$ is such that all the matrices in the group $\Omega(T_\circ)$ are of the form denoted above, provided that $g \in \mathrm{GL}(V)$ and $W_\circ$ is spanned by the last d vectors of the canonical basis. Let then $g \in \mathrm{GL}(V)$ such that $T_\circ = T_\diamond^g$. Since $W_\diamond g = W(T_\diamond^g) = W(T_\circ)$, then $\mathrm{Span}\{e_{m+1}, ..., e_n\}g = \mathrm{Span}\{e_{m+1}, ..., e_n\}$ and also $\mathrm{Span}\{e_{m+1}, ..., e_n\}g^{-1} = \mathrm{Span}\{e_{m+1}, ..., e_n\}$. Consequently, we have

$$g = \begin{pmatrix} G_1 & G_2 \\ \mathbb{0}_{d,m} & G_3 \end{pmatrix}$$

and

$$g^{-1} = \begin{pmatrix} G_1^{-1} & G_2' \\ \mathbb{0}_{d,m} & G_3^{-1} \end{pmatrix},$$

for some $G_1 \in (\mathbb{F}_2)^{m\times m}$, $G_2, G_2' \in (\mathbb{F}_2)^{m\times d}$ and $G_3 \in (\mathbb{F}_2)^{d\times d}$. Consequently, if

$M \in \Omega(T_\diamond)$, it holds

$$M^g = \begin{pmatrix} G_1^{-1} & G_2' \\ \mathbb{0}_{d,m} & G_3^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{1}_m & \Sigma_{m\times d} \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix} \begin{pmatrix} G_1 & G_2 \\ 0 & G_3 \end{pmatrix} = \begin{pmatrix} \mathbb{1}_m & \Sigma'_{m\times d} \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix}$$

and therefore the claim follows from $\Omega(T_\circ) = \Omega(T_\diamond^g) = \Omega(T_\diamond)^g$. ∎

Remark 5.11. $\Sigma_a = 0$ for each $a \in W_\circ$. Moreover, since $M_{a\circ b} = M_a M_b$, we obtain

$$M_{a\circ b} = \begin{pmatrix} \mathbb{1}_m & \Sigma_a + \Sigma_b \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix}.$$

Remark 5.12. Thanks to Remark 5.3 we know that $x \circ a = x\tau_a = xM_a + a$. From this we get

$$(a+b)\circ c = aM_c + bM_c + c = (aM_c + c) + (bM_c + c) + c = (a\circ c) + (b\circ c) + c.$$

Notice that it is not distributive. Writing $a = \sum \xi_i e_i$ we get

$$a \circ b = \begin{cases} \sum_{\xi_i \neq 0} b \circ e_i & \text{if weight(a) is odd} \\ (\sum_{\xi_i \neq 0} b \circ e_i) + b & \text{otherwise.} \end{cases}$$

This fact allows us to compute $a \circ b$ in polynomial time. Moreover, we see that the matrices Σ_i for $1 \leq i \leq m$ completely characterize $\circ$. We can therefore write

$$M_{e_i} = \begin{pmatrix} \mathbb{1}_m & \Sigma_{e_i} \\ \mathbb{0}_{d,m} & \mathbb{1}_d \end{pmatrix} = \left(\begin{array}{c|c} & \mathbf{b}_{i,1} \\ \mathbb{1}_m & \vdots \\ & \mathbf{b}_{i,m} \\ \hline \mathbb{0}_{d,m} & \mathbb{1}_d \end{array}\right)$$

denoting by $b_{i,j}$ the last d components of the j-th row of M_{e_i}. The b_{ij} can be seen as elements of $\mathbb{F}_{2^d}$ and stored in a matrix $\Theta_\circ = \{b_{ij}\} \in (\mathbb{F}_{2^d})^{m\times m}$. This matrix completely defines the operation $\circ$. It can be proven that

Theorem 5.13. *A matrix $\Theta_\circ \in (\mathbb{F}_{2^d})^{m\times m}$, defined by*

$$\Theta_\circ = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{b}_{2,1} & \cdots & \boldsymbol{b}_{m,1} \\ \boldsymbol{b}_{2,1} & \boldsymbol{0} & \cdots & \boldsymbol{b}_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{b}_{m,1} & \boldsymbol{b}_{m,2} & \cdots & \boldsymbol{0} \end{pmatrix}$$

is associated to a $\circ$ operation such that $\mathrm{T}_+ < \mathrm{AGL}(V,\circ)$ and $W_\circ = \mathrm{span}\{e_{m+1}, ..., e_n\}$ if and only if $\Theta_\circ$ is zero-diagonal, symmetric and no $\mathbb{F}_2$-linear combination of columns of $\Theta_\circ$ is the null vector. In this case, $\Theta_\circ$ is called defining matrix and the operation $\circ$ is defined by letting $\Sigma_{e_i} = \Theta_\circ[\cdot, i]$ for each $1 \leq i \leq m$.

Proof. Let as above $d = n - m \leq n - 2$. Thanks to Theorem 5.10 for each

$1 \leq i \leq n-d$ we can write

$$M_{e_i} = \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_i} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix}.$$

We then build the matrix $\Theta_\circ$ in the above statement by filling its columns with the rows of the matrices Σ_{e_i}. Since $\mathrm{T}_\circ$ is 2-elementary, for each $1 \leq i \leq n-d$ it holds $e_i \circ e_i = 0$, which means $\mathbf{b}_{i,i} = 0$. In addition, since the operation $\circ$ is commutative, for each $1 \leq i, j \leq n-d$ we have $e_i \circ e_j = e_j \circ e_i$, and hence $\mathbf{b}_{i,j} = \mathbf{b}_{j,i}$. Finally, let us assume that an$\mathbb{F}_2$-linear combination of columns of Θ is the null vector. Without loss of generality, let us assume $\Sigma_{e_1} + \Sigma_{e_2} = 0$. From this it follows that

$$M_{e_1 \circ e_2} = M_{e_1} M_{e_2} = \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_1} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_2} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} =$$
$$= \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_1} + \Sigma_{e_2} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} = \mathbb{1}_n,$$

which implies $e_1 \circ e_2 = w$, for some $w \in W_\circ$, i.e. $e_1 = e_2 \circ w = e_2 + w$. This proves that $e_1 + e_2 \in W_\circ$, which is a contradiction, since e_1 and e_2 can never be weak keys. ∎

Remark 5.14. As a consequence, as little as $m(m-1)/2$ values are needed to define such an operation.

Theorem 5.15. *Let $\circ$ be defined as above. For each $x, y \in V$ there exists $\epsilon_{x,y} \in U_\circ$ such that $x+y = x \circ y + \epsilon_{x,y}$, with $\epsilon_{x,y} = x \cdot y = (0, ..., 0, (x_1, ..., x_m)\Sigma_y)$. Moreover $U_\circ \subseteq W_\circ$.*

Proof. Let us fix $x, y \in V$ and study in greater detail the error $\epsilon_{x,y} = x \cdot y$. Since $V = W_\circ^\perp \bigoplus W_\circ$, we can write $x = (\overline{x}, \widetilde{x})$ with $\overline{x} \in (\mathbb{F}_2)^{n-d}$ and $\widetilde{x} \in (\mathbb{F}_2)^d$. First of all notice that, if $x \in W_\circ$ then $x \cdot y = 0$. In fact, in this case

$$x \cdot y = x + y + x \circ y = x + y + x + y = 0.$$

In the general case we have

$$x \cdot y = xM_y + y + x + y =$$
$$(\overline{x}, \widetilde{x}) \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_y \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} + x = (0, \overline{x}\Sigma_y) \in W_\circ,$$

which does not depend on $\widetilde{x}$, the component of x in the space of weak keys. As the first $n-d$ coordinates of $x \cdot y$ are null, we also proved that any error is part of $W_\circ$. ∎

Corollary 5.16. *It follows that $U_\circ$ is composed of all the possible vectors $w \in W_\circ$ whose last d components are all the possible $\mathbb{F}_2$-linear combination of the rows of the matrices Σ_x.*

Remark 5.17. From the fact that $x \cdot y \in U_\circ \subseteq W_\circ$, it follows that $x \cdot y \cdot z = 0$ for each $x, y, z \in V$.

Proposition 5.18. *For each $x, k, \Delta \in V$ it holds*

$$(k + x) \circ ((x \circ \Delta) + k) = \Delta + k \cdot \Delta.$$

Proof. Let $x, k, \Delta \in V$. Rewriting the above equation we obtain

$$\begin{gathered}(x + k) \circ ((x \circ \Delta) + k) = (x + k) \circ (x + \Delta + x \cdot \Delta + k) = \\ = x + k + x + \Delta + x \cdot \Delta + k + (x + k) \cdot (x + \Delta + x \cdot \Delta + k) = \\ \Delta + x \cdot \Delta + x \cdot x + x \cdot \Delta + x \cdot x \cdot \Delta \\ + x \cdot k + k \cdot x + k \cdot \Delta + k \cdot \Delta \cdot x + k \cdot k = \\ \Delta + k \cdot \Delta,\end{gathered}$$

since $x \cdot x = x + x + x \cdot x = 0$ and all the triple products vanish because of Remark 5.17. ∎

The importance of this equation in studying the interaction of $\circ$ differenced with the key addition layers has already been explained in Section 4.1. We will now introduce some deeper results about KDT.

Definition 5.19. *The Key Distribution Table (KDT) of an operation $\circ$ is the integer table defined by*

$$KDT^\circ[\delta_I, \delta_O] := \#\left\{k \in V | \delta_I + k \cdot \delta_I = \delta_O\right\}.$$

Proposition 5.20. *The KDT° table is symmetric, that is*

$$KDT^\circ[\Delta_1, \Delta_2] = KDT^\circ[\Delta_2, \Delta_1]$$

for each $\Delta_1, \Delta_2 \in V$.

Proof. Let us fix $\Delta_1, \Delta_2, k \in V$ and suppose that $\Delta_1 + k \cdot \Delta_1 = \Delta_2$. Then it holds

$$\begin{gathered}\Delta_2 + k \cdot \Delta_2 = \Delta_1 + k \cdot \Delta_1 + k \cdot (\Delta_1 + k \cdot \Delta_1) = \\ = \Delta_1 + k \cdot \Delta_1 + k \cdot \Delta_1 + k \cdot k \cdot \Delta_1 = \Delta_1,\end{gathered}$$

again thanks to Remark 5.17. Therefore $KDT^\circ[\Delta_1, \Delta_2] = KTD^\circ[\Delta_2, \Delta_1]$. ∎

Remark 5.21. As already pointed out, the more zero entries the KDT has, the easier it is to control the effect of the key addition layer. It turns out that the number of zero entries is strictly related to the dimension of the weak key space $W_\circ$.

Lemma 5.22. *For each $a \in V$, it holds*

$$\mathrm{rk}(\Sigma_a) \leq \min(n - d - 1, d).$$

where $d = \dim(W_\circ)$ and Σ_a is the unique matrix associated to a through Theorem 5.10.

Proof. If $a \in W_\circ$ there is nothing to prove, since $\Sigma_a = \mathbb{0}_{n-d,d}$. If not, then $a = (\overline{a}, \widetilde{a})$ with $\overline{a} \in (\mathbb{F}_2)^{n-d}$ and $\widetilde{a} \in (\mathbb{F}_2)^d$. Moreover, $\overline{a} \neq 0$. Since $0 = a \circ a = aM_a + a$, it follows that $a(M_a + \mathbb{1}_n) = 0$. This implies, thanks to the description of M_a given in Theorem 5.10, that

$$a \in \ker \begin{pmatrix} 0 & \Sigma_a \\ 0 & 0 \end{pmatrix}.$$

Therefore $\overline{a} \in \ker(\Sigma_a)$. From this it follows

$$\mathrm{rk}(\Sigma_a) = \dim(\mathrm{Im}(\Sigma_a)) = n - d - \dim(\ker(\Sigma_a)) < n - d - 1.$$

Now, if $n-d-1 \leq d$, the result clearly holds. Otherwise we have $d < n-d-1 < n-d$, while $\mathrm{rk}(\Sigma_a) \leq d = \min(n-d-1, d)$. ∎

Remark 5.23. This bound reaches minimum values for extremal values of d, which are (thanks to Theorem 5.6) $d = 1$ and $d = n - 2$.

Theorem 5.24. *The number of non zero entries in each row of the key distribution table KDT° is upper bounded by $2^{\min(n-d-1,d)}$.*

Proof. Given a fixed $\Delta \in V$, the number of non zero entries in the row $KDT^\circ[\Delta, \cdot]$ depends on the values of $k \in V$. It holds $k \cdot \Delta = \overline{k}\Sigma_\Delta \in \mathrm{Im}(\Sigma_\Delta)$, and

$$\dim(\mathrm{Im}(\Sigma_\Delta)) = \mathrm{rk}(\Sigma_\Delta) \leq \min(n - d - 1, d)$$

thanks to Lemma 5.22. The thesis then immediately follows. ∎

Corollary 5.25. *For every fixed $\delta_I \in V$, it holds*

$$KDT^\circ[\delta_I, \delta_o] \in \left\{0, 2^{n-\mathrm{rk}(\Sigma_{\delta_I})}\right\}.$$

Proof. Let $\delta_o \in V$ such that $KDT^\circ[\delta_I, \delta_O] \neq 0$.Two keys $k_1, k_2 \in V$ are such that $k_1 \cdot \delta_I = k_2 \cdot \delta_O$ if $\overline{k_1}$ and $\overline{k_2}$ are in the same class modulo $\ker(\Sigma_{\delta_I})$. Recalling that the value of $k \cdot \Delta$ does not depend on the last d components of k, then $KDT^\circ[\delta_i, \delta_o]$ is the number of elements contained in each class modulo $\ker(\Sigma_{\delta_I})$ multiplied by 2^d. Therefore

$$\begin{aligned} KDT^\circ[\delta_I, \delta_o] &= 2^d 2^{\dim(\ker(\Sigma_{\delta_i}))} = \\ &= 2^d 2^{n-d-\mathrm{rk}(\Sigma_{\delta_i})} = 2^{n-\mathrm{rk}(\Sigma_{\delta_i})}. \end{aligned} \tag{5.1}$$

∎

5.2 The case $d = n - 2$

We will now investigate deeper the case $d = n - 2$. As shown in Theorem 5.24, such a choice for d allows us to have the best bound of 2 non-zero entries for each row of the KDT and hence a greater control on how differences propagates through the key addition layer. Notice that the bound grows exponentially with

d, until $d \leq n - d - 1$.
Let us fix for this section an operation $\circ$ such that $d = dim(W_\circ) = n - 2$. Thanks to Theorem 5.13, its representing matrix can be written as

$$\Theta = \begin{pmatrix} \mathbf{0} & \mathbf{b} \\ \mathbf{b} & \mathbf{0} \end{pmatrix}$$

and hence depends on a sigle nonzero element $\mathbf{b} \in (\mathbb{F}_2)^{n-2}$. Observe that $2^{n-2} - 1$ such operations exist, since we do not count the trivial one which conicides with $+$. The matrices associated with the first two basis vectors are respectively

$$M_{e_1} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix} \mathbf{0} \\ \mathbf{b} \end{matrix} \\ \hline \mathbb{0}_{n-2,2} & \mathbb{1}_{n-2} \end{array}\right), \; M_{e_2} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix} \mathbf{b} \\ \mathbf{0} \end{matrix} \\ \hline \mathbb{0}_{n-2,2} & \mathbb{1}_{n-2} \end{array}\right),$$

while $M_{e_j} = \mathbb{1}_n$ for $j \geq 3$. Moreover, applying Corollary 5.16, we obtain $U_\circ = \{\mathbb{0}, u\}$ where $u = (0, 0, \mathbf{b}) \in (\mathbb{F}_2)^n$.
We are now ready to present a very important and useful result, due to [7], which affirms that all the translation groups for sums $\circ$ with $d = n - 2$ are conjugated by an element $g \in \mathrm{GL}(V)$. Notice that this fact is much stronger than Theorem 1.6, since there we can choose $g \in \mathrm{Sym}(V)$. Moreover, thanks to Theorem 5.7 we may restrict ourselves to prove this theorem for translation groups defining operations characterized as above, without loss of generality.
Let for now $T_\circ$ and $T_\diamond$ be elementary abelian regular subgroups of $\mathrm{AGL}(V, +)$ defining two operations $\circ$ and $\diamond$ respectively such that $\dim(W_\circ) = \dim(W_\diamond) = n-2$ and $W_\circ = W_\diamond = \mathrm{span}\{e_3, ...e_n\}$. Denote by $\mathbf{b}_\circ$ and $\mathbf{b}_\diamond$ the defining vectors for the two operations.

Lemma 5.26. *If $\mathbf{b}_\circ$ and $\mathbf{b}_\diamond$ have the same Hamming weight, i.e. the same number of nonzero coordinates when seen as elements of $(\mathbb{F}_2)^d$, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$.*

Proof. Let us denote $T_\circ = \langle \tau_{e_1}^\circ, ..., \tau_{e_n}^\circ \rangle$ and $T_\diamond = \langle \tau_{e_1}^\diamond, ..., \tau_{e_n}^\diamond \rangle$. If $\mathbf{b}_\circ$ and $\mathbf{b}_\diamond$ have the same Hamming weight, i.e. the same number of non-zero coordinates, then there exists a permutation matrix $P \in (\mathbb{F}_2)^{(n-2)\times(n-2)}$ such that $\mathbf{b}_\circ P = \mathbf{b}_\diamond$. Let $P' \in (\mathbb{F}_2)^{n\times n}$ be the permutation matrix defined as

$$P' := \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & P & \\ 0 & 0 & & & \end{pmatrix}.$$

Notice that when we multiply a matrix M by P' on the right we are permuting the last $n - 2$ columns of M. On the other hand, multiplying M by P'^{-1} on the left permutes the last $n - 2$ rows of M. Hence, we have

$$P'^{-1}\tau_{e_i}^\circ P' = P'^{-1}M_{e_i}^\circ P'\sigma_{e_i P'} = \tau_{e_i P'}^\diamond = \tau_{e_i \pi}^\diamond,$$

where π is the index permutation induced by P'. But this implies $P'^{-1}T_\circ P' = T_\diamond$. Hence, the two groups corresponding to vectors with the same weight are conjugated. ∎

Lemma 5.27. *Let*

$$\boldsymbol{b}_\circ = (\underbrace{1,...,1}_{i},0,...,0)$$

and

$$\boldsymbol{b}_\circ = (\underbrace{1,...,1}_{i+1},0,...,0)$$

for $1 \leq i \leq n-3$. *Then there exists* $g \in \mathrm{GL}(V)$ *such that* $T_\diamond = T_\circ^g$.

Proof. With the same notation we used above, let $P \in (\mathbb{F}_2)^{n\times n}$ be the matrix whose j-th row $P_j = e_j$ if $j \neq i+2$ and $P_{i+2} = e_{i+2} + e_{i+3}$, i.e.

$$P := \begin{pmatrix} 1 & 0 & \cdots & & \cdots & 0 \\ 0 & 1 & & & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \cdots & & 1 & 1\cdots & 0 \\ 0 & \cdots & & 0 & 1\cdots & 0 \\ 0 & \cdots & & & \cdots & 0 \end{pmatrix}.$$

Notice that $P^{-1} = P$. Note also that multiplying a matrix M by P no the right we are updating its $(i+3)$-th column by summing up its $(i+2)$-th and $(i+3)$-th columns. On the other hand, multiplying a matrix M by $P^{-1} = P$ on the left we are updating its $(i+2)$-th row by summing up its $(i+2)$-th and $(i+3)$-th rows. Therefore it holds

$$P\tau^\circ_{e_j}P = PM^\circ_{e_j}P\sigma_{e_jP} = \tau^\diamond_{e_j}$$

for $j \neq i+2$ and

$$P\tau^\circ_{e_{i+2}+e_{i+3}}P = \tau^\diamond_{e_{i+2}}.$$

Notice that $\tau^\circ_{e_{i+2}+e_{i+3}}\tau^\circ_{e_{i+3}} = \tau_{e_{i+2}}$. This implies that

$$\langle \tau^\circ_{e_1},...,\tau^\circ_{e_{i+1}},\tau^\circ_{e_{i+2}+e_{i+3}},\tau^\circ_{e_{i+3}},...,\tau^\circ_{e_n}\rangle = T_\circ.$$

Therefore we have $PT_\circ P = T_\diamond$. Notice that, together with Lemma 5.26, we have proved that if $\mathbf{b}_\circ$ and $\mathbf{b}_\diamond$ are such that their Hamming weights differ by one, the associated groups $T_\circ$ and $T_\diamond$ are conjugated in $\mathrm{GL}(V)$. ∎

Theorem 5.28. *Let* $T_\circ$ *and* $T_\diamond$ *elementary abelian regular subgroups of* $\mathrm{AGL}(V,+)$ *defining two operations* $\circ$ *and* $\diamond$ *respectively such that* $\dim(W_\circ) = \dim(W_\diamond) = n-2$. *Then, there exists* $g \in \mathrm{GL}(V)$ *such that* $T_\diamond = T_\circ^g$.

Proof. As already observed, we can restrict ourselves without loss of generality to the case of $W_\circ = W_\diamond = \mathrm{span}\,\{e_3,...e_n\}$. Let us denote by d_1 and d_2 the Hamming weight of $\mathbf{b}_\circ$ and $\mathbf{b}_\diamond$ and assume, again without loss of generality,

that $d_1 < d_2$. We can define

$$\mathbf{b}_j = (\underbrace{1, ..., 1}_{j}, 0, ..., 0)$$

for $d_1 \leq j \leq d_2$ and denote by T_j the translation group associated with the sum defined by $\mathbf{b}_j$. Thanks to Lemma 5.26, $T_\circ$ is conjugated in $\mathrm{GL}(V)$ to T_{d_1} and $T_\diamond$ to T_{d_2}. Moreover, applying Lemma 5.27 we have that T_j is conjugated to T_{j+1} for each j and this completes the proof. ∎

Thanks to this result we are now allowed to fix a single operation $\circ$ (or equivalently, a single nonzero element $\mathbf{b} \in (\mathbb{F}_2)^{n-2}$), and reduce to that (up to conjugation) many computations. This is for example the case of the last section of this work, in which we will analyze differential properties of some permutations. However, this statement is no longer guaranteed to hold if we let the dimension of the weak key space grow.
The other important result we will prove for this case is due to [12] and will rule the composition of the group of automorphisms $H_\circ$ of our operation.

Lemma 5.29. *For each $\lambda \in H_\circ$ it holds $W_\circ\lambda = W_\circ$ and $U_\circ\lambda = U_\circ$.*

Proof. Let $\lambda \in H_\circ$ and let us prove that $W_\circ\lambda = W_\circ$. Let $a \in W_\circ$ and $b \in V$. We want $a\lambda \circ b = a\lambda + b$. But

$$a\lambda \circ b\lambda = (a \circ b)\lambda = (a + b)\lambda = a\lambda + b\lambda$$

and since λ is invertible, $W_\circ\lambda = W_\circ$.
On the other hand, if $a \in U_\circ$ then $a = b \cdot c$ for some $b, c \in V$. Then $a\lambda = (b \cdot c)\lambda = b\lambda \cdot c\lambda$, hence $a\lambda \in U_\circ$. ∎

Theorem 5.30. *Let $\boldsymbol{b} \in (\mathbb{F}_2)^d$ (with $d = n - 2$) the defining vector of $\circ$, and $\lambda \in (\mathbb{F}_2)^{n \times n}$. The following are equivalent:*

- *λ is compatible with $\circ$;*
- *there exist $A \in \mathrm{GL}((\mathbb{F}_2)^2, +)$, $D \in \mathrm{GL}((\mathbb{F}_2)^d, +)$, and $B \in (\mathbb{F}_2)^{2 \times d}$ such that*

  $$\lambda = \begin{pmatrix} A & B \\ \mathbb{0}_{d,2} & D \end{pmatrix}$$

  and $\boldsymbol{b}D = \boldsymbol{b}$.

Proof. Let us write λ into the block form

$$\lambda = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with the dimensions of the blocks as indicated above, and suppose $\lambda \in H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$. Since, from Lemma 5.29, $W_\circ\lambda = W_\circ$ and $W_\circ$ is spanned by the last d vectors of the canonical basis, it must hold $C = \mathbb{0}_{d,2}$. As a consequence, A and D must be invertible (since $\lambda \in H_\circ$). Finally, since from

Lemma 5.29 again $U_\circ\lambda = U_\circ$ and as previously observed $U_\circ = \{\mathbb{0}, u\}$, we obtain $\mathbf{b}D = \mathbf{b}$.
Conversely, let us assume that the above conditions hold. Recalling that $H_\circ = Aut(V, +, \cdot)$ we want to prove that given $x, y \in V$ it holds

$$(x \cdot y)\lambda = x\lambda \cdot y\lambda.$$

If $x \in W_\circ$ also $x\lambda \in W_\circ$ and the equation is trivial since both sides are zero. It is then sufficient to consider the case $x = e_1$ and $y = e_2$. It is easy to check that both the products $e_1 \cdot e_2$ and $e_1\lambda \cdot e_2\lambda$ are nonzero and hence equal to u, so the above equation holds and this completes the proof. ∎

As already observed many times, in order to perform a successful differential attack we need that the mixing layer $\lambda \in H_\circ$. However, as shown in Section 2.2, not all $\lambda \in H_\circ$ may be used into the mixing layer of a cipher; indeed, λ is required also to be secure in the standard setting, while providing diffusion for the cipher (i.e. activating as many S-boxes as possible). If both these conditions are satisfied, we found a λ that can actually be the mixing layer of a cipher, and hence one or more ciphers against which our attack may outperform the classic one. An interesting example of a distinguish attack built in this way is given in [12]. In the next sections, we try to reproduce the main results of the case $d = n - 2$, especially Theorem 5.28 and Theorem 5.30, in two new different settings. We will do this in order to be able to include more λ's and consequently more possible ciphers in our analysis. The first one is $d = n - 3$, a natural continuation of our analysis. As an immediate drawback, we observe that the bound of Theorem 5.24 is twice the one we achieved in the previous case. On the other hand, we obtain new possible sums, split into two different classes of conjugation with slightly different conditions on $H_\circ$. In the second setting we will operate on parallel sums, i.e. sums that acts in parallel on smaller blocks of components. Each sum has $d = n - 2$. The importance of this case is clear once again from Section 2.2; since usually the S-boxes are small and act in parallel, while the mixing layer is the same, studying a sum that operates on each block separately and for which λ is linear allows us to to perform an attack on bigger ciphers. We start looking at sums acting on two parallel blocks, but then generalize the results to m parallel sums. The choice $d = n - 2$ offers all the advantages this condition brings in the classic setting. Moreover, an equivalent of Theorem 5.28 allows us to fix the same sum for all the parallel blocks (up to conjugation).

5.3 The case $d = n - 3$

Let us fix for now an operation $\circ$ such that $d = \dim(W_\circ) = n - 3$. Thanks to Theorem 5.13, its representing matrix can be written as

$$\Theta_\circ = \begin{pmatrix} \mathbf{0} & \mathbf{b}_{2,1} & \mathbf{b}_{3,1} \\ \mathbf{b}_{2,1} & \mathbf{0} & \mathbf{b}_{3,2} \\ \mathbf{b}_{3,1} & \mathbf{b}_{3,2} & \mathbf{0} \end{pmatrix}$$

such that no $\mathbb{F}_2$-linear combination of columns of $\Theta_\circ$ is the null vector (notice also the simmetry). Non-identity matrices are associated to the first three basis vectors, more precisely

$$M_{e_1} = \left(\begin{array}{c|c} & \mathbf{0} \\ \mathbb{1}_3 & \mathbf{b}_{2,1} \\ & \mathbf{b}_{3,1} \\ \hline \mathbb{0}_{n-3,3} & \mathbb{1}_{n-3} \end{array}\right),\ M_{e_2} = \left(\begin{array}{c|c} & \mathbf{b}_{2,1} \\ \mathbb{1}_3 & \mathbf{0} \\ & \mathbf{b}_{3,2} \\ \hline \mathbb{0}_{n-3,3} & \mathbb{1}_{n-3} \end{array}\right),\ M_{e_3} = \left(\begin{array}{c|c} & \mathbf{b}_{3,1} \\ \mathbb{1}_3 & \mathbf{b}_{3,2} \\ & \mathbf{0} \\ \hline \mathbb{0}_{n-3,3} & \mathbb{1}_{n-3} \end{array}\right)$$

while as always $M_{e_j} = \mathbb{1}_n$ for $j > 3$.

Remark 5.31. Thanks to Theorem 5.24 the number of nonzero entries for each row of the KDT° is upper bounded by 4. Notice that this is twice the bound for the case $d = n - 2$.

Notation 5.32. Given $i, j < 3$ (so that e_i and e_j are components of the strong key space) we have

$$e_i \cdot e_j = e_i \circ e_j + e_i + e_j = e_i M_j + e_j + e_i + e_j = (0, 0, \mathbf{b}_{i,j})$$

thanks to the structure of matrices M_i as pointed out above. We can therefore introduce the notation

$$u_{ij} = u_{ji} := e_i \cdot e_j = (0, 0, \mathbf{b}_{i,j}).$$

Remark 5.33. Thanks to Corollary 5.16, we know that $U_\circ$ is composed of all the vectors $w \in W_\circ$ whose last d components are all the possible $\mathbb{F}_2$-linear combinations of the vectors $\mathbf{b}_{i,j}$. At least two of these vectors are granted to be independent thanks to the constraints on $\Theta_\circ$, hence $\dim(U_\circ) \geq 2$. Moreover, all these vectors are spanned by u_{12}, u_{13} and u_{23} and this implies $\dim(U_\circ) \leq 3$.

It turns out that $\dim(U_\circ)$ is a key information in explaining the behaviour of the operation $\circ$. As we have just seen it can be 2 or 3. This confirms the (already known) fact that if $n = \dim(V) = 4$ there are no possible sums with $d = n - 3$. If $n = 5$, $\dim(W_\circ) = 2$ and consequently $\dim(U_\circ) = 2$, since $U_\circ \subseteq W_\circ$. If $n \geq 6$, $\dim(W_\circ) \geq 3$ and so $\dim(U_\circ)$ can effectively be 2 or 3, with $U_\circ = W_\circ$ only for $n = 6$ and $\dim(U_\circ) = 3$.
Let us now study matrices $\lambda \in H_\circ$. Like the previous case, we can write

$$\lambda = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with $A \in (\mathbb{F}_2)^{3\times 3}$, $B \in (\mathbb{F}_2)^{3\times(n-3)}$, $C \in (\mathbb{F}_2)^{(n-3)\times 3}$ and $D \in (\mathbb{F}_2)^{(n-3)\times(n-3)}$. Again, Lemma 5.29 immediately implies $C = 0$ (and consequently A and D invertible) and $U_\circ D = U_\circ$. This first part is independent from $\dim(U_\circ)$. Now we want to investigate deeper each of the two cases, in order to obtain necessary and sufficient conditions for λ to belong to $H_\circ$.

Observation 5.34. Let us focus first on the case $\dim(U_\circ) = 2$. $U_\circ$ is generated by two error vectors among u_{12}, u_{13} and u_{23}. We can assume, without loss of

generality, that u_{12} and u_{13} are the two independent vectors, and complete them to a basis of $W_\circ$ $\{u_{12}, u_{13}, w_3, ..., w_{n-3}\}$ with $w_i \in W_\circ \backslash U_\circ$. Again thanks to Lemma 5.29, the image of u_{12} and u_{13} must be two independent vectors of $U_\circ$; note that we have in total 6 ways to map them, 3 choices for the first one (all the 3 nonzero vectors in $U_\circ$) and 2 for the second one (the 2 remaining nonzero vectors). On the other hand, the w_i must be mapped into $W_\circ \backslash \langle U_\circ, w_j \rangle$ for each $j < i$ (since D must be invertible). The image of u_{12} and u_{13} create additional constraints on the composition of A, since it must satisfy the compatibility equations

$$u_{12}\lambda = e_1\lambda \cdot e_2\lambda,$$

$$u_{13}\lambda = e_1\lambda \cdot e_3\lambda,$$

$$u_{23}\lambda = e_2\lambda \cdot e_3\lambda.$$

Conversely, if all these constraints are satisfied for $\lambda \in (\mathbb{F}_2)^{n\times n}$, then λ is compatible with $\circ$. Indeed, we need to prove that given $x, y \in V$ it holds $(x \cdot y)\lambda = x\lambda \cdot y\lambda$. If $x \in W_\circ$, then, by construction of λ, $x\lambda \in W_\circ$, so when we have a dot product between an element of $W_\circ$ and an element outside $W_\circ$ both sides of the compatibility equation are zero. For linearity we than have that $(x \cdot y)\lambda = x\lambda \cdot y\lambda$ if and only if

$$((x_1, x_2, x_3, 0, ...0) \cdot (y_1, y_2, y_3, 0, ...0))\lambda = (x_1, x_2, x_3, 0, ...0)\lambda \cdot (y_1, y_2, y_3, 0, ...0)\lambda$$

if and only if it holds for all possible combinations of $x = e_i$ and $y = e_j$ with $i, j \leq 3$. But these are precisely the three constraints we put on A.

Observation 5.35. Let us now study the case $\dim(U_\circ) = 3$ instead. Now $U_\circ$ is spanned by u_{12}, u_{13} and u_{23}. Since the dimension of $U_\circ$ is now equal to the dimension of the strong key space, we have an easier way to look at it. For the block A, we can choose any invertible 3×3 matrix. This fixes the image of e_1, e_2 and e_3 and consequently u_{12}, u_{13}, u_{23}; they are also independent by linearity, since if for example $\lambda u_{12} + \lambda u_{13} = 0$ then $\lambda(u_{12} + u_{13}) = 0$ and since λ is also invertible this would imply $\dim(U_\circ) = 2$. We can then choose like before $\{u_{12}, u_{13}, u_{23}, w_4, ..., w_{n-3}\}$ as a basis for $W_\circ$ and the freedom on D, once A is fixed, is left from multiple ways to map the w_i into $W_\circ \backslash \langle U_\circ, w_j \rangle$. It is possible to work also in the opposite direction: we may fix the image of u_{12}, u_{13} and u_{23}, and then A is consequently fixed (we have three independent compatibility equations). Also in this case, for each mapping of u_{ij} we are allowed to freely map the w_i into vectors of $W_\circ \backslash \langle U_\circ, w_j \rangle$.
Vice versa, like the previous case, if all these constraints are satisfied for $\lambda \in (\mathbb{F}_2)^{n\times n}$, then λ is compatible with $\circ$, since the first part (with the application of Lemma 5.29) remains unchanged, and for the second one we can once again restrict ourselves to look at the cases $x = e_i$ and $y = e_j$ with $i, j \leq 3$. These are exactly the compatibility equations that we impose on D if we start constructing λ from A or on A if we start from D. This completes the proof of the equivalent of Theorem 5.15 for the case $d = n - 3$.

Theorem 5.36. *Let $\lambda \in (\mathbb{F}_2)^{n\times n}$. The following are equivalent:*

- *λ is compatible with $\circ$;*

- *there exist* $A \in GL((\mathbb{F}_2)^3, +)$, $D \in GL((\mathbb{F}_2)^{n-3}, +)$, *and* $B \in (\mathbb{F}_2)^{3\times(n-3)}$ *such that*

$$\lambda = \begin{pmatrix} A & B \\ \mathbb{0}_{n-3,3} & D \end{pmatrix}$$

and if $\dim(U_\circ) = 2$ *A and D follow the constraints described in Observation 5.34, if* $\dim(U_\circ) = 3$ *they follow the constraints described in Observation 5.35.*

Example 5.37. Theorem 5.36 gives us a quick way to check the dimension of $H_\circ$. Let us check it in some different cases.

- $n = 5$; in this case, as observed before, the only possible case is $\dim(U_\circ) = \dim(W_\circ) = 2$. We have only 6 possibilities for D (3 for the first error and 2 for the second one; in this case no w_i are added to form a base), and each of them through compatibility equations allows 4 different possibilities for A (by trivial computation) for a total of 24 possible couples (A, D). We have then 64 possibilities for B (that has no constraints, and size 3×2) for a total of 1536 possible matrices in $H_\circ$.

- $n = 6$, $\dim(U_\circ) = 2$; let us fix (u_{12}, u_{13}, w_1) as a basis for $W_\circ$, with u_{12} and u_{13} a basis for $U_\circ$. Like before, u_{12} and u_{13} can be mapped in 6 ways. Each of these 6 ways allows 4 possible A's like before, since the compatibility equations does not depend on the image of w_1 (thanks to Remark 5.34). Note that this fact is true also in higher dimensions: for each D, we will always have 4 possible A's. Moreover, for each way we still have 4 possible images for w_1 (the size of $W_\circ \backslash U_\circ$) and hence 4 possible D's. In conclusion, we have 6*4=24 possibilities for D, and for each D 4 possibilities for A, so 96 possible couples A/D; from the 512 possible B's we obtain 49152 different elements in $H_\circ$.

- $n = 6$, $\dim(U_\circ) = 3$; we have 168 possible A's, each one fixing the image of u_{12}, u_{13}, u_{23}. Since these three vectors form a basis for $W_\circ = U_\circ$, A fixes D. We have then 512 possibilities for B for a total of 86016 possible matrices in $H_\circ$.

- $n = 7$, $\dim(U_\circ) = 2$; let's fix $(u_{12}, u_{13}, w_1, w2)$ as a basis for $W_\circ$. We have 3 possibilities for u_{12}, 2 for u_{13}, then 12 for w_1 ($W_\circ \backslash U_\circ$) and finally 8 for w_2 ($W_\circ \backslash \langle u_{12}, u_{13}, w_1 \rangle$). We than have 576 possible D's, with 4 A's each (as noted before), 4096 possible B's for 9437184 possible matrices in total.

- $n = 7$, $\dim(U_\circ) = 3$; again, we have all the possible 168 invertible 3×3 matrices for A, and each one fixes u_{12}, u_{13}, u_{23}. We can complete this to a base for $W_\circ$ by adding w_4, which have 8 possible images ($W_\circ \backslash U_\circ$) for a total of 1344 couples A/D. With 4096 B's we obtain 5505024 possible elements for $H_\circ$.

And so on.

Remark 5.38. From $n = 6$, depending on $\dim(U_\circ)$, we have two possible cardinalities for $H_\circ$. This fact, that does not occur if $n = 5$, implies that we will have at least two different conjugacy classes for $H_\circ$ in higher dimensions, in accordance with the computation made in [7]. Thanks to that result, we can also show that conjugacy classes depends only on $\dim(U_\circ)$.

Lemma 5.39. *Let $T_\circ$ and $T_\diamond$ elementary abelian regular subgroups of $AGL(V,+)$ defining two operations $\circ$ and $\diamond$ respectively, and such that $T_\diamond = T_\circ^g$ for $g \in \mathrm{GL}(V)$. Then $H_\diamond = H_\circ^g$.*

Proof. We have that $\mathrm{AGL}(V,\circ)$ and $\mathrm{AGL}(V,\diamond)$ are the normalizers of $T_\circ$ and $T_\diamond$ respectively. Since $T_\diamond = T_\circ^g$, we have $\mathrm{AGL}(V,\diamond) = \mathrm{AGL}(V,\circ)^g$ and consequently $\mathrm{GL}(V,\diamond) = \mathrm{GL}(V,\circ)^g$, being $\mathrm{GL}(V,\circ)$ the stabilizer of 0 in $\mathrm{AGL}(V,\circ)$. Finally, the intersection with $\mathrm{GL}(V,+)$ is preserved since $g \in \mathrm{GL}(V,+)$. ∎

Theorem 5.40. *Let $T_\circ$ and $T_\diamond$ elementary abelian regular subgroups of $\mathrm{AGL}(V,+)$ defining two operations $\circ$ and $\diamond$ respectively such that $\dim(W_\circ) = \dim(W_\diamond) = n-3$. Then, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$ if and only if $\dim(U_\circ) = \dim(U_\diamond)$.*

Proof. Up to conjugation, suppose that $W_\circ$ and $W_\diamond$ are generated by $\{e_4, ..., e_n\}$ and that $U_\circ$ and $U_\diamond$ are generated by e_4, e_5 if $\dim(U_\circ) = 2$ and e_4, e_5, e_6 if $\dim(U_\circ) = 3$. This implies that the matrices associated with e_i can be written as

$$\begin{pmatrix} M_i^\circ & \mathbb{0}_{6,n-6} \\ \mathbb{0}_{n-6,6} & \mathbb{1}_{n-6} \end{pmatrix}, \begin{pmatrix} M_i^\diamond & \mathbb{0}_{6,n-6} \\ \mathbb{0}_{n-6,6} & \mathbb{1}_{n-6} \end{pmatrix}$$

respectively, with M_i° and $M_i^\diamond$ are 6×6 matrices, since columns from 4 to n are made by elements of $U_\circ$ and so are 0 from the 7th component. Moreover, for $i > 3$ M_i° is the identity as usual. We can then consider this sum as composed by two parallel sums, the first one acting on the first six components and defined by M_i° and $M_i^\diamond$ and the second one acting on the last $n-6$ components being the usual XOR (its matrices are all identities).
Now call $\overline{T}_\circ$ and $\overline{T}_\diamond$ the translation groups associated to the alternative sum defined my M_i° and $M_i^\diamond$. Thanks to the classification made in [7] we know that for $n = 6$ and $d = 3$ we have two conjugacy classes. Moreover, thanks to Lemma 5.39 and Example 5.37 these two conjugacy classes corresponds exactly to the two cases $\dim(U_\circ) = 2$ and $\dim(U_\circ) = 3$. If we denote by $\overline{U}_\circ$ the set of error associated by the operation defined by $\overline{T}_\circ$, this means that, since $\dim(\overline{U}_\circ) = \dim(\overline{U}_\diamond)$ ($U_\circ$ is left fixed by our restriction by construction) there exist $\overline{g} \in \mathrm{GL}_6(V)$ such that $\overline{T}_\diamond = \overline{T}_\circ^{\overline{g}}$. By taking

$$g = \begin{pmatrix} \overline{g} & \mathbb{0}_{6,n-6} \\ \mathbb{0}_{n-6,6} & \mathbb{1}_{n-6} \end{pmatrix}$$

we have $T_\diamond = T_\circ^g$ as required. ∎

Corollary 5.41. *Given two sum $\circ$ and $\diamond$ with $\dim(W_\circ) = \dim(W_\diamond) = n-3$, the corresponding groups of compatible maps $H_\circ$ and $H_\diamond$ are conjugated by $g \in GL(V)$ if and only if $\dim(U_\circ) = \dim(U_\diamond)$.*

Proof. It follows from application of Lemma 5.39 to Theorem 5.40. ∎

5.4 Parallel Sums

As we already seen in Section 2.2, the S-boxes act in parallel on j components of V each and then the mixing layer operates at the same time on all the components, with the goal of spreading differences activating as many S-boxes as possible. For this reason, we want to focus on sums (and hence differences) acting in parallel like S-boxes, in order to obtain conditions for the linearity of the mixing layer λ with respect to these kind of sums. In this way, we can exploit known properties of alternative sums in low dimension (e.g. targeting 4 bit S-boxes) while attacking a bigger cipher, composed by different parallel components. For this reason, we use for each parallel component an alternative sum with $d = n - 2$, which is the best known and most promising setting. Like the case $d = n - 3$, our main targets are Theorems 5.30 and 5.28. While the second result is just a parallelization of the one [7] present in the classic setting, for the first one we obtain different conditions on matrices λ and more specifically we are allowed to modify in a certain sense the condition $C = 0$ we had in the standard $n-2$ case. The fact that we must use λ with a big bottom-left zero block has as a consequence that often a cipher vulnerable with respect to our alternative sum was also vulnerable to different standard attacks, thus destroying our advantage. Removing that condition gives us the opportunity to outperform results obtained considering just a single sum.

5.4.1 Two parallel sums

We start our study with a sum operating on two parallel blocks, of n components each. Let $V = (\mathbb{F}_2)^{2n}$, $x, y \in V$ and denote by x_1 and x_2 the first and the last n components of x and y respectively. We want to define $\circ$ such that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \circ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \circ_1 y_1 \\ x_2 \circ_2 y_2 \end{pmatrix}.$$

Most of our general setting can be transposed to this kind of sum with slight modification. For example, if we take $\circ_1$ and $\circ_2$ as in Theorem 5.10, since standard +-translation naturally acts in parallel (on each component) we obtain for a generic element x the same equation $\tau_x = M_x \sigma_x$, with

$$M_x = \begin{pmatrix} M_{x_1}^{\circ_1} & \mathbb{0} \\ \mathbb{0} & M_{x_2}^{\circ_2} \end{pmatrix}$$

where $M_{x_1}^{\circ_1}, M_{x_2}^{\circ_2}$ are the matrices defined in Theorem 5.10 for x_1 with sum $\circ_1$ and x_2 with sum $\circ_2$ respectively. Of course in this case the weak key space is spanned by the $2d$ components going from $n - d$ to n and from $2n - d$ to $2n$. Thanks to Theorem 5.13, also our new $\circ$ operation can be stored with $(n-d-1)(n-d)$ values with a matrix Θ_i for each $\circ_i$ and computed in polynomial time.

From now on, we focus on the case $d = n - 2$ for both $\circ_1$ and $\circ_2$. As already observed, this gives us access to many useful tools that we can apply on each parallel component. What is left to us is to understand how this components interact with each other. Due to the structure of our sum it is quite simple to obtain an analogous of Theorem 5.28.

Theorem 5.42. *Let $\circ, \diamond$ be two parallel operation, defined by $\circ_1, \circ_2$ and $\diamond_1, \diamond_2$ respectively, such that for all $\circ_i, \diamond_i$ it holds $\dim(W) = n - 2$. Then, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$.*

Proof. Thanks to Theorem 5.28 applied to each of the two sums, we obtain two permutation matrices P_1 sending $\circ_1$ into $\diamond_1$ and P_2 sending $\circ_2$ into $\diamond_2$ via conjugation. We can thus build a matrix

$$P = \begin{pmatrix} P_1 & \mathbb{0} \\ \mathbb{0} & P_2 \end{pmatrix}$$

that sends $\circ$ into $\diamond$ in the same way, since we can write all the M_x in the form given above. ∎

Corollary 5.43. *Let $H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$ and $H_\diamond = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \diamond)$, with $\circ$ and $\diamond$ as above. Then $H_\diamond = H_\circ^g$ for $g \in \mathrm{GL}(V, +)$ given by Theorem 5.42.*

Proof. Since $\mathrm{AGL}(V, \circ)$ is the normalizer of $T_\circ$ and $T_\diamond = T_\circ^g$, we have $\mathrm{AGL}(V, \diamond) = \mathrm{AGL}(V, \circ)^g$ and consequently $\mathrm{GL}(V, \diamond) = \mathrm{GL}(V, \circ)^g$, being $\mathrm{GL}(V, \circ)$ the stabilizer of 0 in $\mathrm{AGL}(V, \circ)$. Finally, the intersection with $\mathrm{GL}(V, +)$ is preserved since $g \in \mathrm{GL}(V, +)$. ∎

Thanks to Corollary 5.43 we can restrict, up to conjugation, to the case $\circ_1 = \circ_2$. Now

$$\Theta = \begin{pmatrix} \mathbf{0} & \mathbf{b} \\ \mathbf{b} & \mathbf{0} \end{pmatrix},$$

with $\mathbf{b} \in (\mathbb{F}_2)^d$, is enough to completely characterize our $\circ$ sum. What we want to obtain now is an equivalent of Theorem 5.30 that will give us a characterization of $H_\circ$. A little preliminar work is needed. As observed above the weak key space for the first sum, that we will denote by $W_{\circ,1}$, is spanned by the components from 3 to n of our space, while the second one ($W_{\circ,2}$) is spanned by components from $n+3$ to $2n$. The definition of weak key space is itself parallel, since it involves only two parallel operations ($\circ$ and $+$). For this reason, it is clear that it holds $W_\circ = W_{\circ,1} \bigoplus W_{\circ,2}$. For the same reason, applying Theorem 5.15 to our case we obtain that $U_\circ$ is generated by $u_1 = (0, 0, \mathbf{b}, 0, 0, \mathbf{0})$ and $u_2 = (0, 0, \mathbf{0}, 0, 0, \mathbf{b})$ and hence it also holds $U_\circ = U_{\circ,1} \bigoplus U_{\circ,2}$.

Lemma 5.44. *For each $\lambda \in H_\circ$, it holds $W_\circ \lambda = W_\circ$ and $U_\circ \lambda = U_\circ$.*

Proof. It follows directly from Lemma 5.29 (the proof is exactly the same in the parallel case). ∎

Theorem 5.45. *Let $\lambda \in (\mathbb{F}_2)^{2n\times 2n}$. Then λ is compatible with $\circ$ if and only if*

$$\lambda = \left(\begin{array}{cc|cc} A_1 & B_1 & A_2 & B_2 \\ C_1 & D_1 & C_2 & D_2 \\ \hline A_3 & B_3 & A_4 & B_4 \\ C_3 & D_3 & C_4 & D_4 \end{array}\right),$$

with

1. *$A_i \in (\mathbb{F}_2)^{2\times 2}$ such that either $A_1, A_4 = 0$ and A_2, A_3 are invertible or vice versa $A_2, A_3 = 0$ and A_1, A_4 are invertible;*

2. *$B_i \in (\mathbb{F}_2)^{2\times n-2}$*

3. *$C_i = \mathbb{0}_{n-2\times 2}$*

4. *$D_i \in (\mathbb{F}_2)^{n-2\times n-2}$ such that either $\boldsymbol{b}D_1 = \boldsymbol{b}D_4 = \boldsymbol{b}$ and $\boldsymbol{b}D_2 = \boldsymbol{b}D_3 = \boldsymbol{0}$ or vice versa $\boldsymbol{b}D_1 = \boldsymbol{b}D_4 = \boldsymbol{0}$ and $\boldsymbol{b}D_2 = \boldsymbol{b}D_3 = \boldsymbol{b}$, following the pattern of A_i (i.e. if A_2, A_3 are invertible $\boldsymbol{b}D_2 = \boldsymbol{b}D_3 = \boldsymbol{b}$). Moreover, the matrix D defined by*

$$D := \begin{pmatrix} D_1 & D_2 \\ D_3 & D_4 \end{pmatrix}$$

must be invertible.

Proof. ($\Rightarrow$) Since, from Lemma 5.44, $W_\circ\lambda = W_\circ$ and $W_\circ$ is spanned by vectors from 3 to n and from $n+3$ to $2n$, we must have $C_i = 0$. Moreover, the matrix D defined above must send $W_\circ$ into $W_\circ$ and hence be invertible.
Let's now study blocks A_i. Except from the error part (components from 3 to n and from $n+3$ to $2n$), which can be added to the image since it always become zero after applying dot product due to the definition of weak key space and the dot product itself, we observe that span $\{e_1, e_2\}$ can remain fixed or be switched with span $\{e_{n+1}, e_{n+2}\}$. Other linear combinations of these spaces are not possible. To see this, suppose that $e_1\lambda = e_1 + e_{n+1}$ (other cases are equivalent). For linearity with respect to $\cdot$ we obtain

$$0 = (e_1 \cdot e_{n+1})\lambda = e_1\lambda \cdot e_{n+1}\lambda$$

which implies that $e_{n+1}\lambda$ must have components 2 and $n+2$ equals to zero, since they will produce non-zero errors when dotted against $e_1 + e_{n+1}$. We also have $0 = e_1\lambda \cdot e_{n+2}\lambda$ and so the same holds for e_{n+2}. Since $W_\circ$ is left fixed, we have that $e_2\lambda$ must generate components 2 and $n+2$ in the image space, being λ invertible, and this is impossible. So there are two possible cases. If span $\{e_1, e_2\}$ and span $\{e_{n+1}, e_{n+2}\}$ are mapped to themselves, we have $A_2, A_3 = 0$ and A_1, A_4 invertible. If they are switched we have A_2, A_3 invertible and $A_1, A_4 = 0$.
Again from Lemma 5.44 we have $U_\circ\lambda = U_\circ$. Moreover, from the previous point we have

$$u_1\lambda = (e_1 \cdot e_2)\lambda = e_1\lambda \cdot e_2\lambda$$

which is u_1 in the first case and u_2 in the second. From the shape of u_2 the constraints on D_i follows.
($\Leftarrow$) If $x \in W_\circ$, then $x\lambda \in W_\circ$, so when we have a dot product between an element of $W_\circ$ and an element outside $W_\circ$ both sides of the compatibility equation are zero. For instance

$$0 = (e_1 \cdot e_3)\lambda = e_1\lambda \cdot e_3\lambda = 0.$$

Inside $W_\circ$ the equation holds again by definition of $W_\circ$. We can restrict ourselves by linearity to check $(e_1 \cdot e_{n+1})\lambda = e_1\lambda \cdot e_{n+1}\lambda$ and $(e_1 \cdot e_2)\lambda = e_1\lambda \cdot e_2\lambda$. In the first case, we observe that the left side is zero by definition of $\cdot$, while the right side is zero by construction of A_i, which sends e_1 and e_{n+1} to different blocks of the parallel sum (as observed before), thus sending their dot product to zero. Hence the equation holds. In the second case, $e_1 \cdot e_2 = u_1$, and both $e_1\lambda$ and $e_2\lambda$ are sent to the same block of the parallel sum, leading to $e_1\lambda \cdot e_2\lambda = u_1$ if A_1 is invertible, $e_1\lambda \cdot e_2\lambda = u_2$ if A_1 is zero. But if A_1 is invertible, $\mathbf{b}D_1 = \mathbf{b}$ and so $u_1\lambda = u_1$; otherwise $u_1\lambda = u_2$. In both cases, the equation holds. ■

5.4.2 m parallel sums

Most of the results obtained above does not exploit the fact that we are using two sums, and can be easily generalized to the case with m parallel sums.
Let $V = (\mathbb{F}_2)^{m\times n}$, and let $x \in V$. We split x in m vectors $x_1, ..., x_m$ of n components each. We want to study a sum $\circ$ such that

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \circ \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} x_1 \circ_1 y_1 \\ \vdots \\ x_m \circ_m y_m \end{pmatrix}.$$

Like the case of two sums every element x defines the corresponding translation $\tau_x = M_x\sigma_x$, with

$$M_x = \begin{pmatrix} M_{x_1}^{\circ_1} & \cdots & \mathbb{0} \\ \vdots & \ddots & \vdots \\ \mathbb{0} & \cdots & M_{x_m}^{\circ_m} \end{pmatrix}$$

where $M_{x_i}^{\circ_i}$ are the matrices defined in Theorem 5.10 for x_i with sum $\circ_i$. Restricting again to the case $d = n - 2$, we easily obtain also the equivalent of Theorem 5.42 and 5.43.

Theorem 5.46. *Let $\circ, \diamond$ be two parallel operation, defined by $\circ_1, ..., \circ_m$ and $\diamond_1, ..., \diamond_m$ respectively, such that for all $\circ_i, \diamond_i$ it holds $dim(W(T)) = n - 2$. Then, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$.*

Corollary 5.47. *Let $H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$ and $H_\diamond = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \diamond)$, with $\circ$ and $\diamond$ as above. Then $H_\diamond = H_\circ^g$.*

This allows us to restrict again, up to conjugation, to the case $\circ_1 = ... = \circ_m$,

characterized by $\mathbf{b} \in (\mathbb{F}_2)^d$ and

$$\Theta = \begin{pmatrix} \mathbf{0} & \mathbf{b} \\ \mathbf{b} & \mathbf{0} \end{pmatrix}.$$

It is now convenient to introduce some new notation. First of all, we slightly change the numeration of components. Let e_1^1 the first basis vector of the first parallel space (e_1), e_2^1 the second one, and so on up to e_n^1. $e_1^2, ..., e_n^2$ will be the component of the second parallel space, with $e_1^2 = e_{n+1}$ in the standard notation. This goes on until the last parallel space, whose components are $e_1^m, ..., e_n^m$ (instead of $e_{m(n-1)+1}, ..., e_{mn}$). We can also numerate the parallel spaces, with V_i which is spanned by components e_j^i for $j = 1, ..., n$. Our $\circ$ sum is acting on every V_i. We assumed that this sum has a weak key space spanned by the last $n-2$ components, and this is true for each one of the V_i. We may then define the i-th weak key space $W_{\circ,i}$ spanned by $e_3^i, ..., e_n^i$ and call for brevity the i-th strong key space the subspace generated by components e_1^i and e_2^i. We also write

$$x = (\overline{x}_1, \widetilde{x}_1, ..., \overline{x}_m, \widetilde{x}_m)$$

where $\overline{x}_i$ are the components in the i-th strong key space we just defined, and $\widetilde{x}_i$ are the ones in the i-th weak space. Finally, we call $\widetilde{x} = (\widetilde{x}_1, ..., \widetilde{x}_m)$ the vector made of the weak part of each parallel component, and $\overline{x} = (\overline{x}_1, ..., \overline{x}_m)$ the one made of the strong parts.
In this setting, Lemma 5.44 clearly holds with no modifications. We can again obtain a generalization of Theorem 5.45, with only small differences from the case with only two sums.

Theorem 5.48. *Let $\lambda \in (\mathbb{F}_2)^{(n\times m)\times(n\times m)}$. Then λ is compatible with $\circ$ if and only if it can be splitted in blocks*

$$\lambda = \left(\begin{array}{cc|c|cc} A_{11} & B_{11} & & A_{1m} & B_{1m} \\ C_{11} & D_{11} & \cdots & C_{1m} & D_{1m} \\ \hline \multicolumn{2}{c|}{\vdots} & \ddots & \multicolumn{2}{c}{\vdots} \\ \hline A_{m1} & B_{m1} & & A_{mm} & B_{mm} \\ C_{m1} & D_{m1} & \cdots & C_{mm} & D_{mm} \end{array}\right),$$

with

1. *$A_{ij} \in (\mathbb{F}_2)^{2\times 2}$ such that for each row and each column there is one and only one non-zero A_{ij}; moreover, all the non-zero A_{ij} must be invertible*

2. *$B_{ij} \in (\mathbb{F}_2)^{2\times(n-2)}$*

3. *$C_{ij} = \mathbb{0}_{(n-2)\times 2}$*

4. *$D_{ij} \in (\mathbb{F}_2)^{(n-2)\times(n-2)}$ such that if A_{ij} is zero $\mathbf{b}D_{ij} = \mathbf{0}$, and if A_{ij} is*

invertible $\boldsymbol{b}D_{ij} = \boldsymbol{b}$*. Moreover, the matrix* D *defined by*

$$D := \begin{pmatrix} D_{11} & \cdots & D_{1m} \\ \vdots & \ddots & \vdots \\ D_{m1} & \cdots & D_m \end{pmatrix}$$

must be invertible.

Proof. ($\Rightarrow$) Again from Lemma 5.44 $W_\circ\lambda = W_\circ$ implies $C_{ij} = 0$. As a consequence, $W_\circ$ is generated by $\widetilde{x}D$, hence D must be invertible.
Let us now focus on A_{ij}. The argument used in the proof of Theorem 5.45 allows us to prove that every vector with one non-zero strong component $\overline{x}_i$ will be sent by λ in a vector y with one and only one non-zero strong component $\overline{x}_j$. Let us see how. We know that in general $\overline{e}^j \cdot \overline{e}^k = 0$ (here by $\overline{e}^j$ we denote a generic vector in the j-th strong space, which can be $e_1^j, e_2^j, e_1^j + e_2^j$). This implies that $\overline{e}^j\lambda \cdot \overline{e}^k\lambda = 0$ for each $j \neq k$. Since we are studying the A_{ij}, we only focus on strong spaces. Let us suppose that the strong space components of the image of $\overline{e}^j$ through λ are not contained into a single subspace V_i, but into some of them. Without loss of generality, we may suppose that they are included into two subspaces V_{k_1} and V_{k_2}. The dimension of the image of $\overline{e}^j$ is two, while the dimension of the whole subspace spanned by $\overline{e}^{k_1}$ and $\overline{e}^{k_2}$ is four. For this reason, we need at least one other subspace $\overline{e}^k$ whose image is contained in $\overline{e}^{k_1}$ and different from the image of $\overline{e}^j$. This implies that for some vectors it will result $\overline{e}^j\lambda \cdot \overline{e}^k\lambda \neq 0$, which is absurd. As a consequence, if $\overline{e}^i$ is mapped into $\overline{e}^j$, A_{ij} must be invertible, while all the A_{ik} and A_{lj} are zero for $k \neq j$ and $l \neq i$.
The conditions on D_{ij} are the same of the case with two sums, and in the same way follows directly from the ones on the A_{ij}. Indeed, if e_1^i and e_2^i are sent to a single subspace e^j, the same is true for $u_i = e_1^i \cdot e_2^i$. Applying Lemma 5.44 it must be $\mathbf{b}D_{ij} = \mathbf{b}$ and $\mathbf{b}D_{ik} = \mathbf{b}D_{lj} = \mathbf{0}$ for each $k \neq j$, $l \neq i$.
($\Leftarrow$) For the vice-versa, notice that in the proof of Theorem 5.45 we never exploited the fact that we had just two sums. For linearity, we may restrict ourselves to check only two cases, which are the same of Theorem 5.45: $(e_1^j \cdot e_2^j)\lambda = e_1^j\lambda \cdot e_2^j\lambda$ (vectors coming from the same strong subspace) and $(e_1^j \cdot e_2^k)\lambda = e_1^j\lambda \cdot e_2^k\lambda$ (vectors coming from different subspaces). Both equations are satisfied thanks to the structure of λ, as already shown in the previous proof. ∎

Chapter 6

Analysis of Optimal 4-bit S-boxes

In this last chapter, we will show in greater detail what we suggested in Section 4.2, i.e. that our alternative sums are capable of significantly increase the differential uniformity of S-boxes. Our target will be the optimal 4-bit S-boxes presented in Chapter 3. Recall that we obtained 16 out of 302 equivalence classes of optimal permutations, and that we were able to do so thanks to Proposition 1.80 which stated that differential uniformity is invariant under affine equivalence. However, this is not true in general for $\circ$-differential uniformity, since the affine equivalence we considered was obtained through linear maps with respect to $+$. We are then left with 16 classes of optimal permutations, each one actually composed by about 2^{36} single permutations, with (potentially) different $\circ$-differential uniformity. Of course extensive computation of differential properties of this much S-boxes is not feasible. Some preliminary work is hence required.
Since we are dealing with 4-bit S-boxes (and so $n = 4$), our natural choice is a sum $\circ$ such that $d = n-2 = 2$. First of all, this is suggested from the aforementioned considerations about the optimality of this case. Moreover, notice that 4-bit maps are not likely to be used a cipher themselves, but instead to constitute a parallel confusion layer. Thanks to Section 5.4, we already know how sums with $d = n - 2$ behave when put to work in parallel. For $n = 4$, we have to choose among 105 sums for $d = n - 2$. Notice that the fact, extensively used through all this work, that those sums are conjugated each other is of no help here. However, we may restrict for now to consider only sums characterized in Theorem 5.10; we are then only left with the choice of $\mathbf{b} \in \{(0,1),(1,0),(1,1)\}$, i.e. with three possible sums. Let us for now fix one of them, for example the one corresponding to $\mathbf{b} = (1,0)$. Our goal is to somehow replicate the result of [22], avoiding the repeated computation of functions that have the same DDT with respect to both the classic $+$ sum and our $\circ$ one.

Proposition 6.1. *Given f a permutation on V and $g_1, g_2 \in \mathrm{GL}(V, \circ)$ it holds*

$$\delta^{\circ}_{g_1 \cdot f \cdot g_2}(a,b) = \delta^{\circ}_{f}(g_2(a), g_1^{-1}(b))$$

Proof. Given $g_1, g_2 \in \mathrm{GL}(V, \circ)$ we have

$$\delta^{\circ}_{g_1 \cdot f \cdot g_2}(a, b) = \#\left\{g_1 \cdot f \cdot g_2(x \circ a) \circ g_1 \cdot f \cdot g_2(x) = b\right\} = \#\left\{f(g_2(x) \circ g_2(a)) \circ f(g_2(x)) = g_1^{-1}(b)\right\}$$

because both g_1 and g_2 are linear with respect to $\circ$. Since $g_2 \in \mathrm{GL}(V, \circ)$, $g_2(x)$ goes through all the elements of V and can hence be replaced by x. We obtain

$$\delta^{\circ}_{g_1 \cdot f \cdot g_2}(a, b) = \delta^{\circ}_{f}(g_2(a), g_1^{-1}(b))$$

which is the desired equation. ∎

Since $H_{\circ}(V) \subseteq \mathrm{GL}(V, \circ)$, Proposition 6.1 holds for every $g \in H_{\circ}$. As a consequence, multiplying on the left or on the right a function for an element $g \in H_{\circ}$ preserves δ°-differential uniformity as well as the δ^{+}-differential uniformity, since it also holds $H_{\circ}(V) \subseteq \mathrm{GL}(V, +)$. Moreover, from the proof of this proposition we get the extra property that when multiplying by elements in $H_{\circ}$ the rows of DDT° are shuffled, but the highest elements of each row remain unchanged, with the only difference that the max of row a now becomes the max of row $g_2(a)$ and so on. This fact can be very useful, since we are mainly interested in the highest differential ($\delta^{\circ}(f)$), but there may be other extremely high differentials on other rows which can considerably empower our attack.

Proposition 6.2. *Given σ_c a translation on V (with respect to XOR) and f a permutation of V then $f, \sigma_c \cdot f$ and $f \cdot \sigma_c$ have the same differential δ-uniformity with respect to $\circ$.*

Proof. Since the sum $\circ$ is defined by a translation group $T_{\circ} < AGL(V, +)$ such that $T_{+} < AGL(V, \circ)$, for each XOR translation σ_c there exists $M_c \in GL(V, \circ)$ such that $\sigma_c = M_c \tau_c$, with τ_c translation with respect to $\circ$.
We have $f\sigma_c(x) = f(xM_c\tau_c)$, but since $M_c \in GL(V, \circ)$ xM_c goes through all the elements of V and can be replaced by x without altering the DDT as observed before. The differential uniformity equation becomes $f((x \circ a)\tau_c) \circ f(x\tau_c) = f(x \circ c \circ a) \circ f(x \circ c)$. We can now replace $x \circ c$ with x, again with no effect on the DDT since τ_c is a translation.
On the other hand, $\sigma_c f = M_c \tau_c f$ but since M_c is linear with respect to $\circ$ we can multiply the whole equation by M_c^{-1} moving the effect on the b's. We then obtain the same DDT with the columns shuffled. The equation becomes $f(x \circ a) \circ c \circ f(x) \circ c = f(x \circ a) \circ f(x)$, which is the equation for f. ∎

Proposition 6.2 exploits the requirements we imposed on our sum in order to get rid of translations in affine equivalence. Thanks to this, we can reduce our inspection of the optimal class represented by f to elements of the form $g_1 \cdot f \cdot g_2$ for $g_1, g_2 \in \mathrm{GL}(V, +)$. Merging this result with Proposition 6.1 we can furthermore restrict to choose g_1 and g_2 in $\mathrm{GL}(V, +)$ quotiented by $H_{\circ}$. Finally, for each of the 16 optimal S-boxes we are left to check only 10816 permutations, which is absolutely affordable. Notice that for each different $\circ$ sum, also the automorphism group $H_{\circ}$ end hence the maps g_1 and g_2 that we must consider

are different.
All these computation are made through the MAGMA Software [4]. Moreover, the implementation of optimal S-boxes is due to [20]; for this reason, the equivalence classes are denoted by P_n, following the numeration given in [20], and the chosen representatives are different but affine equivalent to the ones given above. These are the hexadecimal description of the maps used:

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
P_{16}	0	1	2	B	4	C	9	F	8	5	D	6	7	3	E	A
P_{20}	0	1	2	7	4	3	B	D	8	6	C	E	5	A	F	9
P_{35}	0	1	2	B	4	C	D	F	8	5	9	6	3	7	E	A
P_{102}	0	1	2	C	4	E	F	9	8	5	D	7	6	3	A	B
P_{128}	0	1	2	5	4	B	A	6	8	C	E	D	F	9	3	7
P_{220}	0	1	2	C	4	6	F	9	8	D	B	3	E	5	7	A
P_{227}	0	1	2	E	4	C	7	F	8	B	6	D	3	9	A	5
P_{243}	0	1	2	A	4	3	C	9	8	F	E	6	D	7	5	B
P_{245}	0	1	2	A	4	3	7	B	8	5	C	D	F	9	6	E
P_{249}	0	1	2	5	4	7	A	F	8	3	6	B	9	C	D	E
P_{254}	0	1	2	6	4	C	9	5	8	F	D	3	E	B	7	A
P_{262}	0	1	2	B	4	E	F	D	8	5	6	9	3	7	C	A
P_{267}	0	1	2	3	4	E	A	D	8	6	C	5	B	F	7	9
P_{275}	0	1	2	6	4	D	A	7	8	C	E	F	3	B	9	5
P_{280}	0	1	2	D	4	9	5	3	8	C	E	A	B	7	6	F
P_{296}	0	1	2	E	4	A	7	F	8	5	9	D	6	C	B	3

The results obtained are summarized in Tables 6.1, 6.2 and 6.3 at the end of this chapter. Each row represents an S-box, while each column contains the number of permutations affine equivalent to that S-box with a specific δ°-differential uniformity. We can consider these as overall good results. Surprisingly, for each sum some optimal S-boxes become 16-uniform, which is great. In many other cases, we at least managed to obtain 12-uniformity. Since, as already explained, difference may come out from the key addition layer either unchanged or modified by a single error (the only non zero element of $U_\circ$, which in our setting is the vector $(0, 0, \mathbf{b})$) it is probably more appropriate to compare the standard δ-differential uniformity of an S-box with the half of its δ°-differential uniformity. Anyway, being able to pass from 4 to 6 and sometimes even to 8 is a significant improvement.
We conclude by showing two concrete examples of good DDT° computations, together with the description of maps used to generate them. For each one, i, g_1 and g_2 are given. The S-box is then obtained by composition as $g_1 P_i g_2$, where P_i is one of the optimal permutations listed above. The sum considered in both is the one corresponding to $\mathbf{b} = (1, 0)$.

Example 6.3. For this first example, we will see a 16-uniform S-box. It is affine equivalent to the optimal permutation P_{102} through the maps g_1 and g_2 defined by

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
g_1	0	B	4	F	8	3	C	7	6	D	2	9	E	5	A	1
g_2	0	D	F	2	4	9	B	6	3	E	C	1	7	A	8	5

And its DDT° is listed below.

$\circ$	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1	·	·	·	·	4	4	4	4	·	·	·	·	·	·	·	·
2	·	·	·	·	4	·	·	4	4	·	·	4	·	·	·	·
3	·	4	4	·	·	·	·	·	·	·	·	·	·	4	4	·
4	·	4	·	4	·	·	·	·	·	·	·	·	·	4	·	4
5	·	·	·	·	·	4	·	4	4	·	4	·	·	·	·	·
6	·	·	·	·	·	·	4	4	·	·	4	4	·	·	·	·
7	·	·	4	4	·	·	·	·	·	·	·	·	·	·	4	4
8	·	·	·	·	4	4	·	·	4	4	·	·	·	·	·	·
9	·	·	4	4	·	·	·	·	·	·	·	·	·	·	4	4
A	·	4	·	4	·	·	·	·	·	·	·	·	·	4	·	4
B	·	·	·	·	4	·	4	·	·	4	·	4	·	·	·	·
C	·	·	·	·	·	·	·	·	·	·	·	·	16	·	·	·
D	·	·	·	·	·	·	·	·	4	4	4	4	·	·	·	·
E	·	·	·	·	·	4	4	·	·	4	4	·	·	·	·	·
F	·	4	4	·	·	·	·	·	·	·	·	·	·	4	4	·

It means that any time we have two input whose $\circ$ difference is 1100, no matter what the inputs are, the $\circ$ difference between the outputs will always be 1100. Notice that in general it is not required that the input difference it is equal to the output difference.

Example 6.4. Our second example targets an S-box affine equivalent to P_{16}, through the maps

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
g_1	0	C	4	8	1	D	5	9	2	E	6	A	3	F	7	B
g_2	0	B	7	C	9	2	E	5	6	D	1	A	F	4	8	3

Its DDT° is then

$\circ$	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1	·	·	·	·	·	4	·	·	·	·	·	4	·	8	·	·
2	·	·	·	2	·	·	2	·	2	4	·	·	·	·	4	2
3	·	·	2	4	·	·	4	2	·	2	·	·	·	·	2	·
4	·	·	·	·	8	4	·	·	·	·	·	4	·	·	·	·
5	·	·	·	·	·	·	·	·	·	·	·	8	8	·	·	·
6	·	·	2	·	·	·	·	2	4	2	·	·	·	·	2	4
7	·	·	4	2	·	·	2	4	2	·	·	·	·	·	·	2
8	·	·	4	2	·	·	2	4	2	·	·	·	·	·	·	2
9	·	·	2	·	·	·	·	2	4	2	·	·	·	·	2	4
A	·	4	·	·	·	·	·	·	·	·	·	·	8	4	·	·
B	·	8	·	·	4	·	·	·	·	·	4	·	·	·	·	·
C	·	·	·	·	4	·	·	·	·	·	12	·	·	·	·	·
D	·	4	·	·	·	8	·	·	·	·	·	·	·	4	·	·
E	·	·	2	4	·	·	4	2	·	2	·	·	·	·	2	·
F	·	·	·	2	·	·	2	·	2	4	·	·	·	·	4	2

We see that it is 12-uniform. Moreover, remarkably 6 rows have a maximum value of 8. This fact may help a lot when mounting a differential attack, since we can explore many different trails with high probability. Finally, we see that the differential table is sparse, i.e. there are few nonzero values. This gives us a greater control on $\circ$-difference propagation through this S-box.

Table 6.1: δ°-differential uniformity of optimal S-boxes; $\circ$ is defined by $\mathbf{b} = (0, 1)$

	2	4	6	8	10	12	14	16
P_{16}	0	1124	7052	2463	141	36	0	0
P_{20}	0	1112	7433	2123	148	0	0	0
P_{35}	0	1077	7104	2452	147	36	0	0
P_{102}	0	681	6935	2825	363	0	0	12
P_{128}	0	918	7757	1956	173	12	0	0
P_{220}	0	890	7561	2217	148	0	0	0
P_{227}	0	1116	7091	2410	151	48	0	0
P_{243}	0	1127	7093	2399	151	46	0	0
P_{245}	0	903	7611	2153	131	18	0	0
P_{249}	0	840	6571	3058	297	40	0	10
P_{254}	0	1081	7269	2311	155	0	0	0
P_{262}	0	782	6662	2985	359	16	0	12
P_{267}	0	1188	7284	2184	160	0	0	0
P_{275}	0	776	6712	2947	353	16	0	12
P_{280}	0	806	7990	1804	216	0	0	0
P_{296}	0	1105	7771	1824	116	0	0	0

Table 6.2: δ°-differential uniformity of optimal S-boxes; $\circ$ is defined by $\mathbf{b} = (1, 0)$

	2	4	6	8	10	12	14	16
P_{16}	0	1117	7050	2472	141	36	0	0
P_{20}	0	1114	7439	2111	152	0	0	0
P_{35}	0	1072	7115	2444	149	36	0	0
P_{102}	0	676	6916	2833	379	0	0	12
P_{128}	0	923	7768	1936	177	12	0	0
P_{220}	0	889	7563	2216	148	0	0	0
P_{227}	0	1112	7094	2413	149	48	0	0
P_{243}	0	1128	7097	2396	147	48	0	0
P_{245}	0	903	7649	2109	135	20	0	0
P_{249}	0	853	6510	3094	299	48	0	12
P_{254}	0	1094	7279	2288	155	0	0	0
P_{262}	0	778	6680	2969	361	16	0	12
P_{267}	0	1205	7296	2159	156	0	0	0
P_{275}	0	784	6740	2921	347	14	0	10
P_{280}	0	815	7996	1793	212	0	0	0
P_{296}	0	1109	7771	1816	120	0	0	0

Table 6.3: δ°-differential uniformity of optimal S-boxes; $\circ$ is defined by $\mathbf{b} = (1, 1)$

	2	4	6	8	10	12	14	16
P_{16}	0	1123	7051	2467	139	36	0	0
P_{20}	0	1117	7432	2117	150	0	0	0
P_{35}	0	1064	7105	2462	149	36	0	0
P_{102}	0	676	6976	2793	361	0	0	10
P_{128}	0	907	7740	1982	175	12	0	0
P_{220}	0	894	7554	2222	146	0	0	0
P_{227}	0	1121	7089	2413	145	48	0	0
P_{243}	0	1114	7086	2419	149	48	0	0
P_{245}	0	896	7627	2144	129	20	0	0
P_{249}	0	865	6571	3040	282	48	0	10
P_{254}	0	1091	7263	2309	153	0	0	0
P_{262}	0	766	6722	2953	347	16	0	12
P_{267}	0	1205	7296	2159	156	0	0	0
P_{275}	0	785	6705	2953	347	14	0	12
P_{280}	0	812	7991	1799	214	0	0	0
P_{296}	0	1105	7771	1824	116	0	0	0

Ringraziamenti

Mi è doveroso dedicare questo spazio del mio elaborato alle persone che hanno contribuito, con il loro instancabile supporto, alla realizzazione dello stesso.
In primis, un ringraziamento particolare al mio correlatore, prof. Calderini, per avermi fatto scoprire nuove strade, oltre che per i suoi indispensabili consigli e le conoscenze trasmesse durante tutto il percorso di stesura dell'elaborato. Con lui, desidero ringraziare anche il mio relatore, prof. Canonaco, per la sua grande disponibilità e per avermi trasmesso nel corso di questi anni la passione per la matematica e per l'algebra.
Ringrazio infinitamente i miei genitori, che mi hanno sempre sostenuto, anche nei momenti più difficili, i miei fratelli, che mi sono sempre stati vicino, e i nonni, che non hanno mai fatto mancare un abbraccio.
Non posso non rigraziare chi ha contribuito a farmi arrivare fin qui, partendo da lontano: Bianca, Chiara, Fra, e specialmente Raffa, per tutto quello che abbiamo passato insieme. Le persone più grandi, che mi hanno dato molti consigli, come 'zio' Fausto e Willy, allenatore oltre che professore. E poi Brock e Spock, mentori ma anche e soprattutto grandi amici.
Un ringraziamento speciale ad Ari, una persona speciale.
Infine, ma non certo meno importanti, un ringraziamento doveroso a Dotti e Pitucci, per avermi accompagnato in questi cinque anni. Con loro, desidero ringraziare tutti i Pidotti, Aber, per essere sempre stato una roccia (dalle dimensioni smisurate) a cui appoggiarsi in ogni difficoltà, Adi, per aver sempre avuto un ritmo sonno/veglia forse peggiore del mio, Ago, per le tante avventure in cui non ne abbiamo mai fatta una giusta, Ariel, per essere sempre stato il più grasso di tutti, tanto da vincere la sfida dei 40 Kinder fetta al Dotti, Boa, per avermi venduto sua madre per 3000 lire, Carletto, per i pomeriggi passati a soffrire davanti alla F1, Despa, per le scampagnate in moto, Furia, per essersi preso cura con costanza e determinazione della sanificazione della mia testa e del mio tragitto fino alla nave, Giggino, per la sua risata malefica, Goblin, per la sua grande ospitalità, volontaria e non, Norman, per i suoi lungimiranti consigli, Semola, per l'indicibile quantità di tonno che giace nel suo corpo, Tucano, per aver sempre vegliato su di noi, e Zuppa, per avermi organizzato la festa di compleanno migliore di sempre. Ovviamente, oltre a questo c'è molto altro, difficile da rendere su un foglio di carta. Ma dato che questo foglio di carta lo leggeranno pochissime persone, credo possa bastare.
Per concludere, un ringraziamento sentito anche a tutti quelli che non ho potuto citare qui, ma che in molti modi hanno accompagnato e allietato il mio percorso fin qui.

Bibliography

[1] F. Abazari, B. Sadeghiyan, Cryptanalysis with ternary difference: applied to block cipher PRESENT, *Int. J. Inf. Electron. Eng.* 2 (3) (2012) 441.

[2] T.A. Berson, Differential cryptanalysis mod 2^{32} with applications to MD5, in: *Workshop on the Theory and Application of Cryptographic Techniques*, Springer, (1992) pp. 71–80.

[3] E. Biham, A. Shamir, Differential cryptanalysis of DES-like cryptosystems, *J. Cryptol.* 4 (1) (1991) 3–72.

[4] W. Bosma, J. Cannon, C. Playoust, The Magma algebra system. I: The user language, *J. Symbolic Comput.*, 24 (1997) 235–265.

[5] C. Brunetta, M. Calderini, M. Sala, On hidden sums compatible with a given block cipher diffusion layer, *Discrete Math.* 342 (2) (2019) 373–386.

[6] M. Calderini, On Boolean functions, symmetric cryptography and algebraic coding theory (Doctoral dissertation, University of Trento) (2015).

[7] M. Calderini, R. Civino, M. Sala, On properties of translation groups in the affine general linear group with applications to cryptography, *J. of Algebra* 569 (2021) 658–680

[8] M. Calderini, M. Sala, Elementary abelian regular subgroups as hidden sums for cryptographic trapdoors, *arXiv preprint* arXiv:1702.00581 (2017).

[9] M. Calderini, M. Sala, On differential uniformity of maps that may hide an algebraic trapdoor, in: *International Conference on Algebraic Informatics*, Springer, (2015) 70–78.

[10] A. Caranti, F. Dalla Volta, M. Sala, Abelian regular subgroups of the affine group and radical rings, *Publ. Math. (Debr.)* 69 (3) (2006) 297–308.

[11] C. Carlet, Boolean functions for cryptography and coding theory, *Cambridge University Press* (2021).

[12] R. Civino, C. Blondeau, M. Sala, Differential attacks: using alternative operations, *Des. Codes Cryptogr.* 87 (2–3) (2019) 225–247.

[13] D. Coppersmith, The Data Encryption Standard (DES) and its strength against attacks, *IBM Journal of Research and Development*, 38 (3) (1994) 243–250.

[14] D. Coppersmith, E. Grossman, Generators for certain alternating groups with applications to cryptography, *SIAM J. Appl. Math.* 29 (4) (1975) 624–627.

[15] N. Courtois, J. Pieprzyk, Cryptanalysis of block ciphers with overdefined systems of equations, in: *Zheng, Y. (ed.) ASIACRYPT 2002. LNCS*, 2501 (2020) 267–287.

[16] J. Daemen, V. Rijmen, *AES proposal: Rijndael*, Tech. report, NIST, 1998, http://www.nist.gov/aes.

[17] J. D. Dixon, Maximal abelian subgroups of the symmetric groups, *Canadian Journal of Mathematics*, 23(3) (1971) 426-438.

[18] A. Dotti, P. Coelho, T. Junior, A. Bessani, F. Pedone, Byzantine fault-tolerant atomic multicast, in: *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, IEEE, (2018) 39-50

[19] H. M. Heys, A tutorial on linear and differential cryptanalysis, *Cryptologia*, 26(3) (2002) 189-221.

[20] M. Iavernaro, On some cryptographic properties of vectorial Boolean functions. Master Thesis, University of Trento (2015).

[21] G. Lachaud, J. Wolfmann, The weights of the orthogonals of the extended quadratic binary Goppa codes, *IEEE Transactions on Information Theory* 36(3) (1990) 686

[22] G. Leander, A. Poschmann, On the classification of 4 bit S-boxes, in: *International Workshop on the Arithmetic of Finite Fields* (2007) 159-176

[23] M. Matsui, Linear cryptanalysis method for DES cipher, in: *Workshop on the Theory and Application of Cryptographic Techniques*, Springer, (1993) 386–397.

[24] National Institute of Standards and Technology, Announcing Development of a Federal Information Standard for Advanced Encryption Standard, *Federal Register*, 62 (1) (1997) 93-94.

[25] K. Nyberg, Perfect nonlinear S-boxes, in: *EUROCRYPT 1991. LNCS*, Springer, 547 (1991) 378–386.

[26] C. E. Shannon, Communication Theory of Secrecy Systems, *Bell System Technical Journal*, 28 (4) (1949) 656–715.

[27] R. Rivest, A. Shamir, Data Encryption Standard (DES), *Federal Information Processing Standards Publications (FIPS PUBS)* 46 (3) (1999)

[28] D.R. Stinson, *Cryptography: theory and practice*, Chapman and Hall/CRC (2005).