Commutative Cryptanalysis as a Generalization of Differential Cryptanalysis

Jules Baudrin[1], Christof Beierle[2], Patrick Felke[3], Gregor Leander[2], Patrick Neumann[2], Léo Perrin[1], and Lukas Stennes[2]

[1]Inria, Paris, France
firstname.lastname@inria.fr
[2]Ruhr University Bochum, Bochum, Germany
firstname.lastname@rub.de
[3]University of Applied Sciences Emden/Leer, Emden, Germany
patrick.felke@hs-emden-leer.de

Abstract

Recently, Baudrin et al. analyzed a special case of Wagner's commutative diagram cryptanalysis, referred to as *commutative cryptanalysis*. For a family $(E_k)_k$ of permutations on a finite vector space G, commutative cryptanalysis exploits the existence of affine permutations $A, B \colon G \to G$, $I \notin \{A, B\}$ such that $E_k \circ A(x) = B \circ E_k(x)$ holds with high probability, taken over inputs x, for a significantly large set of *weak keys* k. Several attacks against symmetric cryptographic primitives can be formulated within the framework of commutative cryptanalysis, most importantly differential attacks, as well as rotational and rotational-differential attacks. Besides, the notion of c-differentials on S-boxes can be analyzed as a special case within this framework. We discuss the relations between a general notion of commutative cryptanalysis, with A and B being arbitrary functions over a finite Abelian group, and differential cryptanalysis, both from the view of conducting an attack on a symmetric cryptographic primitive, as well as from the view of a theoretical study of cryptographic S-boxes.

Keywords: block cipher, differential uniformity, c-differentials, distinguisher, weak keys (MSC: 94A60, 94D10, 06E30)

1 Introduction

Symmetric cryptography is a crucial building block for protecting our everyday communication. The security of symmetric cryptographic primitives is measured by the absence of any discovered attack through years of public scrutiny by the scientific community and should be supported by arguments why known classes of attacks do not apply. On of the most promising and widely-studied attack vectors is *differential cryptanalysis* [8], a statistical attack that aims to break a cipher by

Parts of the results were presented by the second author as an invited talk at WCC 2024.

tracing the propagation of pairs of inputs $(x, x+\alpha)$ for a fixed input difference α. In a nutshell, a family of permutations $(E_k)_k$ is considered broken by differential cryptanalysis if there exists a non-zero input difference α and an output difference β for which the probability (originally taken over all keys k and inputs x) that $E_k(x+\alpha) = E_k(x) + \beta$ holds is higher than one would expect for a permutation chosen uniformly at random.

Strictly speaking, the probability does not have to be taken over the whole key space and an empirical estimation of this probability could already give information about the key. Moreover, a cipher would already be broken if the probability is higher than expected for a significant subset of all possible keys. This case, that is much harder to analyze in general, is referred to as the *weak-key model* and keys for which the probability is large are called *weak keys*. As we will see below, weak-key attacks are, inherently, of interest in the more general setting.

Because differential cryptanalysis poses a serious threat to cryptographic primitives, designers are expected to provide sufficient arguments for the resistance of their proposed ciphers against these attacks. For key-alternating substitution-permutation-networks (SPNs), one of the most employed arguments is the so-called *wide-trail strategy* [17] that estimates the security of a cipher from a security analysis of its building blocks, i.e., their S-boxes (that are functions on a small input size) and linear layer. The Advanced Encryption Standard (AES) [18] is the most prominent example of a symmetric cryptographic primitive employing the wide-trail strategy for arguing its security.

At EUROCRYPT 1993, Nyberg introduced the notion of *differential uniformity* of a function between two finite Abelian groups as a measure for the resistance against differential attacks.

Definition 1 ([35])**.** *Let $S\colon G_1 \to G_2$ be a function between two finite Abelian groups G_1 and G_2. The* differential uniformity *of S is defined as*

$$\Delta_S \coloneqq \max_{\alpha \in G_1 \setminus \{0\}, \beta \in G_2} |\{x \in G_1 \mid S(x+\alpha) = S(x) + \beta\}|.$$

The wide-trail strategy suggests that choosing an S-box with low differential uniformity together with a "suitable" linear layer provides sufficient resistance against differential cryptanalysis. This motivates the study of the differential uniformity (and more general the differential spectrum, i.e., the multiset of values $|\{x \in G_1 \mid S(x+\alpha) = S(x) + \beta\}|$ over all $\alpha \neq 0$ and β) of S-boxes in a kind of isolated manner. Indeed, the definition of differential uniformity triggered a significant amount of research in mathematics and cryptography. The most prominent line of research focuses on functions with optimal values for their differential uniformity in case of G_1 and G_2 being elementary Abelian p-groups. From a cryptographic point of view, the case of $p = 2$ is the most important (as most of the ciphers are defined over an $\mathbb{F}_2$-vector space).[1] In that case, the optimal value on the differential uniformity Δ_S is 2 and in the case of $G_1 = G_2 = \mathbb{F}_2^n$, functions achieving this optimal value are called *almost perfect nonlinear (APN) functions.* In the case of odd p and $G_1 = G_2$, the optimal value on Δ_S is 1 and functions achieving this optimum are called *perfect nonlinear* or *planar functions.* APN and planar functions are also of interest in finite geometry and combinatorics. We refer to the book by Carlet [13, Chapter 11] and the survey by Pott [36] for more information on the significant amount of research conducted in this area. Within the recent years, the notion of differential uniformity was generalized in various ways and studied from a mathematical point of view. A particular kind of generalization, attracting lots of interest, is the notion of *c-differential uniformity.*

[1] Recently, there is a significant amount of research on so-called *arithmetization-oriented primitives*, with many ciphers defined over a field of odd characteristic.

Definition 2 ([20])**.** *Let $S\colon G \to G$ be a function on a finite field G and let $c \in G$. The* c-differential uniformity *of S is defined as*

$$_c\Delta_S \coloneqq \max_{\alpha\in G,\beta\in G,\alpha\neq 0 \text{ if } c=1} |\{x \in G \mid S(x+\alpha) = c \cdot S(x) + \beta\}|.$$

The notion of differential uniformity corresponds to the notion of c-differential uniformity for $c = 1$. Quite some papers appeared studying the general notion of c-differential uniformity (and more generally the c-differential spectrum) of functions, see e.g., [34, Section 5] for a survey. However, to the best of our knowledge, the notion of c-differential uniformity (with $c \neq 1$) was never successfully applied to attack a cryptographic primitive and *an application of the framework of c-differentials to cryptography is yet to be shown.* The preprint [2] already questioned its applicability for building cryptographic attacks due to a non-deterministic propagation of c-differentials through a linear layer and a key addition within a cipher.

c-differentials, as well as other existing cryptographic attacks such as rotational cryptanalysis [27] and rotational-differential cryptanalysis [1] can be formulated within the unifying framework of *commutative cryptanalysis* [37, 4]. In a nutshell, if $(E_k)_k$ is a family of permutation on a finite Abelian group G, commutative cryptanalysis exploits the existence of functions $A, B\colon G \to G$ such that $E_k \circ A(x) = B \circ E_k(x)$ holds with high probability (taken over inputs x) for a significantly large set of weak keys k. In this framework, it is crucial to impose further restrictions on the choice of A and B to avoid trivial properties (e.g., both A and B being the identity). Besides differential attacks, as well as rotational attacks and rotational-differential attacks, we have examples of commutative attacks (with probability 1) in the weak-key model, when A and B are restricted to be affine permutations on a finite vector space G, see [4].

Our Contribution. In this work, we discuss the relations between a general commutative attack and differential cryptanalysis. Our results are grouped into four parts:

1. In Section 3, we discuss the applicability of the general framework of commutative cryptanalysis with respect to conducting an attack against a cryptographic primitive. Under some simplifying assumptions, we outline why a commutative attack more general than a differential attack is *necessarily in the weak-key model*, and is not applicable in the case of independent whitening keys. In particular, we see that c-differentials with $c \neq 1$ belong to the class of distinguishers from this framework that has the least potential for an attack. The main results of this section are stated in Corollaries 1 and 2.

2. Motivated by applications of the commutative framework and the existing examples in the fixed/weak key model, in Section 4 we then study S-boxes that are vulnerable to commutative attacks and show lower bounds on the differential uniformity of such S-boxes. These bounds link the number of inputs for which a commutative property holds to the number of weak-keys in a trail-based attack and to the differential uniformity. As the most important special case, the focus is on commutative properties with affine permutations A, B (in which case G is a finite vector space). In a nutshell, we show that an S-box possessing a non-trivial deterministic affine commutative property (i.e., a non-trivial affine self-equivalence) that allows for many weak keys necessarily has high differential uniformity.

3. We discuss in Section 5 the mathematically interesting question of how to generate affine permutations A, B having the same cycle type and from this the question how to generate all

S-boxes S for which $S \circ A = B \circ S$. The first question is a classical problem studied in a series of papers and our contribution is to give an, as far as we know, first comprehensive survey on this topic. We close this section by analyzing the effect on the self-equivalence if the order of A and thus B is not a power of p.

4. As the last part, we discuss in Section 6 a generalization of differential cryptanalysis using alternative group operations (see [15]) and present links between these kind of generalized attacks, differential cryptanalysis of conjugate ciphers, and the commutative cryptanalysis framework. We show in particular that the "alternative differential" trails considered in [15] can be interpreted as commutative trails, or as regular differential trails of a conjugate cipher. We then provide a detailed analysis of some of the toy ciphers illustrating those earlier results using our own framework.

2 Preliminaries

Throughout this work, let $(G, +)$ be a finite Abelian group. For an element $\alpha \in G$, we denote by $T_\alpha \colon G \to G$ the translation $x \mapsto x + \alpha$. As an important and relevant special case, we will focus on G having the additional structure of a vector space, i.e., $G = \mathbb{F}_p^n$ for a prime p. We denote by $\mathrm{GL}(n, \mathbb{F}_p)$ the general linear group of degree n over $\mathbb{F}_p$. By $\mathrm{AGL}(n, \mathbb{F}_p)$, we denote the set of all affine bijections over $\mathbb{F}_p^n$ and by I the identity in $\mathrm{AGL}(n, \mathbb{F}_p)$. With $e_1, \dots, e_n$ we denote the canonical unit vectors in $\mathbb{F}_p^n$, i.e. $e_1 = (1, 0 \dots, 0)^t, e_2 = (0, 1, 0, \dots, 0)^t, \dots$.

A function $S \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ can be uniquely expressed in its *algebraic normal form*, i.e., a multivariate polynomial expression of the form

$$S(x) = \sum_{u \in \mathbb{F}_p^n} a_u x^u,$$

where $x^u := \prod_{i=0}^{n-1} x_i^{u_i}$ and $a_u \in \mathbb{F}_p^n$. The *algebraic degree* of S is defined as the degree of its algebraic normal form, i.e., $\max\{\sum_{i=0}^{n-1} u_i \mid a_u \neq 0\}$. The affine functions are exactly those of algebraic degree 1. A function of algebraic degree 2 is called *quadratic*.

For a function $S \colon G \to G$ and functions $A, B \colon G \to G$, let

$$\Gamma_S(A, B) := |\{x \in G \mid S \circ A(x) = B \circ S(x)\}|.$$

The notion of differential uniformity then corresponds to the maximum taken over translations A, B of $\Gamma_S(A, B)$, i.e., $\Delta_S = \max_{\alpha \in G \setminus \{0\}, \beta \in G} \Gamma_S(T_\alpha, T_\beta)$. If G is a finite field $\mathbb{F}_q$ and $m_c \colon G \to G, x \mapsto cx$ the multiplication by $c \in G$, for the c-differential uniformity ${}_c\Delta_S$ of S, we have

$${}_c\Delta_S = \max_{\alpha, \beta \in G, \alpha \neq 0 \text{ if } c=1} \Gamma_S(T_\alpha, T_\beta \circ m_c).$$

For a mapping $A \colon G \to G$, we denote by $\mathrm{Fix}(A)$ the set of fixed points of A, i.e., $\{x \in G \mid A(x) = x\}$. For $C \in \mathrm{AGL}(n, \mathbb{F}_p)$, we have

$$|\mathrm{Fix}(C)| = \begin{cases} 0 & \text{if } c_C \notin \mathrm{Im}(I - L_C) \\ p^{\dim \ker(I - L_C)} & \text{otherwise} \end{cases},$$

where $C = L_C + c_C$ with $L_C = C + C(0)$ being linear and $c_C = C(0)$. In case $\mathrm{Fix}(C)$ is non-empty, it forms an affine subspace of $\mathbb{F}_p^n$. By $\mathrm{ord}(C)$, we denote the order of C, i.e.,

$$\mathrm{ord}(C) \coloneqq \min\{i \in \mathbb{N} \setminus \{0\} \mid C^i \coloneqq \overbrace{C \circ C \circ \cdots \circ C}^{i \text{ times}} = I\}.$$

A *block cipher* over G is a finite family of permutations on G, indexed by a key from a finite key space κ. The notion of *pseudorandom permutation security* is formalized as follows (see e.g., [33]).

Definition 3 (Pseudorandom permutation distinguisher)**.** *Let $E = (E_k)_{k\in\kappa}$ be a finite family of permutations on G (indexed by a key k from the finite key space κ). A* PRP-distinguisher *against E is an algorithm $\mathcal{A}$ that interacts with an oracle $\mathcal{O}\colon G \to G$ and outputs a bit $b \in \{0, 1\}$.*

The *CPA-security game* works as follows:

1. With probability $\frac{1}{2}$, the oracle $\mathcal{O}$ is instantiated with E_k for a uniformly random choice $k \in \kappa$. With probability $\frac{1}{2}$, the oracle $\mathcal{O}$ is instantiated with a permutation $P\colon G \to G$ chosen uniformly at random from the set of all permutations on G, denoted $\mathrm{Perm}(G)$.

2. $\mathcal{A}$ runs with oracle access to $\mathcal{O}$ and outputs $b \in \{0, 1\}$.

3. $\mathcal{A}$ wins the security game if
$$b = \begin{cases} 1 & \text{if } \mathcal{O} = E_k \\ 0 & \text{if } \mathcal{O} = P \end{cases}.$$

We write $\mathcal{A}^{\mathcal{O}} = b$ for indicating the event that $\mathcal{A}$ interacts with $\mathcal{O}$ and outputs b.

Definition 4 (Advantage)**.** *Let $E = (E_k)_{k\in\kappa}$ be a finite family of permutations on G. The* advantage *of the PRP-distinguisher $\mathcal{A}$ against E is defined as*

$$\mathrm{Adv}_{\mathcal{A}} \coloneqq \left|\Pr[\mathcal{A}^{E_k} = 1] - \Pr[\mathcal{A}^{P} = 1]\right|,$$

where k is chosen uniformly at random from κ and P is chosen uniformly at random from $\mathrm{Perm}(G)$.

The *PRP-security* of a block cipher is then specified by an upper bound on the advantage over all PRP-distinguishers $\mathcal{A}$ against it, where $\mathcal{A}$ is only allowed a limited amount of computational resources (like number of computation steps, number of oracle queries, or memory).

In the commutative cryptanalysis framework, we focus on one special kind of PRP-distinguisher.

Definition 5 (Commutative Distinguisher)**.** *Let $A, B\colon G \mapsto G$. A* commutative distinguisher *is a PRP-distinguisher $\mathcal{C}(A, B)$ that operates the following way:*

1. $\mathcal{C}(A, B)$ encrypts x and $A(x)$ for a uniformly random choice of $x \in G$.

2. $\mathcal{C}(A, B)$ returns 1 if $\mathcal{O}(A(x)) = B(\mathcal{O}(x))$ and 0 otherwise.

In other words, a commutative distinguisher test for a uniformly random choice of plaintext, whether A commutes with B over $\mathcal{O}$ for this particular input. We stress that for a simplified analysis, a commutative distinguisher is only allowed one choice for $x \in G$, i.e., only two queries to $\mathcal{O}$. The benefit is that we can ignore the data complexity of the distinguisher in the analysis and that we obtain a simple expression of the advantage, as discussed in the next section.

3 On the Advantage of a Commutative Distinguisher

For a permutation P over G and $A, B\colon G \mapsto G$, we denote by $\Pr[A \overset{P}{\to} B]$ the probability that $P(A(x)) = B(P(x))$ over uniform random choices of $x \in G$, i.e.,

$$\Pr[A \overset{P}{\to} B] = \Pr_{x \in G}[P \circ A(x) = B \circ P(x)] = \frac{\Gamma_P(A, B)}{|G|}.$$

For a finite family $E = (E_k)_{k \in \kappa}$ of permutations over G, the *expected commutative probability (ECP)* is defined as

$$\mathrm{ECP}[A \overset{E}{\to} B] \coloneqq \frac{1}{|\kappa|} \sum_{k \in \kappa} \Pr[A \overset{E_k}{\to} B].$$

For a commutative distinguisher $\mathcal{C}(A, B)$, we then have $\Pr[\mathcal{C}(A, B)^P = 1] = \Pr[A \overset{P}{\to} B]$, so that

$$\mathrm{Adv}_{\mathcal{C}(A,B)} = |\mathrm{ECP}[A \overset{E}{\to} B] - \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B \circ P(x)]|.$$

The term $\Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B \circ P(x)]$ can be given explicitly based on the fixed points of A and B, as we show in the following lemma.

Lemma 1. *Let $A, B\colon G \to G$. Then,*

$$\Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B \circ P(x)] = \frac{|G| - |\mathrm{Fix}(A)| - |\mathrm{Fix}(B)| + |\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)|}{|G| \cdot (|G| - 1)}.$$

Proof. First, we note that

$$\Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B \circ P(x)] = \frac{1}{|G|} \cdot \sum_{y \in G} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B(y) \mid P(x) = y].$$

If we consider the restriction of the sum to the fixed points of B, we get

$$\begin{aligned} &\sum_{y \in \mathrm{Fix}(B)} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B(y) \mid P(x) = y] \\ = &\sum_{y \in \mathrm{Fix}(B)} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = y \mid P(x) = y] = |\mathrm{Fix}(B)| \cdot \Pr_{x \in G}[A(x) = x]. \end{aligned}$$

For the remaining part we get

$$\begin{aligned} &\sum_{y \in G \setminus \mathrm{Fix}(B)} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B(y) \mid P(x) = y] \\ = &\sum_{y \in G \setminus \mathrm{Fix}(B)} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B(y) \neq P(x) \mid P(x) = y] \\ = &\Pr_{x \in G}[A(x) \neq x] \cdot \sum_{y \in G \setminus \mathrm{Fix}(B)} \Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B(y) \mid P(x) = y \neq B(y), A(x) \neq x] \\ = &\Pr_{x \in G}[A(x) \neq x] \cdot |G \setminus \mathrm{Fix}(B)| \cdot \frac{1}{|G| - 1}, \end{aligned}$$

where the last step comes from the fact that $A(x) \neq x$ and $P(x) = y$, which means that $P \circ A(x)$ is drawn uniformly at random from $G \setminus \{y\}$. In total, this means that

$$\begin{aligned}
&\Pr_{P \in \mathrm{Perm}(G), x \in G}[P \circ A(x) = B \circ P(x)] \\
=&\frac{1}{|G|^2 \cdot (|G|-1)} \cdot ((|G|-1) \cdot |\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)| + (|G| - |\mathrm{Fix}(A)|) \cdot (|G| - |\mathrm{Fix}(B)|)) \\
=&\frac{1}{|G| \cdot (|G|-1)} \cdot (|G| - |\mathrm{Fix}(A)| - |\mathrm{Fix}(B)| + |\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)|).
\end{aligned}$$

□

In the special case of $G = \mathbb{F}_p^n$, we then get for the distinguishing advantage

$$\begin{aligned}
\mathrm{Adv}_{\mathcal{C}(A,B)} &= \left| \mathrm{ECP}[A \xrightarrow{E} B] - \frac{p^n - |\mathrm{Fix}(A)| - |\mathrm{Fix}(B)| + |\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)|}{p^n \cdot (p^n - 1)} \right| \\
&= \frac{1}{p^n} \cdot \left| \frac{1}{|\kappa|} \sum_{k \in \kappa} \Gamma_{E_k}(A,B) - \frac{p^n - |\mathrm{Fix}(A)| - |\mathrm{Fix}(B)| + |\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)|}{p^n - 1} \right|.
\end{aligned}$$

Because the distinguishing advantage of a commutative distinguisher $\mathcal{C}(A,B)$ depends on the cardinality of the sets of fixed points of A and B, the notion of affine uniformity as defined in [4] is only meaningful if we take the maximum restricted to sets $\mathcal{A} \subseteq \mathrm{AGL}(n, \mathbb{F}_p)^2$ such that for all $(A,B),(C,D) \in \mathcal{A}$, we have $|\mathrm{Fix}(A)| = |\mathrm{Fix}(C)|$, and $|\mathrm{Fix}(B)| = |\mathrm{Fix}(D)|$. This is consistent with the notions of differential uniformity and c-differential uniformity (except for the cases where $\alpha = 0$ or $\beta = 0$, which have to be analyzed separately).

3.1 Commutative Distinguishers over Iterated Permutations

In the following, we consider the permutation $F \colon G \to G$, where $F = F_3 \circ F_2 \circ F_1$ for permutations $F_1, F_2, F_3 \colon G \to G$. Let $A, B \colon G \mapsto G$. We now study how the the quantity $\Gamma_F(A,B)$ can be expressed by means of commutative trails. For $C_1, C_2 \colon G \to G$, we define

$$\begin{aligned}
\Gamma_{F_1,F_2,F_3}(A, C_1, C_2, B) \coloneqq |\{x \in G \,|& F_1 \circ A(x) = C_1 \circ F_1(x), \\
& F_2 \circ F_1 \circ A(x) = C_2 \circ F_2 \circ F_1(x), \\
& F \circ A(x) = B \circ F(x)\}|
\end{aligned}$$

as the number of inputs x following the *commutative trail* $A \to C_1 \to C_2 \to B$. Since for any $P \colon G \to G$, we have $G = \bigcup_{\gamma \in G} \{x \mid P \circ A(x) = P(x) + \gamma\}$, we obtain that $\Gamma_F(A,B)$ can be expressed as a sum over $\Gamma_{F_1,F_2,F_3}(A, T_\gamma, T_\delta, B)$ for all translations T_γ, T_δ, i.e.,

$$\Gamma_F(A,B) = \sum_{\gamma \in G} \sum_{\delta \in G} \Gamma_{F_1,F_2,F_3}(A, T_\gamma, T_\delta, B).$$

If we consider the keyed permutation $F^{(k_1,k_2)} \colon G \to G$, where $F^{(k_1,k_2)} = F_3 \circ T_{k_2} \circ F_2 \circ T_{k_1} \circ F_1$, see Figure 1, we can generalize the well-known trail formula from [29] on the expected differential probability over iterated ciphers (with independent round keys) as follows.

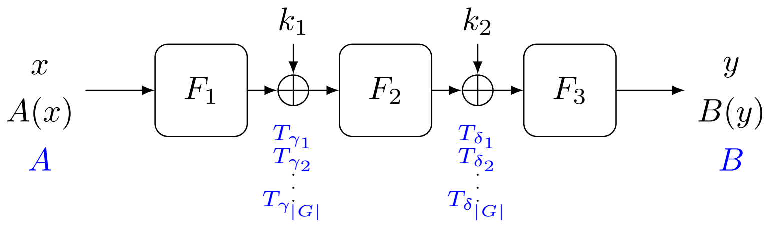

Figure 1: A commutative distinguisher over an iterated cipher containing commutative trails.

Proposition 1 (Trail Formula)**.** *Let $F = (F^k)_{k\in G\times G}$ be the family of permutations defined by $F^{(k_1,k_2)} = F_3 \circ T_{k_2} \circ F_2 \circ T_{k_1} \circ F_1$ for permutations $F_1, F_2, F_3 \colon G \to G$ and let $A, B \colon G \to G$. We have*

$$\sum_{k\in G\times G} \Gamma_{F^k}(A, B) = \sum_{\gamma\in G}\sum_{\delta\in G} \Gamma_{F_1}(A, T_\gamma) \cdot \Gamma_{F_2}(T_\gamma, T_\delta) \cdot \Gamma_{F_3}(T_\delta, B),$$

or equivalently,

$$\mathrm{ECP}[A \xrightarrow{F} B] = \frac{1}{|G|^2} \sum_{k\in G\times G} \Pr[A \xrightarrow{F^k} B] = \sum_{\gamma\in G}\sum_{\delta\in G} \Pr[A \xrightarrow{F_1} T_\gamma] \cdot \Pr[T_\gamma \xrightarrow{F_2} T_\delta] \cdot \Pr[T_\delta \xrightarrow{F_3} B].$$

Proof. This is an immediate consequence of the fact that

$$\sum_{(k_1,k_2)\in G\times G} \Gamma_{T_{k_1}\circ F_1, T_{k_2}\circ F_2, F_3}(A, T_\gamma, T_\delta, B) = \Gamma_{F_1}(A, T_\gamma) \cdot \Gamma_{F_2}(T_\gamma, T_\delta) \cdot \Gamma_{F_3}(T_\delta, B),$$

for any $\gamma, \delta \in G$. □

For the case of $F_1 = F_3 = \mathrm{id}_G$ we get the following corollary, which gives an upper bound on the advantage of a commutative distinguisher against an Even-Mansour construction.

Corollary 1. *Let $F = (F^k)_{k\in G\times G}$ be the family of permutations defined by $F^{(k_1,k_2)} = T_{k_2} \circ R \circ T_{k_1}$ for a permutation $R \colon G \to G$ and let $\mathcal{C}(A, B)$ be a commutative distinguisher against F. Then,*

$$\mathrm{Adv}_{\mathcal{C}(A,B)} \le \max\left\{\frac{\Delta_R}{|G|} - \frac{1}{|G|-1}, \frac{1}{|G|-1}\right\}.$$

Moreover, if one of $A - \mathrm{id}_G$ or $B - \mathrm{id}_G$ is a permutation, we have $\mathrm{Adv}_{\mathcal{C}(A,B)} = 0$.

Proof. Applying Proposition 1 to the case where $F_1 = F_3 = \mathrm{id}_G$ and $F_2 = R$ yields

$$\mathrm{ECP}[A \xrightarrow{F} B] = \sum_{\gamma\in G}\sum_{\delta\in G} \Pr_{x\in G}[(A - \mathrm{id}_G)(x) = \gamma] \cdot \Pr_{x\in G}[(B - \mathrm{id}_G)(x) = \delta] \cdot \Pr[T_\gamma \xrightarrow{R} T_\delta]$$

$$= \frac{|\mathrm{Fix}(A)| \cdot |\mathrm{Fix}(B)|}{|G|^2} + \sum_{\gamma\in G\setminus\{0\}} \sum_{\delta\in G\setminus\{0\}} \Pr_{x\in G}[(A - \mathrm{id}_G)(x) = \gamma] \cdot \Pr_{x\in G}[(B - \mathrm{id}_G)(x) = \delta] \cdot \Pr[T_\gamma \xrightarrow{R} T_\delta],$$

where the second equality holds because $\Pr[T_\gamma \xrightarrow{R} T_\delta] = 1$ if $\gamma = \delta = 0$ and $\Pr[T_\gamma \xrightarrow{R} T_\delta] = 0$ if exactly one of γ or δ is zero. We can bound above $\Pr[T_\gamma \xrightarrow{R} T_\delta]$ by $\frac{\Delta_R}{|G|}$, which yields

$$
\begin{aligned}
J &\coloneqq \sum_{\gamma \in G\setminus\{0\}} \sum_{\delta \in G\setminus\{0\}} \mathrm{Pr}_{x\in G}[(A-\mathrm{id}_G)(x)=\gamma]\cdot \mathrm{Pr}_{x\in G}[(B-\mathrm{id}_G)(x)=\delta]\cdot \Pr[T_\gamma \overset{R}{\to} T_\delta]\\
&\leq \frac{\Delta_R}{|G|} \sum_{\gamma \in G\setminus\{0\}} \sum_{\delta \in G\setminus\{0\}} \mathrm{Pr}_{x\in G}[(A-\mathrm{id}_G)(x)=\gamma]\cdot \mathrm{Pr}_{x\in G}[(B-\mathrm{id}_G)(x)=\delta]\\
&= \frac{\Delta_R}{|G|} \sum_{\gamma \in G\setminus\{0\}} \mathrm{Pr}_{x\in G}[(A-\mathrm{id}_G)(x)=\gamma] \sum_{\delta \in G\setminus\{0\}} \mathrm{Pr}_{x\in G}[(B-\mathrm{id}_G)(x)=\delta]\\
&= \frac{\Delta_R}{|G|}\cdot\frac{|G|-|\mathrm{Fix}(A)|}{|G|}\cdot\frac{|G|-|\mathrm{Fix}(B)|}{|G|}.
\end{aligned}
$$

Further, we have

$$
\frac{|\mathrm{Fix}(A)||\mathrm{Fix}(B)|}{|G|^2} - \frac{|G|-|\mathrm{Fix}(A)|-|\mathrm{Fix}(B)|+|\mathrm{Fix}(A)||\mathrm{Fix}(B)|}{|G|(|G|-1)} = \frac{(|G|-|\mathrm{Fix}(A)|)(|G|-|\mathrm{Fix}(B)|)}{-|G|^2(|G|-1)}.
$$

Let us define $K \coloneqq \frac{(|G|-|\mathrm{Fix}(A)|)(|G|-|\mathrm{Fix}(B)|)}{|G|^2(|G|-1)}$. We obtain for the advantage

$$
\mathrm{Adv}_{\mathcal{C}(A,B)} = |-K+J| = \begin{cases} J-K & \text{if } J \geq K\\ K-J & \text{if } J < K. \end{cases}
$$

In the first case, $\mathrm{Adv}_{\mathcal{C}(A,B)} = J-K \leq \frac{(|G|-|\mathrm{Fix}(A)|)(|G|-|\mathrm{Fix}(B)|)}{|G|^2}\cdot\left(\frac{\Delta_R}{|G|}-\frac{1}{|G|-1}\right) \leq \frac{\Delta_R}{|G|}-\frac{1}{|G|-1}$. In the second case, $\mathrm{Adv}_{\mathcal{C}(A,B)} = K-J \leq K \leq \frac{1}{|G|-1}$.

Suppose now that one of $A-\mathrm{id}_A$ or $B-\mathrm{id}_G$ is invertible. Without loss of generality, let us assume $B-\mathrm{id}_G$ is invertible. Then, $\mathrm{Pr}_{x\in G}[(B-\mathrm{id}_G)(x)=\delta]$ is equal to $1/|G|$ independently of δ and we get

$$
\begin{aligned}
\mathrm{ECP}[A \overset{F}{\to} B] &= \frac{1}{|G|}\sum_{\gamma\in G}\mathrm{Pr}_{x\in G}[(A-\mathrm{id}_G)(x)=\gamma]\sum_{\delta\in G}\Pr[T_\gamma \overset{R}{\to} T_\delta]\\
&= \frac{1}{|G|}\sum_{\gamma\in G}\mathrm{Pr}_{x\in G}[(A-\mathrm{id}_G)(x)=\gamma] = \frac{1}{|G|}.
\end{aligned}
$$

Moreover, we have $|\mathrm{Fix}(B)|=1$, so that

$$
\frac{|G|-|\mathrm{Fix}(A)|-|\mathrm{Fix}(B)|+|\mathrm{Fix}(A)||\mathrm{Fix}(B)|}{|G|(|G|-1)} = \frac{1}{|G|}
$$

and thus $\mathrm{Adv}_{\mathcal{C}(A,B)} = 0$. □

If we assume R in Corollary 1 to be a keyed permutation as well, we can take F as being a key-alternating block cipher (with independent round keys). We then obtain the following bounds on the advantage.

Corollary 2. *Let $\kappa = G \times \kappa_m \times G$ and let $E = (E_k)_{k\in\kappa}$ be a finite family of permutations defined by $E_{(k_1,k_m,k_2)} = T_{k_2} \circ E_m^{k_m} \circ T_{k_1}$ for a finite family of permutations $(E_m^{k_m})_{k_m\in\kappa_m}$. Let $\mathcal{C}(A,B)$ be a commutative distinguisher against E. Then,*

$$\mathrm{Adv}_{\mathcal{C}(A,B)} \leq \max\left\{ \max_{\gamma,\delta\in\mathbb{F}_p^n,\gamma\neq 0} \mathrm{ECP}[T_\gamma \overset{E_m}{\to} T_\delta] - \frac{1}{|G|-1}, \frac{1}{|G|-1} \right\}.$$

Moreover, if one of $A - \mathrm{id}_G$ or $B - \mathrm{id}_G$ is a permutation, we have $\mathrm{Adv}_{\mathcal{C}(A,B)} = 0$*.*

Proof. The proof is similar to the proof of Corollary 1, the only difference is that we express $\mathrm{ECP}[A \overset{F}{\to} B]$ as

$$\frac{1}{|\kappa_m|} \sum_{k_m\in\kappa_m} \sum_{\gamma\in G} \sum_{\delta\in G} \mathrm{Pr}_{x\in G}[(A - \mathrm{id}_G)(x) = \gamma] \cdot \mathrm{Pr}_{x\in G}[(B - \mathrm{id}_G)(x) = \delta] \cdot \mathrm{Pr}[T_\gamma \overset{E_m^{k_m}}{\to} T_\delta]$$

$$= \sum_{\gamma\in G} \sum_{\delta\in G} \mathrm{Pr}_{x\in G}[(A - \mathrm{id}_G)(x) = \gamma] \cdot \mathrm{Pr}_{x\in G}[(B - \mathrm{id}_G)(x) = \delta] \cdot \frac{1}{|\kappa_m|} \sum_{k_m\in\kappa_m} \mathrm{Pr}[T_\gamma \overset{E_m^{k_m}}{\to} T_\delta],$$

and for $\gamma \neq 0, \delta \neq 0$, we have $\frac{1}{|\kappa_m|}\sum_{k_m\in\kappa_m} \mathrm{Pr}[T_\gamma \overset{E_m^{k_m}}{\to} T_\delta] \leq \max_{\gamma,\delta\in\mathbb{F}_p^n,\gamma\neq 0} \mathrm{ECP}[T_\gamma \overset{E_m}{\to} T_\delta]$. □

Those bounds show that the security of a key-alternating block cipher (on average over all round keys) against commutative attacks is the same as the security against differential attacks. This fact is caused by the addition of the whitening keys and is not surprising, as a similar result has been shown in the context of t-wise independence [31, Lemma 2] (the proof is found in the full version [32]). However, that result uses the probability distribution of all differentials with fixed input difference – a quantity that is typically infeasible to compute. In contrast, the adversary we consider makes use of one differential only – an assumption which is in line with most differential attacks in current literature.

What we showed further is that, using c-differentials with $c \neq 1$, a block cipher cannot be distinguished from a random permutation at all (because the advantage would be zero since $T_\beta \circ m_c - I$ has full rank).

To summarize, a commutative cryptanalysis attack more general than a differential attack only makes sense when considering the *fixed-key* or *weak-key* model.

4 Relations Between $\Gamma_S(A, B)$ and Δ_S

As outlined in [4], for the case of $G = \mathbb{F}_2^n$ and $A, B \in \mathrm{AGL}(n, \mathbb{F}_2)$, there exist examples of commutative attacks over SPNs in the weak-key model, i.e., if the round keys of the cipher fulfill certain properties. Our goal now is to study more generally what a high value of $\Gamma_S(A, B)$, where A, B allow many weak keys, means for the differential uniformity of S.

4.1 On the Number of Weak Keys

We first analyze the commutative property over a key addition, as already studied for the case of $G = \mathbb{F}_2^n$ in [4, Section 4.1]. Here, we put a slightly different focus and consider the more general case of $G = \mathbb{F}_p^n$ for a prime p.

Let $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$. For cryptanalytic attacks, we would like that $\Gamma_{T_k}(A, B)$ is large for as many keys (or constants) k as possible, so that we can build an iterative commutative trail for many keys. Let us denote $A = L_A + c_A$ and $B = L_B + c_B$, where L_A and L_B are linear maps and c_A, c_B constants. Note that we have

$$\begin{aligned} & T_k \circ A(x) = B \circ T_k(x) \\ \Leftrightarrow \quad & (L_A - L_B)(x) = (L_B - I)(k) + c_B - c_A. \end{aligned} \tag{1}$$

In the case where $L_A = L_B = I$, i.e., differential cryptanalysis, this is equivalent to $0 = c_B - c_A$, i.e.,

$$\Gamma_{T_k}(A, B) = \begin{cases} p^n & \text{if } c_A = c_B \\ 0 & \text{else} \end{cases},$$

independently of the key k. Hence, for the choice of $c_A = c_B$, the transition over the key addition holds with probability 1 for all keys k. This is the best situation from an attacker's point of view.

In the case where $L_A = I$ and L_B corresponds to multiplication with an element $c \in \mathbb{F}_{p^n} \setminus \{0, 1\}$, i.e., c-differentials, this corresponds to $(1 - c) \cdot x = (c - 1) \cdot k + c_B - c_A$, i.e., $x = \frac{c-1}{1-c} \cdot k + \frac{c_B - c_A}{1-c}$. Hence, we have $\Gamma_{T_k}(A, B) = 1$ for each key k, so c-differentials are on the opposite side of the spectrum and cannot be used to construct exploitable iterative trails over the key addition.

Intuitively, what we require in order to build a potential attack is that Equation 1 has a high number of solutions $(x, k) \in (\mathbb{F}_p^n)^2$, i.e., the matrix

$$M_{A,B} \coloneqq [L_A - L_B \mid I - L_B]$$

is of low rank. Indeed, if $L_A - L_B$ is of low rank, we obtain that $\Gamma_{T_k}(A, B)$ is large for some suitable k (and appropriate c_A, c_B). If $L_B - I$ is of low rank, then we can control $\Gamma_{T_k}(A, B)$ for many keys k.

The only case in which $M_{A,B}$ is of rank 0 is if $L_A = L_B = I$, i.e., the case of differential attacks. In case of c-differentials, we have $\mathrm{rank}(M_{A,B}) = n$, which is maximally unfortunate from the attacker's point of view.

Now, we restrict to the case where $\Gamma_{T_k}(A, B) = p^n$, i.e., if k is a weak key, we want that the commutative property holds over the key addition with probability one. From Equation 1, we obtain that there exists k such that $\Gamma_{T_k}(A, B) = p^n$ if and only if $L_A = L_B$. In that case, k is a weak key if and only if $(L_B - I)(k) = c_A - c_B$, which implies that the number of weak keys equals p^{n-d_B}, where $d_B \coloneqq \mathrm{rank}(L_B - I)$. Hence, to maximize the number of weak keys, we want d_B to be low. This observation is precisely what was already shown in [4, Corollary 1] for the case of $p = 2$.

4.2 Bounds on $\Gamma_S(A, B)$ with Respect to Δ_S

Suppose we are given a function $S\colon G \to G$ (e.g., an S-box, a cryptographic permutation, or a fixed-key instance of a block cipher) and mappings $A, B\colon G \to G$, one can deduce an upper bound on the quantity $\Gamma_S(A, B)$ based on the differential uniformity of S.

Proposition 2. *Let $S, A, B\colon G \to G$. Then,*

$$\Gamma_S(A, B) \leq \begin{cases} |\mathrm{Im}(A - \mathrm{id}_G)| \cdot |\mathrm{Im}(B - \mathrm{id}_G)| \cdot \Delta_S & \textit{if } \mathrm{Fix}(A) = \emptyset \\ (|\mathrm{Im}(A - \mathrm{id}_G)| - 1) \cdot |\mathrm{Im}(B - \mathrm{id}_G)| \cdot \Delta_S + |\mathrm{Fix}(A)| & \textit{else} \end{cases}.$$

Proof. Let us denote by A' the mapping $A - \mathrm{id}_G$ and by B' the mapping $B - \mathrm{id}_G$. Further, for $a, b \in G$, we define the sets $\mu_a \coloneqq \{x \in G \mid A'(x) = a\}$ and $\nu_b \coloneqq \{x \in G \mid B'(S(x)) = b\}$. We then have

$$\begin{aligned}
&\{x \in G \mid S(A(x)) = B(S(x))\} = \{x \in G \mid S(A'(x) + x) = B'(S(x)) + S(x)\} \\
&= \bigcup_{a \in \mathrm{Im}(A')} \bigcup_{b \in \mathrm{Im}(B')} \{x \in \mu_a \cap \nu_b \mid S(A'(x) + x) = B'(S(x)) + S(x)\} \\
&= \bigcup_{a \in \mathrm{Im}(A')} \bigcup_{b \in \mathrm{Im}(B')} \{x \in \mu_a \cap \nu_b \mid S(x + a) = S(x) + b\},
\end{aligned}$$

hence

$$\begin{aligned}
\Gamma_S(A, B) &= |\bigcup_{a \in \mathrm{Im}(A')} \bigcup_{b \in \mathrm{Im}(B')} \{x \in \mu_a \cap \nu_b \mid S(x + a) = S(x) + b\}| \\
&= \sum_{a \in \mathrm{Im}(A')} \sum_{b \in \mathrm{Im}(B')} |\{x \in \mu_a \cap \nu_b \mid S(x + a) = S(x) + b\}|.
\end{aligned}$$

In case that $0 \notin \mathrm{Im}(A')$, i.e., $\mathrm{Fix}(A) = \emptyset$, we immediately get $\Gamma_S(A, B) \leq |\mathrm{Im}(A')| \cdot |\mathrm{Im}(B')| \cdot \Delta_S$. Otherwise, we get

$$\begin{aligned}
\Gamma_S(A, B) &\leq (|\mathrm{Im}(A')| - 1) \cdot |\mathrm{Im}(B')| \cdot \Delta_S + \sum_{b \in \mathrm{Im}(B')} |\{x \in \mu_0 \cap \nu_b \mid 0 = b\}| \\
&= (|\mathrm{Im}(A')| - 1) \cdot |\mathrm{Im}(B')| \cdot \Delta_S + |\mu_0 \cap \nu_0|,
\end{aligned}$$

and the result follows since $|\mu_0 \cap \nu_0| \leq |\mathrm{Fix}(A)|$. □

For the case of G being a finite vector space and A, B affine mappings, we get the following corollary by applying the rank-nullity theorem.

Corollary 3. *Let p be a prime and let $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ and $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$, where $A = L_A + c_A$, $B = L_B + c_B$ with L_A, L_B linear. Let d_A and d_B denote the rank of $L_A - I$ and $L_B - I$, respectively. We then have*

$$\Gamma_S(A, B) \leq \begin{cases} p^{d_A + d_B} \cdot \Delta_S & \text{if } c_A \notin \mathrm{Im}(I - L_A) \\ (p^{d_A} - 1) \cdot p^{d_B} \cdot \Delta_S + p^{n - d_A} & \text{else} \end{cases}. \tag{2}$$

Suppose $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ has a non-trivial affine self-equivalence $S \circ A = B \circ S$, plugging $\Gamma_S(A, B) = p^n$ into Equation (2) yields $p^{n - (d_A + d_B)} \leq \Delta_S$ in both cases. If we want to allow many weak keys, d_A and d_B should be rather small, as discussed in Section 4.1. Hence, the differential uniformity of such S would be relatively high (unless n is very small).

Remark 1. Suppose $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ has a non-trivial affine self-equivalence $S \circ A = B \circ S$. For $\Delta_S \leq p$ we obtain $d_A + d_B \geq n - 1$. In [5], it was shown that if $S\colon \mathbb{F}_2^8 \to \mathbb{F}_2^8$ is an APN permutation with a non-trivial linear self equivalence $S \circ L_A = L_B \circ S$, then there are essentially only two possible classes of (L_A, L_B) to consider ([5, Theorem 4]). With our argument, we can exclude the second class from consideration, as $d_A = d_B = 3$, which is a contradiction to $d_A + d_B \geq n - 1 = 7$.

Example 1. If we identify $\mathbb{F}_p^n$ by the finite field $\mathbb{F}_{p^n}$ and take L as the mapping $x \mapsto x^{p^i}$, where i divides n, we get $\ker(L - I) = \mathbb{F}_{p^i}$, hence $\operatorname{rank}(L - I) = n - i$. If $S \colon \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ is such that its interpolating polynomial in $\mathbb{F}_{p^n}[X]$ only has coefficients in the subfield $\mathbb{F}_{p^i}$, we have $\Gamma_S(L, L) = p^n$. There are examples of such S with $\Delta_S \leq p$ (e.g., $x \mapsto x^2$ fulfills $\Delta_S = 1$ for p odd, for $p = 2$, the mapping $x \mapsto x^3$ fulfills $\Delta_S = 2$). If n is even, we can take $i = n/2$, in which case $\operatorname{rank}(L - I) + \operatorname{rank}(L - I) = n > n - 1$.

4.2.1 The Differential-Affine Case

The case where only one of A or B is a translation and the other an arbitrary affine bijection was discussed in [7, Example 1]. We now want to study this special case in more detail.

Let $A = T_\alpha$ for $\alpha \neq 0$ and $B = L_B + c_B \in \mathrm{AGL}(n, \mathbb{F}_p)$ with L_B linear and $d_B = \operatorname{rank}(L_B - I)$. For $S \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$, Equation (2) yields that $\Gamma_S(A, B) \leq p^{d_B} \cdot \Delta_S$. In the case that (A, B) defines a self-equivalence of S, i.e., $\Gamma_S(A, B) = p^n$, we obtain $\Delta_S \geq p^{n-d_B}$. In case S is a permutation, this lower bound on the differential uniformity of S can be improved, as we show in the following.

First, we observe the following properties of the order of the linear parts.

Lemma 2. *Let $A \in \mathrm{AGL}(n, \mathbb{F}_p)$, let L_A be the linear part of A, as well as $c_A = A(0)$. Then $A^{\mathrm{ord}(L_A)} = T_c$ for $c := \sum_{i=0}^{\mathrm{ord}(L_A)-1} L_A^i(c_A)$ and*

$$\mathrm{ord}(A) = \begin{cases} \mathrm{ord}(L_A) & \text{if } c = 0 \\ p \cdot \mathrm{ord}(L_A) & \text{otherwise.} \end{cases}$$

Proof. The proof follows from noting that

$$\begin{aligned} A^{\mathrm{ord}(L_A)}(x) &= L_A^{\mathrm{ord}(L_A)}(x) + \sum_{i=0}^{\mathrm{ord}(L_A)-1} L_A^i(c_A) \\ &= x + \sum_{i=0}^{\mathrm{ord}(L_A)-1} L_A^i(c_A) \end{aligned}$$

for all $x \in \mathbb{F}_p^n$ and that the order of T_c is either 1 (if $c = 0$) or p (if $c \neq 0$). □

This directly implies that if $\mathrm{ord}(A) \neq \mathrm{ord}(L_A)$ we can consider $S \circ T_c = S \circ A^{\mathrm{ord}(L_A)} = B^{\mathrm{ord}(L_A)} \circ S$, instead of $S \circ A = B \circ S$, and arrive in the differential-affine case.

Focusing on the case in which A is a translation, again, this implies that the order of the linear part L_B of B needs to divide p.

Lemma 3. *Let $S \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be bijective. Let $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$ such that $S \circ A = B \circ S$. Then, $\mathrm{ord}(A) = \mathrm{ord}(B)$. Especially, if $A = T_\alpha$ for $\alpha \in \mathbb{F}_p^n \setminus \{0\}$, and denoting by L_B the linear map $B - B(0)$, then $L_B^p = I$.*

Proof. From $S \circ T_\alpha = B \circ S$ it follows that $S \circ A \circ S^{-1} = B$ holds for all $x \in \mathbb{F}_p^n$. We then have $S \circ A^r \circ S^{-1} = B^r$, which implies $\mathrm{ord}(A) = \mathrm{ord}(B)$. If $A = T_\alpha$ then Lemma 2 implies that $\mathrm{ord}(L_B)$ is either $p = \mathrm{ord}(T_\alpha)$ or 1. In both cases, we get that $L_B^p = I$. □

We then obtain an upper bound on d_B.

Lemma 4. *Let $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be bijective. Let $\alpha \in \mathbb{F}_p^n \backslash \{0\}$ and $B \in \mathrm{AGL}(n, \mathbb{F}_p)$ with $B = L_B + c_B$ for a linear mapping L_B and $c_B \in \mathbb{F}_p^n$, such that $S \circ T_\alpha = B \circ S$. Then, $d_B = \mathrm{rank}(L_B - I) \leq \left\lfloor \frac{n(p-1)}{p} \right\rfloor$.*

Proof. By Lemma 3, we obtain $(L_B - I)^p = L_B^p - I = 0$, hence $\mathrm{rank}((L_B - I)^p) = 0$. By Sylvester's rank inequality, this yields

$$0 = \mathrm{rank}((L_B - I)^p) \geq p \cdot \mathrm{rank}(L_B - I) - (p-1)n.$$

The result follows by rearranging terms and the fact that $\mathrm{rank}(L_B - I)$ is an integer value. □

Plugging this upper bound on d_B into $\Delta_S \geq p^{n-d_B}$ yields $\Delta_S \geq p^{n-\lfloor \frac{np-n}{p} \rfloor} = p^{\lceil n - \frac{np-n}{p} \rceil} = p^{\lceil \frac{n}{p} \rceil}$. In summary, we obtain the following corollary which can be seen as a generalization of [7, Item 1 of Lemma 10].[2]

Corollary 4. *Let $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$. Let $A, B \in \mathrm{AGL}(n, \mathbb{F}_p^n)$, where $A = L_A + c_A$ and $B = L_B + c_B$ for linear mappings L_A, L_B and $c_A, c_B \in \mathbb{F}_p^n$, respectively, such that* $\mathrm{ord}(A) \neq \mathrm{ord}(L_A)$ *and $S \circ A = B \circ S$. Then, $\Delta_S \geq p^{n-\mathrm{rank}\left(L_B^{\mathrm{ord}(L_A)} - I\right)}$. If S is bijective, then $\Delta_S \geq \max\left\{p^{\lceil \frac{n}{p} \rceil}, p^{n-\mathrm{rank}\left(L_B^{\mathrm{ord}(L_A)} - I\right)}\right\}$.*

Example 2. The (bijective) 5-bit S-box $S\colon \mathbb{F}_2^5 \to \mathbb{F}_2^5$ given in [7, Example 1] is an example of an S-box for which exists $\alpha \neq 0$ and $B \in \mathrm{AGL}(n, \mathbb{F}_p)$ with $d_B = 1$ such that $\Gamma_S(T_\alpha, B) = 2^5$, but differential uniformity strictly lower than 2^5 and non-maximal linearity.

4.2.2 Tightness and Application

Let m be a positive integer and $f_1, f_2\colon \mathbb{F}_p^m \to \mathbb{F}_p^m$. Let $n \coloneqq 2m$. We study the permutation on $\mathbb{F}_p^n$ defined as $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n, (l, r) \mapsto (l', r')$ with $l' = \ell + f_1(r)$, $r' = r + f_2(l')$. Such a form of S is known as the *2-round Feistel construction*, see Figure 2. As we explain now, by suitable choices for the functions f_1 and f_2, we can guarantee the existence of $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$ with A being a translation such that $\Gamma_S(A, B)$ takes the maximal value p^n on the one hand, but fulfilling $\Delta_S < p^n$ on the other hand.

Proposition 3. *Let $f_1, f_2\colon \mathbb{F}_p^m \to \mathbb{F}_p^m$ and $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be the 2-round Feistel construction defined by f_1 and f_2 (with $n = 2m$). Let $\alpha \in \mathbb{F}_p^m$. If f_2 has algebraic degree 2, then the mapping $B_\alpha\colon \mathbb{F}_p^n \to \mathbb{F}_p^n, (x, y) \mapsto (x + \alpha, y + f_2(x + \alpha) - f_2(x))$ is in $\mathrm{AGL}(n, \mathbb{F}_p)$ and we have*

1. *$S \circ T_{(\alpha,0)} = B_\alpha \circ S$, i.e., $\Gamma_S(T_{(\alpha,0)}, B_\alpha) = p^n$.*
2. *$\Delta_S = p^m \cdot \max\{\Delta_{f_1}, \Delta_{f_2}\}$.*

Proof. The fact that B_α is affine follows because $x \mapsto f_2(x + \alpha) - f_2(x)$ is affine if f_2 is of algebraic degree 2. Because B_α is a permutation, we have $B_\alpha \in \mathrm{AGL}(n, \mathbb{F}_p)$. Statement 1 then follows from a simple computation (see also Figure 2 for the propagation of the difference $(\alpha, 0)$ through the Feistel construction). Statement 2 follows since the two-round Feistel construction defined by f_1 and f_2 is CCZ-equivalent to a parallel application of f_1 and f_2. We recall that $S\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ and $T\colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ are called *CCZ-equivalent* [14] if there exists $A \in \mathrm{AGL}(2n, \mathbb{F}_p)$ such that $\{(x, T(x)) \mid x \in \mathbb{F}_p^n\} = A(\{(x, S(x)) \mid x \in \mathbb{F}_p^n\})$ and that CCZ-equivalence preserves the differential uniformity. □

[2] The remaining results of [7, Lemma 10] are easily generalizable, too, but out of scope for this work.

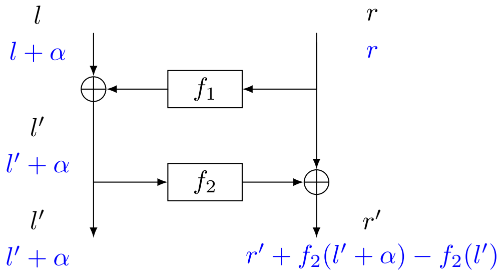

Figure 2: A 2-round Feistel construction.

Example 3. Let $p=2$ and identify $\mathbb{F}_2^m$ by $\mathbb{F}_{2^m}$. For f_1 and f_2 take the APN function $x \mapsto x^3$ over $\mathbb{F}_{2^m}$. Then, $\operatorname{rank}(L_{B_\alpha}-I)=\operatorname{rank}(x \mapsto \alpha x^2+\alpha^2 x)=m-1$ and $\Delta_S=2^{m+1}=2^{n-\operatorname{rank}(L_{B_\alpha}-I)}$.

Example 4. Let p be odd and identify $\mathbb{F}_p^m$ by $\mathbb{F}_{p^m}$. For f_1 and f_2 take the planar function $x \mapsto x^2$ over $\mathbb{F}_{p^m}$ (from the planarity property, we have $\Delta_{f_1}=\Delta_{f_2}=1$). Then, $\operatorname{rank}(L_{B_\alpha}-I)=\operatorname{rank}(x \mapsto 2\alpha x)=m$ and $\Delta_S=p^m=p^{n-\operatorname{rank}(L_{B_\alpha}-I)}$.

These examples show the tightness of the bound given in Corollary 4. Instead of using APN, resp., planar functions for f_1 and f_2, one can construct S with various tradeoffs between low differential uniformity and low rank of $L_{B_\alpha}-I$.

If we allow having an arbitrary mapping $A \in \mathrm{AGL}(n,\mathbb{F}_p)$ in the input instead of a translation, we could extend the 2-round Feistel by one more round in the input, to get a *3-round Feistel construction*. This way, it is possible to obtain a permutation S with $\Gamma_S(A,B)=p^n$ for some $A,B \in \mathrm{AGL}(n,\mathbb{F}_p)$, but differential uniformity lower than 2^{m+1}, resp, p^m, as is the limit for the 2-round Feistel construction.

Several ciphers from the literature use S-boxes based on 3-round Feistel networks, and those indeed have non-trivial commutants. It is the case of the S-box of iScream [25], as pointed out in [4], as well as the ZUC stream cipher (that is part of the 3GPP standard) [22].

5 Classification of Linear and Affine Permutations Sharing the Same Cycle Type

Computing a bijective S-Box S with $\Gamma_S(A,B)=p^n$ and A,B affine permutations can be done as follows. Choose two affine permutations A,B which share the same cycle type. Then determine the cycle structure and compute the corresponding S-boxes via Algorithm 1.

One question arising in this context is how to construct all affine bijective mappings which share the same cycle type. Once this question is settled, all S-Boxes with the desired property can in principle be computed. Indeed those kind of affine mappings exist and its classification has been studied since the late 50's / early 60's at least. For instance Elspas [21] and Crowell [16] studied the linear case, while Wang [38] and Fripertinger [23] focused on the affine one. Proposition 9 and 10, and their proofs appear in [9, Prop. 2.1]. We give a survey about the whole theory, which has not been done before, to our best knowledge. Here, we put a focus on concrete constructions. To do so we rewrote the above two propositions and add some details to the proofs. Moreover, we

Algorithm 1 S-box Generation

```
Require: $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$ of the same cycle type
Ensure: Permutation $S$ with $S \circ A = B \circ S$
for all divisors $\ell \in \mathbb{N}$ of $\mathrm{ord}(A)$ do
    $C_\ell^{(A)} \leftarrow \{x \in \mathbb{F}_p^n \mid \mathrm{ord}_A(x) = \ell\}$
    $C_\ell^{(B)} \leftarrow \{x \in \mathbb{F}_p^n \mid \mathrm{ord}_B(x) = \ell\}$
end for
initialise empty S-box $S$
for all divisors $\ell \in \mathbb{N}$ of $\mathrm{ord}(A)$ do
    $X \leftarrow C_\ell^{(A)}$                          ▷ $X$: inputs not used yet
    $Y \leftarrow C_\ell^{(B)}$                          ▷ $Y$: outputs not used yet
    while $X \neq \emptyset$ do
        $x \leftarrow_\$ X$                  ▷ sample (not necessarily) random element
        $y \leftarrow_\$ Y$                              ▷ sample random element
        for $j \in \{0, 1, \ldots, \ell - 1\}$ do
            $X \leftarrow X \setminus \{A^j(x)\}$
            $Y \leftarrow Y \setminus \{B^j(y)\}$
            $S(A^j(x)) \leftarrow B^j(y)$                ▷ define S-box
        end for
    end while
end for
```

apply these results to generate A, B with the desired property. We first recall some definitions and results on matrices and polynomials.

Definition 6. *Let $P(X) \in \mathbb{F}_p[X]$ be a nonzero polynomial. If $P(0) \neq 0$, then the least positive integer e for which $P(X)$ divides $X^e - 1$ is called the* order *of $P(X)$ and denoted by* $\mathrm{ord}(P(X))$. *If $P(0) = 0$ then $P(X) = X^h G(X), G(0) \neq 0$, where $h \in \mathbb{N}$ and $G(X)$ are uniquely determined;* $\mathrm{ord}(P(X))$ *is then defined to be* $\mathrm{ord}(G(X))$.

Remark 2. If $P(X)$ is an irreducible polynomial of order e then $P(X)^d$ has order $p^t e$, where t is smallest integer s.t. $p^t \geq d$ (see e.g. [30], Theorem 3.8., p. 86). Hence if two irreducible polynomials have the same order then also any power of them. In the sequel we will denote with e instead of $\mathrm{ord}(P(X))$ if $P(X)$ is clear from the context for ease of notation.

Definition 7. *Given $A \in GL(n, \mathbb{F}_p)$. The* characteristic polynomial $\chi_A(X)$ *is defined as* $\det(XI - A)$. *The minimal polynomial $m_A(X) = X^d + m_{d-1}X^{d-1} + \cdots + m_0$ is the polynomial of lowest positive degree with the property that $m_A(A) = 0$.*

Definition 8. *The matrix*

$$\begin{pmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

is called the companion matrix *of the polynomial $P(X) = X^n + \sum_{i=0}^{n-1} a_i X^i$.*

Proofs for the following theorem and proposition can be found in e.g. [24, pp. 190-192].

Theorem 1. *Given $A \in GL(n, \mathbb{F}_p)$ with characteristic polynomial $\chi_A(X)$. If $\chi_A(X)$ decomposes into the product $\prod_{i=1}^{r'} P_i(X)^{e_i}$ of prime factors $P_i(X)$ then A is similar to*

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots \ddots & \vdots & \\ 0 & 0 & \dots & A_r \end{pmatrix}.$$

Thereby A_j is the companion matrix of a $P_i^{t_j}(X), t_j \leq e_i$. Moreover for each $P_i(X)$ the powers t_j sum up to e_i for the corresponding companion matrices A_j of $P_i(X)^{t_j}$. This matrix is unique for A apart from the order of the blocks and called Weierstraß normal form.

Proposition 4. *Given $A \in GL(n, \mathbb{F}_p)$ where $\chi_A(X)$ is equal to the minimal polynomial $m_A(X)$. Then there exists a vector v s.t. $A^0 v, \dots, A^{n-1} v$ forms a basis of $\mathbb{F}_p^n$. Such a vector is called* cyclic. *If A is a companion matrix of $P(X)$ with $P(0) \neq 0$ then e_1 is cyclic.*

A direct consequence of Theorem 1 is the following.

Corollary 5. *Given $A \in GL(n, \mathbb{F}_p)$ where $\chi_A(X) = \prod_{i=1}^{r} P_i(X)^{e_i} = m_A(X)$. Then we have the following isomorphisms of algebras:*

$$\mathbb{F}_p[A] := \left\{ \sum_{i=0}^{l} A^i | l \in \mathbb{N} \right\} \cong \mathbb{F}_p[X]/(\chi_A(X)) \cong \prod_{i=1}^{m} \mathbb{F}_p[X]/(P_i(X)^{e_i}).$$

Remark 3. Note, that $\mathbb{F}_p[X]/(P(X)^e)$ is a local ring. Indeed its ideals are

$$0 = (P(X)^e) \subset \left(P(X)^{e-1}\right) \subset \dots \subset (P(X))$$

and the latter is maximal. Hence the invertible elements $\mathbb{F}_p[X]/(P(X)^e)^*$ are exactly those not lying in $(P(X))$ and any element $U(X) \in \mathbb{F}_p[X]/(P(X)^e)$ has a representation of the form $E(X)P(X)^t, 1 \leq t \leq e$ with $E(X) \in \mathbb{F}_p[X]/(P(X)^e)^*$.

Proposition 5. *Given $A \in \mathrm{GL}(n, \mathbb{F}_p)$ with $\chi_A(X) = m_A(X)$. The linear map $\psi_{A,v} : \mathbb{F}_p[X]/(\chi_A(X)) \to \mathbb{F}_p^n, P(X) \mapsto P(A)v$ for a cyclic vector v is a bijection.*

Proof. By Proposition 4 and Corollary 5 the mapping $\varphi_{A,v} : \mathbb{F}_p[A] \to \mathbb{F}_p^n, \quad P(A) \mapsto P(A)v$ is a bijection and $\mathbb{F}_p[A]$ isomorphic to $\mathbb{F}_p[X]/(\chi_A(X))$. □

Definition 9. *Given an affine mapping $A = T_a \circ L_A \in \mathrm{AGL}(n, \mathbb{F}_p)$ with L_A linear.*

1. *For polynomials $P(X), U(X) \in \mathbb{F}_p[X]$ we define the mappings $\varphi_{P(X)} : \mathbb{F}_p[X]/(P(X)) \to \mathbb{F}_p[X]/(P(X))$, $Q(X) \mapsto XQ(X)$ and $\varphi_{P(X),U(X)} : \mathbb{F}_p[X]/(P(X)) \to \mathbb{F}_p[X]/(P(X)), Q(X) \mapsto XQ(X) + U(X)$.*

2. *We also set $\varphi_A : \mathbb{F}_p^n \to \mathbb{F}_p^n, v \mapsto Av = L_A v + a$.*

3. *If additionally $\chi_{L_A}(X) = m_{L_A}(X)$ then for $u \in \mathbb{F}_p^n$ and a cyclic vector v of L_A, we have $u = \left(\sum_{i=0}^{n-1} u_i L_A^i\right) v$ for uniquely determined $u_0, \ldots, u_{n-1} \in \mathbb{F}_p$. We denote with $U_{L_A}(X)$ the polynomial $\sum_{i=0}^{n-1} u_i X^i$. Vice Versa given $U(X) = \sum_{i=0}^{n-1} u_i X^i \in \mathbb{F}_p[X]$ we denote with $u_{L_A,v} = \left(\sum_{i=0}^{n-1} u_i L_A^i\right) v$. If L_A and v is clear from the context we just write $U(X)$ or u respectively.*

Item 2 is introduced to ease notation in the sequel.

Corollary 6. *Given $A \in \mathrm{GL}(n, \mathbb{F}_p)$ with $\chi_A(X) = m_A(X)$. Then $\varphi_{\chi_A}(U(X)) = \psi_{A,e_1}^{-1} \circ \varphi_A \circ \psi_{A,e_1}(U(X))$.*

Definition 10. *Let $A \in GL(n, \mathbb{F}_p)$ as in Theorem 1 with $\chi_A(X) = \prod_{i=1}^{r'} P_i(X)^{e_i}$. Let $\mathbb{F}_p^n$ be decomposed as $\mathbb{F}_p^n = \bigoplus_{i=1}^{r} \mathbb{X}_i$ so that each block A_i has $\mathbb{X}_i$ as domain and codomain. Let $j \in \{1, \cdots, r\}$. We define the $P_j(X)$-*primary component *$\mathbb{Y}_j$ as the direct sum $\bigoplus_{i \in J_j} \mathbb{X}_i$ where $J_j = \{i \mid \chi_{A_i}(X) = P_j^{t_j}(X) \text{for some } t_j\}$. Let $x \in \mathbb{F}_2^n$ be decomposed as $x = (x_1, \ldots, x_r)$ with $x_i \in \mathbb{X}_i$. We similarly say that x belongs to $\mathbb{Y}_j$ if:*

$$\forall i \in \{1, \cdots r\}, x_i \neq 0 \implies \chi_{A_i}(X) = P_j^{t_j}(X).$$

Remark 4. It is well known that $\mathbb{F}_p^n = \bigoplus_{i=1}^{m} \mathbb{X}_i$.

5.1 The Linear Case

Proposition 6. *Given $A, B \in \mathrm{GL}(n, \mathbb{F}_p)$ where the characteristic polynomials of A and B have the prime factor decomposition $\chi_A(X) = \prod_{i=1}^{r} P_i(X)^{e_i}$ and $\chi_B(X) = \prod_{i=1}^{l} Q_i(X)^{h_i}$ respectively. A and B share the same cycle type if and only if $l = r$ and for every $P_i(X)$ there exists a $Q_j(X)$ with $e_i = h_j$ and $\mathrm{ord}(P_i(X)) = \mathrm{ord}(Q_j(X))$.*

Proof. Let us first consider the case that $\chi_A(X) = P(X)^e$ with $P(X)$ irreducible, $\deg(P(X)) = d$ and $\chi_A(X)$ is equal to the minimal polynomial of A. By assumption the matrix A is similar to the companion matrix of $P(X)^e$. Thus, without loss of generality, we can assume that A is already the companion matrix as the similarity relation preserves the cycle structure. From Proposition 4 we have that $A^0 e_1, \ldots, A^{n-1} e_1$ forms a basis of $\mathbb{F}_p^n$, where $n := de$. Given $u \in \mathbb{F}_p^n$, then $u = \sum_{i=0}^{n-1} u_i A^i e_1 = \left(\sum_{i=0}^{n-1} u_i A^i\right) e_1$ for uniquely determined $u_0, \ldots, u_{n-1} \in \mathbb{F}_p$. Computing the cycle of u, i.e. computing $A^0 u, \ldots, A^d u$ until $A^i u = u$ is equivalent to compute

$$A^i \left(\sum_{i=0}^{n-1} u_i A^i\right) e_1 = \left(\sum_{i=0}^{n-1} u_i A^i\right) e_1.$$

Note, that since A is bijective we do not have to deal with preperiods. Computing the cycle of u is equivalent to compute the cycle of $U(X)$ among the mapping $\varphi_{P^e(X)}$ due to Proposition 5 and Corollary 6. This boils down to considering when $X^i(U(X)) = U(X)$. If $U(X) \in \mathbb{F}[X]/(P(X)^e)^*$ then the length of its cycle is the order X which is the order of $P(X)^e$. If $U(X)$ is not invertible, then $U(X) = E(X)(P(X))^t, 1 \leq t \leq e$ with $E(X) \in \mathbb{F}[X]/(P(X)^e)^*$ by Remark 3. Thus the cycle length of $U(X)$ is equal to the order of X in $\mathbb{F}[X]/(P(X)^{e-t})$, which is the order of $P(X)^{t-e}$. Indeed since $E(X) \in \mathbb{F}_p[X]/(P(X)^e)^*$, it holds that $X^i(U(X)) = U(X)$ if and only if $(X^i - 1)P(X)^t = 0$ in $\mathbb{F}_p[X]/(P(X)^{e-t})$, i.e. $P(X)^{e-t}$ divides $(X^i - 1)$. Hence if $P(X)$ and $Q(X)$ have the same order then any matrix B similar to the companion matrix of $Q(X)^e$ will share the same cycle type with A by Remark 2.

Now consider the general case, i.e. matrices with Weierstraß normal form

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots \ddots & \vdots & \\ 0 & 0 & \dots & A_r \end{pmatrix},$$

where $\chi_A(X) = \prod_{i=1}^{r} P_i(X)^{e_i}$. Any $u \in \mathbb{F}_p^n$ can uniquely be written as $u = \sum_{i=1}^{r} u_i$, where $u_i \in \mathbb{X}_i$ according to Definition 10. Hence the length of the cycle of u is equal to $\operatorname{lcm}(l_1, \dots, l_r)$, where l_i is the length of the cycle of u_i with respect to A_i. Thus this case boils down to Case 1 as each A_i can be treated independently. This proves the statement for Case 2 and finally the proposition. □

In order to construct two $A, B \in \mathrm{GL}(n, \mathbb{F}_p)$ with the same cycle type, but *not* similar, one can start with a matrix A in Weierstraß normal form, where $m_A(X) = \chi_A(X)$. Then some or all irreducible polynomials $P_i(X)$ of $\chi_A(X)$ are exchanged by some polynomials $Q_i(X)$ with $Q_i(X) \neq P_i(X)$ and $\operatorname{ord}(P_i(X)) = \operatorname{ord}(Q_i(X))$. The corresponding Weierstraß normal form B is computed based on A by modifying some of the blocks A_j according to the change of the irreducible factors. Then, any pair $(A', B') = (SAS^{-1}, TBT^{-1}), S, T \in \mathrm{GL}(n, \mathbb{F}_p)$ is a solution.

5.2 The Affine Case

Proposition 7 ([26])**.** *Given an affine mapping $A = T_a \circ L_A \in \mathrm{AGL}(n, \mathbb{F}_p)$ with L_A linear. Let L_A be similar to*

$$\begin{pmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots \ddots & \vdots & \\ 0 & 0 & \dots & B_r \end{pmatrix},$$

where w.l.o.g. the $(X-1)$-primary component, if it exists, corresponds to the last blocks. Then there exists a $w \in \mathbb{F}_p^n$ s.t. $T_{-w} \circ A \circ T_w = T_u \circ L_A$, where u belongs to the $(X-1)-$primary component.

Proof. According to Remark 4, we decompose $\mathbb{F}_p^n$ as $\mathbb{F}_p^n = V \oplus U$, where U is the $(X-1)-$primary component or $U = \{0\}$. Thus $a = v + u, v \in V, u \in U$. It is $\chi(L_A)(X) = \prod_{i=1}^{r} P_i(X)^{e_i}(X-1)^e$ and $\chi(L_A|_V)(X) = \prod_{i=1}^{r} P_i(X)^{e_i}$. Therefore $L_A - I$ is invertible over V and there exists an element $w \in V$ with $(L_A - I)(w) = -v$. It follows that $T_{-w} \circ A \circ T_w(X) = L_A(X) + L_A(w) - w + u + v = T_u \circ L_A(X)$ as requested. □

Corollary 7 ([38])**.** *Given $A \in \mathrm{AGL}(n, \mathbb{F}_p)$, where $(X-1) \nmid \chi(L_A)$. Then the cycle structure of A equals the one of L_A.*

Proof. In this case $U = \{0\}$ and thus $u = 0$. □

Hence, to generalize Proposition 6 to the affine case it is sufficient to consider affine mappings of the form $T_u \circ L_A$, where u is an element of the $(X-1)-$ primary component or 0. The parts not affected by the $(X-1)$-primary component are covered by Proposition 6. Therefore we can restrict to consider the subcase $A = T_u \circ L_A$ where L_A is a companion matrix with minimal polynomial $(X-1)^e$. Let $A \in \mathrm{GL}(n, \mathbb{F}_p)$ a matrix with cyclic vector $e_1 \in \mathbb{F}_p^n$. To prove the linear case we used that $\varphi_{\chi_A(X)}$ and φ_A share the same cycle structure. We will generalize this fact to the affine case.

Proposition 8. *Let $P(X)$ be a polynomial of degree n, $a \in \mathbb{F}_p^n$ and $A = T_u \circ L_A$, where L_A is the companion matrix of $P(X)$. Then $\varphi_{P(X),U(X)}$ and φ_A share the same cycle type.*

Proof. By Proposition 5, ψ_{A,e_1} is bijective and thus ψ^{-1}_{A,e_1} exists. It is

$$A(\psi_{L_A,e_1}(Q(X))) = L_A(q(A)e_1) + u \tag{3}$$

and therefore $\psi^{-1}_{L_A,e_1} \circ \varphi_A \circ \psi_{L_A,e_1}(Q(X)) = XQ(X) + U(X)$. It follows that $\psi^{-1}_{A,e_1} \circ \varphi_A \circ \psi_{A,e_1} = \varphi_{P(X),U(X)}$ and eventually that φ_A and $\varphi_{P(X),U(X)}$ share the same cycle structure. □

We will now show that $L_A + u$ and $B = L_B + u' \in \mathrm{AGL}(n, \mathbb{F}_p)$ share the same cycle structure only if $m_{L_B}(X)$ is also equal to $(X-1)^e$ and u' fulfills certain conditions depending on u.

Corollary 8. *Let $u \in \mathbb{F}_p^n$. Then u belongs to a cycle of length ℓ of A if and only if $U(X)$ belongs to a cycle of length ℓ of φ_{A,e_1}.*

Proposition 9 ([9, 21])**.** *Let $q > 1$ be a power of prime p. Let $e \geq 1$ and s be its size (in base p) $s := \lceil \log_p(e) \rceil$. Let $G(X) = P(X)^e$ be a power of an irreducible polynomial $P(X) \neq X - 1$ of degree d. Then the permutation $\varphi_{G(X)}$ has the following cycle count:*

cycle length	*1*	$\mathrm{ord}(P(X))$	$\mathrm{ord}(P(X))p^i,\ i \in [\![1, s-1]\!]$	$\mathrm{ord}(P(X))p^s$
cardinality	*1*	$\frac{q^d-1}{\mathrm{ord}(P(X))}$	$\frac{q^{dp^i}-q^{dp^{i-1}}}{\mathrm{ord}(P(X))p^i}$	$\frac{q^{de}-q^{dp^{s-1}}}{\mathrm{ord}(P(X))p^s}$

Proof. First, $P(X) = 0$ is a fixed point. Let $Q(X) \neq 0$ and $i \geq 1$ such that $\varphi^i(Q(X)) = Q(X)$. It is

$$\varphi^i_{G(X)}(Q(X)) = Q(X) \iff (X^i - 1)(Q(X)) = 0. \tag{4}$$

Decomposing $Q(X)$ as $Q(X) = Q(X)'P(X)^j$ with $q'(X)$ invertible and $0 \leq j < e$, we obtain

$$\varphi^i_{G(X)}(Q(X)) = Q(X) \iff (X^i - 1)P(X)^j = 0 \iff X^i - 1 \in (P(X)^{e-j}), \tag{5}$$

so the least i verifying this equation is precisely $\mathrm{ord}(P(X)^{e-j})$. With $e - j = 1$, we conclude that the elements lying on cycles of length $\mathrm{ord}(P(X))$ are the ones belonging to $\left(P(X)^{e-1}\right) \setminus \{0\}$; there are $q^d - 1$ of them. Otherwise, $e - j > 1$ and $\mathrm{ord}(P(X)^{e-j}) = p^{\lceil \log_p(e-j) \rceil}\mathrm{ord}(P(X))$. In that case, the elements on cycles of length $p^k\mathrm{ord}(P(X))$ with $1 \leq k < s$ correspond to the elements of $(P(X)^{e-p^k}) \setminus (P(X)^{e-p^{k-1}})$. There are $q^{dp^k} - q^{dp^{k-1}}$ of them. Elements on cycles of length $p^s\mathrm{ord}(P(X))$ are the remaining ones, that is elements of $\mathbb{F}_q[X]/(Q(X)) \setminus \left(P(X)^{e-p^{s-1}}\right)$; there are $q^{de} - q^{dp^{s-1}}$ of them. □

Proposition 10 ([9, 23])**.** *Let $1 < q$ be a power of prime p. Let $e \geq 1$ and s be its size (in base p) $s := \lceil \log_p(e) \rceil$. Let $G(X) = (X-1)^e$. Then the cycle structure of $\varphi_{G(X),U(X)}$ depends on the nature of $U(X)$ as follows:*

1. if $X-1 \mid U$ then its cycle count is:

cycle length	*1*	$p^i,\ i \in [\![1, s-1]\!]$	p^s
cardinality	q	$\frac{q^{p^i}-q^{p^{i-1}}}{p^i}$	$\frac{q^e-q^{p^{s-1}}}{p^s}$

2. if $X-1 \nmid U$ and:

(a) e is a power of p then all cycles are of length p^{s+1} (and there are q^e/p^{s+1} of them),

*(b) e is **not** a power of p then all cycles are of length p^s (and there are q^e/p^s of them).*

Proof. With $\nu(Q(X))$ we denote the $(X-1)$-valuation of $Q(X) \in \mathbb{F}_p[X]$, i.e. the maximal $l \geq 0$ with $(X-1)^l | Q$ and $\nu(0) = \infty$.

Let $i \geq 0$ and consider the equation $\varphi^{p^i}(Q(X)) = Q(X)$ in $\mathbb{F}[X]/(G(X))$. It follows

$$\varphi^{p^i}(Q(X)) = Q(X) \iff X^{p^i}Q(X)+U(X)S(X)_{p^i} = Q \iff (X-1)^{p^i}Q+U(X)S(X)_{p^i} = 0. \quad (6)$$

In particular for $i \geq s$, Eq. (6) becomes

$$0Q(X) + U(X)S(X)_{p^i} = 0 \quad (7)$$

which has either none or all $Q(X)$ as solutions, depending on whether $U(X)S(X)_{p^i} \neq 0$ or not, that is, whether $\nu(U(X)S(X)_{p^i}) < e$ or not.

Case 1) $X-1 \mid U$ With $i = s$, we get $\nu(US_{p^s}) = \nu(U) + \nu(S(X)_{p^s}) \geq 1 + (p^s - 1) \geq e$ so $U(X)S(X)_{p^i} = 0$ and all $Q(X)$ are solutions of Eq. (7). Therefore all $Q(X)$ lies on cycles of length dividing p^s.

For $i \in [\![0, s-1]\!]$, by decomposing $U(X)S(X)_{p^i} = (X-1)^{\nu(U(X)S(X)_{p^i})}Z$ with $X-1 \nmid Z$, Eq. (6) becomes:

$$\varphi^{p^i}(Q(X)) = Q(X) \iff (X-1)^{p^i}\left(Q(X) + (X-1)^{\nu(U(X)S(X)_{p^i})-p^i}Z\right) = 0$$

$$\iff Q(X) \in \mathcal{I} - (X-1)^{\nu(U(X)S(X)_{p^i})-p^i}Z$$

where $\mathcal{I}$ is the ideal generated by $(X-1)^{e-p^i}$. So there are exactly $|\mathcal{I}| = q^{p^i}$ elements located on cycles whose length divides p^i. Therefore for any $i \in [\![0, s-1]\!]$, there are $q^{p^i} - q^{p^{i-1}}$ elements on cycles of length p^i and $q^e - q^{s-1}$ elements on cycles of length p^s.

Case 2b) $X-1 \nmid U(X)$**,** e **not a power of** p In that case, $\nu(U(X)S(X)_{p^s}) = \nu(U(X)) + \nu(S(X)_{p^s}) = 0 + (p^s - 1) \geq e$. Therefore, every $Q(X)$ is solution of Eq. (7) (with $i = s$) and thus every $Q(X)$ lies on cycles of length dividing p^s. However for $i < s$, Eq. (6) becomes

$$U(X)^{-1}(X-1)^{p^i}Q(X) = -S(X)_{p^i} \quad (8)$$

but the left-hand side belongs to the ideal $\left((X-1)^{p^i}\right)$ while on the right-hand side we have $S(X)_{p^i} \in \left((X-1)^{p^i-1}\right) \setminus \left((X-1)^{p^i}\right)$ which means that no $Q(X)$ lies on cycles of length p^i for $i < s$.

Case 2a) $X - 1 \nmid U$**,** $e = p^s$ In that case, $\nu(S(X)_{p^{s+1}}) = p^{s+1} - 1 \geq p^s = e$ so Eq. (7) with $i = s + 1$ admits all $Q(X)$ as solutions. Thus, all $Q(X)$ lie on cycles of length divisible by p^{s+1}. But for $i \leq s$, Eq. (8) has no solution for the same reasons as above, so all $Q(X)$ lie on cycles of length p^{s+1}. □

So from Proposition 9 and 10 it follows that L_A and $L_B + u$ with $\chi_{L_A} = (P(X))^l$ and $\chi(L_B) = (X - 1)^{l\deg(P(X))}$ cannot have the same cycle structure.

Thus given an affine mapping $A = T_u \circ L_A \in \mathrm{AGL}(n, \mathbb{F}_p)$

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots \ddots & \vdots & \\ 0 & 0 & \dots & A_r \end{pmatrix},$$

where w.l.o.g. A_i are companion matrices, the $(X - 1)$-primary component, if exists, is made of the last blocks and u belongs to the $(X - 1)$-primary component. To get another affine mapping B with the same cycle structure one just has to exchange all blocks A_i conforming to Proposition Proposition 6. The blocks belonging to the $(X - 1)$-primary component stay the same. Here u can be exchanged by u', where $U(X)$ and $u'(X)$ fulfill the same condition of Proposition 10. The pairs $(A'$,$B')$, where $A' = MAM^{-1}, B' = NBN^{-1}, M, N \in \mathrm{GL}(n, \mathbb{F})$ cover the sought for general case.

In the next section we will prove some facts about the Weierstraß normal form depending on the order of A. This has interesting implications for the self-equivalence $S \circ A = B \circ S$.

5.3 Affine and Linear Self-Equivalences – Orders Not a Power of p

We will now assume that $S \circ A = B \circ S$ (i.e., $\Gamma_S(A, B) = p^n$) and take a deeper look at the implications of $\mathrm{ord}(A)$ not being a power of p. Hence, let us write $\mathrm{ord}(A) = p^m \cdot d$ with m maximal. Without loss of generality, we can assume that $m = 0$, as we can simply consider A^{p^m} instead of A. Thus, we will assume that $\mathrm{ord}(A) = d$ is not divisible by p. Then, we note that by Lemma 2, $\mathrm{ord}(A) \in \{\mathrm{ord}(L_A), p \cdot \mathrm{ord}(L_A)\}$, implying that $\mathrm{ord}(A) = \mathrm{ord}(L_A)$. But then $\mathrm{ord}(L_A)$ is not a multiple of p, implying that L_A is similar to a certain block-diagonal matrix. The next lemma is a well-known consequence on the theory given before. Still we give the proof to ease following the results.

Lemma 5. *Let L be a linear permutation of* $\mathbb{F}_p^n$ *and let* $k = \dim(\mathrm{Fix}(L))$. *If* $\mathrm{ord}(L)$ *is not a multiple of p, then L is similar to*

$$\begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix},$$

where I *is the* $k \times k$ *identity matrix and* L' *a linear permutation on* $\mathbb{F}_p^{n-k}$ *without any non-trivial fixed points.*

Proof. By Remark 2 we have that $m_L(X)$ is the product of distinct irreducible polynomials of multiplicity 1. Moreover in case of $k \geq 1$, we have $m_L(X) = (X - 1)G(X)$ with $G(1) \neq 0$. Hence in the Weierstraß normal form the Blocks A_i belonging to the $(X - 1)$-primary component are $1 \in \mathrm{GL}(1, \mathbb{F}_p)$. This ends the proof. □

This similarity can easily be translated to affine maps.

Proposition 11. *Let $A \in \mathrm{AGL}(n, \mathbb{F}_p)$ and let $A = L_A + c_A$ with L_A linear and $c_A \in \mathbb{F}_p^n$. If $\mathrm{ord}(A)$ is not a multiple of p, then there exist $Q \in \mathrm{AGL}(n, \mathbb{F}_p)$ such that*

$$Q^{-1}AQ = \begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix}$$

where I is the $k \times k$ identity matrix ($k = \dim(\mathrm{Fix}(L_A))$) and L' a linear permutation on $\mathbb{F}_p^{n-k}$ without any non-trivial fixed points. Additionally, $\mathrm{Fix}(A) = \mathrm{Fix}(L_A) + c_Q$ with $c_Q = Q(0)$.

Proof. We know from Lemma 2 that $\mathrm{ord}(A) \in \{\mathrm{ord}(L_A), p \cdot \mathrm{ord}(L_A)\}$. Since $\mathrm{ord}(A)$ is not a multiple of p, this implies that $\mathrm{ord}(A) = \mathrm{ord}(L_A)$, and $\mathrm{ord}(L_A)$ cannot be a multiple of p. Hence, we know from the previous lemma that there exists a linear permutation L_Q such that

$$L_Q^{-1} L_A L_Q = \begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix},$$

where I is the $k \times k$ identity matrix and L' a linear permutation on $\mathbb{F}_p^{n-k}$ without any non-trivial fixed points. Hence, $L_Q^{-1} A L_Q$ is the affine map defined by

$$x \mapsto \begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix} \cdot x + L_Q^{-1}(c_A).$$

But $\mathrm{ord}(A) = \mathrm{ord}(L_A)$ also implies that

$$0 = \sum_{i=0}^{\mathrm{ord}(L_A)-1} \begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix}^i \cdot L_Q^{-1}(c_A).$$

Hence, the first k coordinates of $L_Q^{-1}(c_A)$ need to be zero. Since L' has no non-trivial fixed points, meaning that $L' - I$ has full rank, there needs to exist a $c'_Q \in 0^k \times \mathbb{F}_p^{n-k}$ such that

$$\begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix} \cdot c'_Q - c'_Q = -L_Q^{-1}(c_A).$$

By defining c_Q as $L_Q(c'_Q)$ and Q as the map $x \mapsto L_Q(x) + c_Q$, we get that

$$Q^{-1}AQ = \begin{pmatrix} I & 0 \\ 0 & L' \end{pmatrix}.$$

At last, we note that $0 = Q^{-1}AQ(0)$ has to hold, which implies that $0 = A(c_Q) - c_Q$. In other words, c_Q is a fixed point of A, and therefore $\mathrm{Fix}(A) = \mathrm{Fix}(L_A) + c_Q$. □

Since $\mathrm{ord}(A) = \mathrm{ord}(B)$ has to hold if S is bijective, and the argument applies for both, A and B, we get the following corollary.

Corollary 9. *Let $S \colon \mathbb{F}_p^n \to \mathbb{F}_p^n$ be bijective and $A, B \in \mathrm{AGL}(n, \mathbb{F}_p)$ with $S \circ A = B \circ S$. Let $L_A = A - A(0)$ be the linear part of A. If $\mathrm{ord}(A) = dp^m$ for d not a multiple of of p, there exist affine maps $Q_A, Q_B \in \mathrm{AGL}(n, \mathbb{F}_p)$ such that for $\hat{S}$ defined as $Q_B \circ S \circ Q_A$ it holds that*

$$\hat{S} \circ \begin{pmatrix} I & 0 \\ 0 & L'_A \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & L'_B \end{pmatrix} \circ \hat{S},$$

where I is the $k \times k$ identity matrix ($k = \dim(\mathrm{Fix}(L_A^{p^m}))$) and L'_A, L'_B are linear permutations on $\mathbb{F}_p^{n-k}$ without any non-trivial fixed points.

Proof. The result follows from the previous proposition and the fact that, if $\mathrm{ord}(A) = dp^m$, then $\mathrm{ord}(A^{p^m}) = d$ and a self-equivalence $S \circ A = B \circ S$ implies the self-equivalence $S \circ A^{p^m} = B^{p^m} \circ S$. □

In other words, if (A, B) is an affine self-equivalence of S with $\mathrm{ord}(A) = \mathrm{ord}(B)$ not being a multiple of p then S is affine equivalent to S' with a linear self-equivalence, and the mappings belonging to the linear self-equivalences are similar to the block-diagonal matrix described above.

6 Differential Attacks Against Conjugate Ciphers

As hinted in [4, Appendix A], commutative cryptanalysis has an intricate relationship with differential cryptanalysis, but also with differential cryptanalysis of *conjugate ciphers*. Such a methodology consists in studying a function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ by considering equations of form

$$H \circ F \circ H^{-1}(x + \Delta^{\mathrm{in}}) = H \circ F \circ H^{-1}(x) + \Delta^{\mathrm{out}};$$

where $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ is a well-chosen bijection. This is in line with the work of Beierle, Canteaut & Leander [6] which first investigated the relationship between non-linear invariants/approximations of F and the linear invariants/approximations of $H \circ F \circ H^{-1}$.

Standard differential cryptanalysis naturally corresponds to $H = I$ and such a generalization is actually interesting only if H is non-linear as the differential properties of a function are preserved by affine equivalence.

Another line of papers [12, 15, 11, 10] tackles a similar problem from a group-theoretic perspective. In particular, Civino, Blondeau & Sala [15] consider equations of a vectorial Boolean function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ of the form

$$F(x \diamond \Delta^{\mathrm{in}}) = F(x) \diamond \Delta^{\mathrm{out}};$$

where $\diamond\colon \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ is an Abelian group operation for $\mathbb{F}_2^n$. This generalizes the usual case which corresponds to the case where $\diamond$ is the bitwise addition.

The objective of this section is to bridge the gap between these three cryptanalysis approaches. The link between commutative cryptanalysis and differential cryptanalysis of conjugates is recalled in Section 6.1, while the link between the latter technique and differential cryptanalysis with alternative group laws is detailed in Section 6.3. Finally, we provide examples extracted from [15, 10, 4, 3] to illustrate these intricate links. In the following, for any bijection $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ and any function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$, we denote by F^H the conjugate of F by H, that is, $F^H := H \circ F \circ H^{-1}$.

6.1 Differential Cryptanalysis of Conjugates and Commutative Cryptanalysis

Let us recall the link between commutation and conjugation that is shown in [4, Appendix A]. Let $F, H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ with H a bijection and $\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}} \in \mathbb{F}_2^n$. A probability-1 differential $\Delta^{\mathrm{in}} \xrightarrow{F^H} \Delta^{\mathrm{out}}$ corresponds to a constant derivative of F^H, *i.e.*:

$$\forall x \in \mathbb{F}_2^n, \quad H \circ F \circ H^{-1}(x + \Delta^{\mathrm{in}}) - H \circ S \circ H^{-1}(x) = \Delta^{\mathrm{out}}. \tag{9}$$

Eq. (9) can equivalently be written as $T_{\Delta^{\mathrm{out}}} \circ H \circ F \circ H^{-1} = H \circ F \circ H^{-1} \circ T_{\Delta^{\mathrm{in}}}$, but also as $T_{\Delta^{\mathrm{out}}}^{H^{-1}} \circ F = F \circ T_{\Delta^{\mathrm{in}}}^{H^{-1}}$, by conjugating the latter expression by H^{-1}.

This proves that the differentials with probability 1 of the conjugate function F^H correspond to commutation relations with probability 1 through the original function F. In the probabilistic case, the same reasoning proves that:

$$\Gamma_{F^H}(T_{\Delta^{\text{in}}}, T_{\Delta^{\text{out}}}) = \Gamma_F(T_{\Delta^{\text{in}}}^{H^{-1}}, T_{\Delta^{\text{out}}}^{H^{-1}}); \tag{10}$$

This is summarized in the following proposition.

Proposition 12 (Conjugation and commutation)**.** *Let $F, H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$, with H a bijection. Studying the differential properties of F^H is equivalent to studying the commutative properties of F with respect to commutants among the group $\{T_c^{H^{-1}} \mid c \in \mathbb{F}_2^n\} = H^{-1}TH$; where the group of translations is denoted by $T := \{T_c \mid c \in \mathbb{F}_2^n\}$.*

Proposition 12 is the differential counterpart of the work of Beierle, Canteaut & Leander [6]. Indeed, commutative properties of a block cipher are related to differential properties of its conjugates, in the same way as non-linear approximations (and in particular invariants) are related to linear properties of the conjugates.

However, Proposition 12 also points out that the considered class of commutants is really *restrictive*. Indeed, $H^{-1}TH$ is a conjugate of the group T and this implies that, like T, $H^{-1}TH$ is an Abelian 2-elementary regular group.

A group G is said to be *2-elementary* if each non-zero element is of order 2. A group $G \subseteq \text{Perm}(\mathbb{F}_2^n)$ is said to be *regular* if it satisfies

$$\forall (x, y) \in \mathbb{F}_2^n, \quad \exists!\ g \in G,\ g(x) = y.$$

The conjugation of T (or the regularity) in particular implies that $H^{-1}TH \setminus \{I\}$ contains only involutions without fixed points.

For these reasons, differential cryptanalysis of a conjugate cipher can only provide *exact* results about (either deterministic or probabilistic) commutations relations $A \xrightarrow{F} B$ where $A, B\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ are strongly constrained. For instance, it cannot exactly handle pairs (A, B) where one of the commutants is not a fixed-point-free involution (because of the 2-elementarity) or pairs for which there exists $x \in \mathbb{F}_2^n$, such that $A(x) = B(x)$ (because of the regularity). Nonetheless, for any of these situations, it is still possible to *approximate* A and B by two elements of a given $H^{-1}TH$ to obtain an *approximation* on the number of solutions of a given relation $A \xrightarrow{F} B$.

6.2 Elementary Regular Subgroups of the Symmetric Group

In the following, 0 always stands for the identity element of the group $(\mathbb{F}_2^n, +)$. In their paper [15], Civino, Blondeau & Sala present a cryptanalysis framework based on alternative group laws. Such an operation is built by first considering a regular 2-elementary Abelian subgroup $G \subseteq \text{Perm}(\mathbb{F}_2^n)$ of the symmetric group. In particular, if G is regular, then $\{g(0) \mid g \in G\}$ is the full space $\mathbb{F}_2^n$. We can then enumerate G as $G = \{g_a \mid a \in \mathbb{F}_2^n\}$ where g_a is the unique function $g \in G$ that satisfies $g(0) = a$. Such a group mimics the group of translations $T := \{T_a\colon x \mapsto x + a\}$ which is indeed made of fixed-point-free involutions (except $T_0 = I$), which commute one with the others, and where the only one that satisfies $T_a(x) = y$ is T_{x+y}. For these reasons, we reserve the notation $\mathcal{T} = \{\mathcal{T}_a \mid a \in \mathbb{F}_2^n\}$ to regular subgroups of $\text{Perm}(\mathbb{F}_2^n)$ where for any $a \in \mathbb{F}_2^n$, we have $\mathcal{T}_a(0) = a$.

Let us clarify this mimicry. First, it is possible to build a group law $\diamond$ for which the group of translations is any regular 2-elementary (Abelian) subgroup of $\text{Perm}(\mathbb{F}_2^n)$. The following proposition is the keystone of [15].

Proposition 13 (Group law based on a regular subgroup)**.** *Let $\mathcal{T} \subseteq \mathrm{Perm}(\mathbb{F}_2^n)$ be a regular Abelian subgroup of the symmetric group. Let us define the operator $\diamond \colon \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ as:*

$$\forall x, y \in \mathbb{F}_2^n, \quad x \diamond y := \mathcal{T}_x(y).$$

Then $(\mathbb{F}_2^n, \diamond)$ is an Abelian group whose identity element $0_\diamond$ satisfies $0_\diamond = 0$. Furthermore $\mathcal{T}$ coincides with its group of translations.

Proof. Let us first observe that $\diamond$ is well-defined. It is a commutative operator because for any $x, y \in \mathbb{F}_2^n$ we have:

$$x \diamond y = \mathcal{T}_x(y) = \mathcal{T}_x(\mathcal{T}_y(0)) = \mathcal{T}_y(\mathcal{T}_x(0)) = \mathcal{T}_y(x) = y \diamond x.$$

Furthermore, for any $x \in \mathbb{F}_2^n$, we have by definition $x \diamond 0 = \mathcal{T}_x(0) = x$ so 0 is the identity element. As $\mathcal{T}_x^{-1} \in \mathcal{T}$, there exists y such that $\mathcal{T}_x^{-1} = \mathcal{T}_y$. We then observe that

$$x \diamond y = \mathcal{T}_x(y) = \mathcal{T}_x(\mathcal{T}_y(0)) = \mathcal{T}_x(\mathcal{T}_x^{-1}(0)) = 0;$$

which makes y the inverse of x. Finally for any $x, y, z \in \mathbb{F}_2^n$, we observe that:

$$x \diamond (y \diamond z) = \mathcal{T}_x(\mathcal{T}_y(z)), \quad \text{and} \quad (x \diamond y) \diamond z = \mathcal{T}_x(y) \diamond z = \mathcal{T}_{\mathcal{T}_x(y)}(z).$$

But $\mathcal{T}_{\mathcal{T}_x(y)}$ satisfies $\mathcal{T}_{\mathcal{T}_x(y)}(0) = \mathcal{T}_x(y)$ and $\mathcal{T}_x \circ \mathcal{T}_y$ belongs to $\mathcal{T}$ and also satisfies $\mathcal{T}_x \circ \mathcal{T}_y(0) = \mathcal{T}_x(y)$. By the regularity of $\mathcal{T}$, we necessarily have that $\mathcal{T}_{\mathcal{T}_x(y)} = \mathcal{T}_x \circ \mathcal{T}_y$ and therefore $\diamond$ is associative. So $(\mathbb{F}_2^n, \diamond)$ is indeed an Abelian group. Furthermore, for any $a \in \mathbb{F}_2^n$, the function $x \mapsto x \diamond a$ coincides by construction with $\mathcal{T}_a$. □

Remark 5. The previous proposition is stated without supposing that $\mathcal{T}$ is 2-elementary. If it is the case, then for any $x \in \mathbb{F}_2^n$, we have $\mathcal{T}_x^{-1} = \mathcal{T}_x$ and therefore x is its own inverse for the group law $\diamond$. In the following, this will always be the case. Furthermore, by construction we have $0_\diamond = 0$. Therefore, we no longer make a distinction between the two identity elements and always use the notation 0 for the identity element.

We can also go further in the parallel between such a group $\mathcal{T}$ and the group of translations thanks to the following well-known result.

Proposition 14. *Up to isomorphism, there exists a single 2-elementary group of order 2^n, which is $\mathbb{F}_2^n$.*

So any 2-elementary regular subgroup $\mathcal{T}$ of $\mathrm{Perm}(\mathbb{F}_2^n)$ is isomorphic to the group of translations T. But due to a result of Dixon [19, proof of Lemma 1], we can be even more precise as two isomorphic regular subgroups of S_n are necessarily conjugate.

Proposition 15 (Isomorphic and conjugate regular subgroups [19])**.** *Let $n \geq 1$ and let S_n be the symmetric group of $[\![0, n-1]\!]$. Let G, K be two regular subgroups of S_n such that there exists a group isomorphism $\phi \colon G \to K$. Then there exists $\sigma \in S_n$ such that $\sigma K \sigma^{-1} = G$.*

Proof. (Adapted from [19, Proof of Lemma 1]) Let us define the bijection $\sigma \colon [\![0, n-1]\!] \to [\![0, n-1]\!]$ as:

$$\forall g \in G, \quad \sigma(g(0)) := \phi(g)(0). \tag{11}$$

The bijection σ is well-defined. Indeed, G is regular so $\{g(0) \mid g \in G\} = [\![0, n-1]\!]$, but we also have $\{\phi(g)(0) \mid g \in G\} = \{k(0) \mid k \in K\} = [\![0, n-1]\!]$ because ϕ is bijective and K is regular. By definition we also note that:

$$\forall g \in G, \quad g(0) = \sigma^{-1}(\phi(g)(0)). \tag{12}$$

Let us enumerate K as $K = \{k_i \mid i \in [\![0, n-1]\!]\}$ where $k_i \in K$ is the unique $k \in K$ such that $k_i(0) = i$.

Let $g \in G$ and let us consider $\sigma \circ g \circ \sigma^{-1}$. Let $i \in [\![0, n-1]\!]$ and let us denote $\widetilde{g} = \phi^{-1}(k_i)$ and observe that by construction we have:

$$\phi(\widetilde{g})(0) = \phi(\phi^{-1}(k_i))(0) = k_i(0) = i. \tag{13}$$

Then it holds that:

$$\begin{align}
\sigma \circ g \circ \sigma^{-1}(i) &= \sigma \circ g \circ \sigma^{-1}\left(\phi\left(\widetilde{g}\right)(0)\right) \tag{14}\\
&= \sigma \circ g\left(\widetilde{g}(0)\right) \tag{15}\\
&= \sigma\left(g \circ \widetilde{g}(0)\right) \tag{16}\\
&= \phi\left(g \circ \widetilde{g}\right)(0) \tag{17}\\
&= \phi(g) \circ \phi(\widetilde{g})(0) \tag{18}\\
&= \phi(g)(i). \tag{19}
\end{align}$$

Eq. (14) comes from Eq. (13); Eq. (15) from Eq. (12); Eq. (16) is only a different bracket grouping; Eq. (17) comes from Eq. (11); Eq. (18) is due to the morphism property of ϕ and finally Eq. (19) is again due to Eq. (13).

All in all, it holds that $\sigma g \sigma^{-1} = \phi(g)$, and this implies that $\sigma G \sigma^{-1} = K$. □

Let $\mathcal{T} = \{\mathcal{T}_a \mid a \in \mathbb{F}_2^n\}$ be a 2-elementary regular subgroup of $\mathrm{Perm}(\mathbb{F}_2^n)$ and $T = \{T_a \colon x \mapsto x + a \mid a \in \mathbb{F}_2^n\}$ be the group of translations for the usual addition law.

There therefore exists σ such that $\sigma \mathcal{T} \sigma^{-1} = T$. This also implies that there exists a bijection $\psi \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ which satisfies:

$$\forall a \in \mathbb{F}_2^n, \quad \sigma \circ \mathcal{T}_{\psi(a)} \circ \sigma^{-1} = T_a. \tag{20}$$

By evaluating the previous equations at point $\sigma(0)$, we obtain:

$$\forall a \in \mathbb{F}_2^n, \quad \sigma \circ \mathcal{T}_{\psi(a)}(0) = \sigma(0) + a, \quad \iff \quad \sigma(\psi(a)) = \sigma(0) + a.$$

In other words, for a given σ such $\sigma \mathcal{T} \sigma^{-1} = T$, there exists a single ψ satisfying Eq. (20) and it is defined as:

$$\forall a \in \mathbb{F}_2^n, \quad \psi(a) := \sigma^{-1}(\sigma(0) + a).$$

Let $c, a \in \mathbb{F}_2^n$. The group $\mathcal{T}$ being Abelian, it holds that $\mathcal{T}_c \circ \mathcal{T}_{\psi(a)} = \mathcal{T}_{\psi(a)} \circ \mathcal{T}_c$, *i.e.*, $\mathcal{T}_c \circ \mathcal{T}_{\psi(a)} \circ \mathcal{T}_c^{-1} = \mathcal{T}_{\psi(a)}$. But as $\mathcal{T}$ is also 2-elementary, any element is its own inverse so $\mathcal{T}_c \circ \mathcal{T}_{\psi(a)} \circ \mathcal{T}_c = \mathcal{T}_{\psi(a)}$.

This implies that for any σ, ψ satisfying Eq. (20), it also holds that for any $c \in \mathbb{F}_2^n$:

$$\forall a \in \mathbb{F}_2^n, \quad \sigma \circ \mathcal{T}_c \circ \mathcal{T}_{\psi(a)} \circ \mathcal{T}_c \circ \sigma^{-1} = T_a. \tag{21}$$

In other words, σ can be replaced by $\sigma \circ \mathcal{T}_c$ for any $c \in \mathbb{F}_2^n$. In particular, with $c = \sigma^{-1}(0)$, we observe that $\sigma \circ \mathcal{T}_{\sigma^{-1}(0)}(0) = \sigma(\sigma^{-1}(0)) = 0$, so without loss of generality, we can always consider σ such that $\sigma(0) = 0$. In that case $\psi = \sigma^{-1}$ and Eq. (20) can be simplified into:

$$\forall a \in \mathbb{F}_2^n, \quad \sigma \circ \mathcal{T}_{\sigma^{-1}(a)} \circ \sigma^{-1} = T_a.$$

We restate this in the following proposition.

Proposition 16. *Let $\mathcal{T} = \{\mathcal{T}_a \mid a \in \mathbb{F}_2^n\}$ be a 2-elementary regular subgroup of* $\mathrm{Perm}(\mathbb{F}_2^n)$ *and $T = \{T_a \colon x \mapsto x + a \mid a \in \mathbb{F}_2^n\}$ be the group of translations for the usual addition law. Then there exists $\sigma \in \mathrm{Perm}(\mathbb{F}_2^n)$ such that:*

$$\forall a \in \mathbb{F}_2^n, \quad \sigma \circ \mathcal{T}_{\sigma^{-1}(a)} \circ \sigma^{-1} = T_a. \tag{22}$$

6.3 Differential Cryptanalysis of Conjugates and $\diamond$-Differential Cryptanalysis

6.3.1 Comparative Study

As already explained, the authors of [15], but also the ones of [10], consider equations of a Boolean function $F \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ of the form:

$$F(x \diamond \Delta^{\mathrm{in}}) = F(x) \diamond \Delta^{\mathrm{out}};$$

where $\diamond \colon \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ is an Abelian group operation defined as in Proposition 13 for a 2-elementary regular subgroup $\mathcal{T} \subseteq \mathrm{Perm}(\mathbb{F}_2^n)$. This is equivalent to saying that they only consider group laws $\diamond$ that are commutative and for which each element $x \in \mathbb{F}_2^n$ is its own inverse. In the following, we translate the framework of [15, 10] into the differential cryptanalysis of some conjugate function.

Remark 6. We stress that the relation between alternative group laws and conjugation is *beyond any doubt* well-known by the authors of [12, 15, 11, 10] and mentioned at multiple times in these papers. The novelties of the following formulation is that it relates this technique to commonly-used tools and notions of standard cryptanalysis. Furthermore, this dictionary used in one way or the other provides more examples to each approach.

The notion of $\diamond$-differential probability is defined in [15] for any ordered pair $(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}}) \in (\mathbb{F}_2^n)^2$ as:

$$p_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}}) := \frac{1}{2^n} |\{x \in \mathbb{F}_2^n \mid F(x \diamond \Delta^{\mathrm{in}}) = F(x) \diamond \Delta^{\mathrm{out}}\}|.$$

We can introduce the size of the associated set of solutions $\Gamma_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}})$ that we define as:

$$\Gamma_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}}) := |\{x \in \mathbb{F}_2^n \mid F(x \diamond \Delta^{\mathrm{in}}) = F(x) \diamond \Delta^{\mathrm{out}}\}|.$$

By Proposition 16, there exists $H \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that:

$$\forall a \in \mathbb{F}_2^n, \quad H \circ \mathcal{T}_{H^{-1}(a)} \circ H^{-1} = T_a.$$

Then, $\Gamma_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}})$ can be given as

$$\begin{aligned}
\Gamma_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}}) &= |\{x \in \mathbb{F}_2^n \mid \mathcal{T}_{\Delta^{\mathrm{out}}} \circ F(x) = F \circ \mathcal{T}_{\Delta^{\mathrm{in}}}(x)\}| \\
&= |\{x \in \mathbb{F}_2^n \mid H^{-1} \circ T_{H(\Delta^{\mathrm{out}})} \circ H \circ F(x) = F \circ H^{-1} \circ T_{H(\Delta^{\mathrm{in}})} \circ H(x)\}| \\
&= \Gamma_F \left(T_{H(\Delta^{\mathrm{in}})}^{H^{-1}}, T_{H(\Delta^{\mathrm{out}})}^{H^{-1}} \right).
\end{aligned}$$

Combined with Eq. (10), we obtain the following proposition.

Proposition 17 ($\diamond$-differential, conjugation & commutation)**.** *Let $(\mathbb{F}_2^n, \diamond)$ be an Abelian group such that $\mathcal{T} := \{x \mapsto x \diamond c \mid c \in \mathbb{F}_2^n\}$ is 2-elementary and regular. Let $H \colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that: $\forall a \in \mathbb{F}_2^n, \quad H \circ \mathcal{T}_{H^{-1}(a)} \circ H^{-1} = T_a$. Then, it holds that*

$$\Gamma_F^{\diamond}(\Delta^{\mathrm{in}}, \Delta^{\mathrm{out}}) = \Gamma_F \left(T_{H(\Delta^{\mathrm{in}})}^{H^{-1}}, T_{H(\Delta^{\mathrm{out}})}^{H^{-1}} \right) = \Gamma_{F^H} \left(T_{H(\Delta^{\mathrm{in}})}, T_{H(\Delta^{\mathrm{out}})} \right). \tag{23}$$

In other words, Proposition 17 states that studying the $\diamond$-differential properties of F is equivalent to either studying the differential properties of its conjugate F^H or the commutation with commutants among the group $T^{H^{-1}}$.

Stated otherwise, the two methodologies from [15, 10] and from [4] coincide: despite the clear difference of flavors, they both study the differential properties of a conjugate cipher E_k^H decomposed as

$$E_k^H = H \circ F_k^{(R)} \circ \ldots \circ F_k^{(1)} \circ H^{-1},$$

by leveraging weaknesses of the conjugate round functions $(F_k^{(r)})^H$ for any r. We also note that both approaches are nourished from the other point-of-view.

In order to use this dictionary in both ways, we clarify that such H is in practice easy to build.

Lemma 6 (Characterization of H)**.** *Let* $(\mathbb{F}_2^n, \diamond)$ *be an Abelian group such that* $\mathcal{T} := \{x \mapsto x \diamond c \mid c \in \mathbb{F}_2^n\}$ *is 2-elementary and regular. Let* $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$*. Then H satisfies* $\forall a \in \mathbb{F}_2^n, \quad H \circ \mathcal{T}_{H^{-1}(a)} \circ H^{-1} = T_a$ *if and only if H is a group isomorphism from* $(\mathbb{F}_2^n, \diamond)$ *to* $(\mathbb{F}_2^n, +)$*.*

Proof. The mapping H satisfies the first condition if and only if it holds that:

$$\forall\ x, y \in \mathbb{F}_2^n, \quad x \diamond y = \mathcal{T}_y(x) = H^{-1} \circ T_{H(y)} \circ H(x).$$

This is naturally equivalent to:

$$\forall\ x, y \in \mathbb{F}_2^n, \quad H(x \diamond y) = T_{H(y)}(H(x)) = H(x) + H(y).$$

□

Because $(\mathbb{F}_2^n, \diamond)$ is actually an n-dimensional $\mathbb{F}_2$-vector space, we can fix ourselves a basis $(b_1, \ldots, b_n)$ such that any element $x \in \mathbb{F}_2^n$ can be uniquely decomposed as $x = y_1 b_1 \diamond y_1 b_1 \diamond \ldots \diamond y_n b_n$, with $y_i \in \mathbb{F}_2$ for all i. By Lemma 6, an isomorphism H must therefore satisfy:

$$\forall x \in \mathbb{F}_2^n, \quad H(x) = \sum_{i=1}^{n} y_i H(b_i).$$

Building such a H is therefore equivalent to selecting a basis $(B_1, \ldots, B_n)$ of $(\mathbb{F}_2^n, +)$, defining $H(b_i) = B_i$ for any i, and expanding the definition by linearity.

6.3.2 About the Weak-Key Space of [15].

The authors of [15] introduced a weak-key space denoted by $W^\diamond$, which is defined as:

$$W^\diamond := \{k \in \mathbb{F}_2^n \mid T_k = \mathcal{T}_k\}.$$

Remark 7. The set $W^\diamond$ is defined as $W^\diamond = \{k \in \mathbb{F}_2^n \mid T_k \in \mathcal{T}\}$ in [11]. Both definitions coincide because $\mathcal{T}$ is regular, so the condition $T_k \in \mathcal{T}$ necessarily implies that T_k and $\mathcal{T}_k$ must coincide because $T_k(0) = k = \mathcal{T}_k(0)$.

From the conjugate point-of-view, $W^\diamond$ can be described as:

$$W^\diamond = \{k \in \mathbb{F}_2^n \mid T_k = \mathcal{T}_k\} = \{k \in \mathbb{F}_2^n \mid T_k = H^{-1} \circ T_{H(k)} \circ H\} = \{k \in \mathbb{F}_2^n \mid T_k^H = T_{H(k)}\}.$$

In other words, $W^\diamond$ is the set of k for which the conjugate of T_k is still a constant addition, with a possibly different constant. But as function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ is affine with L as linear part if and only if it satisfies:

$$\forall \Delta^{\text{in}}, \Delta^{\text{out}} \in \mathbb{F}_2^n,\ \Pr\left[\Delta^{\text{in}} \xrightarrow{F} \Delta^{\text{out}}\right] = \left\{ \begin{array}{ll} 1 & \text{if } \Delta^{\text{out}} = L(\Delta^{\text{in}}) \\ 0 & \quad\text{otherwise} \end{array} \right. ;$$

we obtain the following definition of $W^\diamond$.

Lemma 7 ($W^\diamond$ as a weak-key space)**.** *Let $(\mathbb{F}_2^n, \diamond)$ be an Abelian group such that $\mathcal{T} := \{x \mapsto x \diamond c \mid c \in \mathbb{F}_2^n\}$ is 2-elementary and regular. Let $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that: $\forall a \in \mathbb{F}_2^n, \quad H \circ \mathcal{T}_{H^{-1}(a)} \circ H^{-1} = T_a$. Then,*

$$W^\diamond = \{k \in \mathbb{F}_2^n \mid \forall\ \Delta^{\text{in}}, \Delta^{\text{out}} \in \mathbb{F}_2^n, \quad \Pr\left[\Delta^{\text{in}} \xrightarrow{T_k^H} \Delta^{\text{out}}\right] = \mathbf{1}_{\Delta^{\text{in}}}(\Delta^{\text{out}})\};$$

where $\mathbf{1}_x(y) = 1$ if $x = y$ and 0 otherwise.

The description of $W^\diamond$ in Lemma 7 explains the fact that it is indeed a weak-key space: whenever k belongs to $W^\diamond$, any differential transition through T_k^H is deterministic. Stated otherwise, such a transition only depends on the differences and is independent of the actual *values* of the considered pair.

While Lemma 7 clearly outlines the importance of the set $W^\diamond$ in such a study, its structure can still be clarified. This is the purpose of Lemma 8, which relies on the notion of *linear structures.* In the following, we denote by $D_\Delta F$ the derivative of a function $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ with respect to $\Delta \in \mathbb{F}_2^n$ for the usual addition law, *i.e.* $D_\Delta F(x) = F(x + \Delta) - F(x)$.

Definition 11 (Linear structure)**.** *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$. Let $\Delta \in \mathbb{F}_2^n$. The difference Δ is said to be a* linear structure *of F if the derivative $D_\Delta F$ is a constant function. The set of linear structures of F is denoted by $\mathcal{E}_F$, that is:*

$$\mathcal{E}_F := \{\Delta \in \mathbb{F}_2^n \mid \forall\ x \in \mathbb{F}_2^n, D_\Delta F(x) = F(\Delta) - F(0)\}.$$

Lemma 8 ($W^\diamond$ as linear space of H)**.** *Let $(\mathbb{F}_2^n, \diamond)$ be an Abelian group such that $\mathcal{T} := \{x \mapsto x \diamond c \mid c \in \mathbb{F}_2^n\}$ is 2-elementary and regular. Let $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that: $\forall a \in \mathbb{F}_2^n, \quad H \circ \mathcal{T}_{H^{-1}(a)} \circ H^{-1} = T_a$. Then, $W^\diamond = \mathcal{E}_H$.*

Proof. Starting from the first definition of $W^\diamond$, we observe that:

$$\begin{aligned} W^\diamond &= \{k \in \mathbb{F}_2^n \mid T_k = \mathcal{T}_k\} \\ &= \{k \in \mathbb{F}_2^n \mid T_k = H^{-1} \circ T_{H(k)} \circ H\} \\ &= \{k \in \mathbb{F}_2^n \mid H \circ T_k = T_{H(k)} \circ H\} \\ &= \{k \in \mathbb{F}_2^n \mid \Gamma_H(T_k, T_{H(k)}) = 2^n\} \\ &= \mathcal{E}_H; \end{aligned}$$

where we use the fact that commutations with constant addition corresponds to differential transitions. The last equality holds because the value of a constant derivative $D_k H$ is necessary $H(k) - H(0)$ but we have by construction that $H(0) = 0$. □

In the light of the following standard results, see for instance [28], the notion of linear structure is relatively well understood and gives new insight on this set of weak keys.

Lemma 9 (Standard properties of linear structures)**.** *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$. Then:*

1. *$\mathcal{E}_F$ is a linear space and the restriction of F to $\mathcal{E}_F$ is affine.*
2. *If F is bijective then $\mathcal{E}_{F^{-1}} = F(0) + F(\mathcal{E}_F)$ and in particular $\dim(\mathcal{E}_F) = \dim(\mathcal{E}_{F^{-1}})$.*
3. *[28, Theorem 3] Let $r = \dim(\mathcal{E}_F)$. Then there exists a linear bijection A such that $F \circ A$ can be decomposed as:*

$$F \circ A(x_1, \dots, x_n) = L(x_1, \dots, x_r) + \widetilde{F}(x_{r+1}, \dots, x_n);$$

where $L\colon \mathbb{F}_2^r \to \mathbb{F}_2^n$ is linear and $\widetilde{F}\colon \mathbb{F}_2^{n-r} \to \mathbb{F}_2^n$ satisfies $\mathcal{E}_{\widetilde{F}} = \{0\}$.

Corollary 10 (Dimension of $\mathcal{E}_F$)**.** *Let $F\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$. Then, $\deg(F) \leq n - \dim(\mathcal{E}_F)$. If F is not affine, then $\dim(\mathcal{E}_F) \leq n-2$. In particular, if $\dim(\mathcal{E}_F) = n-2$, we have $\deg(F) = 2$.*

Proof. The fact that for any F, it holds that $\deg(F) \leq n - \dim(\mathcal{E}_F)$ is a consequence of the third item of Lemma 9: F is equivalent to a function which can only be non-linear in its $n - \dim(\mathcal{E}_F)$ last variables. If $\dim(\mathcal{E}_F) = n-2$, then F is of degree at most 2. However, it must be of degree exactly 2 because the linear space of an affine function is the full space $\mathbb{F}_2^n$. □

Upper bound on $W^\diamond = \mathcal{E}_H$. This understanding of linear structures can then be applied to our case. Recall that in order to be an interesting a change of variables, H must be non-linear. In light of Corollary 10, it then satisfies $\dim(\mathcal{E}_H) \leq n-2$. The fact that $\dim(W^\diamond) \leq n-2$ that is stated in [11, Proposition 4.1] can then be seen as a consequence of this more general case.

Lower bound on $W^\diamond = \mathcal{E}_H$. Civino, Blondeau & Sala actually choose H with at least one non-trivial linear structure. This is guaranteed whenever $\mathcal{T}$ is a subgroup of the affine group. The following proposition is adapted from the work of Caranti, Dalla Volta & Sala [12] where it is stated in a more general context.

Proposition 18 (Non-trivial weak-key space [12])**.** *Let $\mathcal{T}$ be a 2-elementary regular subgroup of the affine group $\mathrm{AGL}(n, \mathbb{F}_2)$. Then $\mathcal{T} \cap T \neq \{I\}$, and thus $W^\diamond = \{k \in \mathbb{F}_2^n \mid T_k \in \mathcal{T} \cap T\} \neq \{0\}$.*

Proof. (Adapted from [12]) For any x, we can by hypothesis decompose $\mathcal{T}_x$ as $\mathcal{T}_x = T_x \circ L_x$ where L_x is linear. Because $\mathcal{T}_x^2 = I$, it holds for any $z \in \mathbb{F}_2^n$ that:

$$z = L_x(L_x(z) + x) + x = L_x^2(z) + L_x(x) + x.$$

In particular with $z = 0$, we observe that $L_x(x) = x$. Therefore, we also get $L_x^2 = I$. Let $x, y \in \mathbb{F}_2^n$. Then:

$$\begin{aligned}
\mathcal{T}_x T_y \mathcal{T}_x &= T_x L_x T_y T_x L_x \\
&= T_x L_x T_{y+x} L_x \\
&= T_x T_{L_x(y+x)} L_x L_x \\
&= T_{x+L_x(y+x)} \\
&= T_{L_x(x)+L_x(y+x)} \\
&= T_{L_x(y)}; \qquad (24)
\end{aligned}$$

where we successively use the decomposition of $\mathcal{T}_x$, the fact that $T_yT_x = T_{y+x}$, the fact that $L_xT_{y+x} = T_{L_x(y+x)}L_x$ because L_x is linear, then $L_x^2 = I$ and finally $x = L_x(x)$ and the linearity of L_x again. In particular, we observe that $\mathcal{T}_xT_y\mathcal{T}_x \in T$ for any $x, y \in \mathbb{F}_2^n$. This implies that the function $F\colon \mathcal{T} \times T \to T$ that is defined as:

$$F\colon (\mathcal{T}_x, T_y) \mapsto \mathcal{T}_xT_y\mathcal{T}_x;$$

is actually well-defined. It corresponds to the action by conjugation of $\mathcal{T}$ on the set T as $\mathcal{T}_x^{-1} = \mathcal{T}_x$ for any x in our case. As for any action of a p-group H on a set Z, the orbit-stabilizer theorem states that the number of elements that are H-invariants is equal to $|Z|$ modulo p. In our case, the set Z of $\mathcal{T}$-invariants is defined by:

$$Z := \{T_y \mid \mathcal{T}_xT_y\mathcal{T}_x = T_y \forall x \in \mathbb{F}_2^n\}$$

and it must be of even cardinality. As it contains $T_0 = I$, it must contain at least a non-trivial element. To conclude, we now show that Z is actually equal to $\mathcal{T} \cap T$. Indeed, we have:

$$\begin{aligned} Z &= \{T_y \mid T_{L_x(y)} = T_y,\ \forall\ x \in \mathbb{F}_2^n\} \\ &= \{T_y \mid L_x(y) + y = 0,\ \forall\ x \in \mathbb{F}_2^n\} \\ &= \{T_y \mid L_y(x) + x = 0,\ \forall\ x \in \mathbb{F}_2^n\} \\ &= \{T_y \mid L_y = I\} \\ &= \{T_y \mid \mathcal{T}_y = T_y\} \\ &= \mathcal{T} \cap T. \end{aligned}$$

The third equality holds because for any $x, y \in \mathbb{F}_2^n$, we have:

$$L_x(y) + y = \mathcal{T}_x(y) + y + x = \mathcal{T}_y(x) + y + x = L_y(x) + x.$$

□

Corollary 11. *Let $H\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ be bijective. Let us suppose that $H \circ T_c \circ H^{-1}$ is affine for any $c \in \mathbb{F}_2^n$. Then* $\dim(\mathcal{E}_H) \geq 1$.

The case $\dim(W^\diamond) = \dim(\mathcal{E}_H) = n - 2$. Civino, Blondeau & Sala [15] finally studies the case $\dim(W^\diamond) = n - 2$ in more detail. Let H be a bijection such that $\dim(\mathcal{E}_H) = n - 2$. Because of the third item of Lemma 9, H can be written as $H = H' \circ A^{-1}$, where A is a linear bijection and $H'\colon \mathbb{F}_2^n \to \mathbb{F}_2^n$ a function whose components are made only of constant terms, affine terms and the quadratic term x_1x_2. Note that x_1x_2 must appear in at least one coordinate, but not necessarily all of them. In particular there exists a bijective linear mapping B such that $B \circ H'$ has a single coordinate containing x_1x_2 while all others are affine. However, the differential properties of F^H and the ones of $F^{B\circ H}$ are identical as $F^{B\circ H} = B \circ F^H \circ B^{-1}$. This implies that the choices of change of variables H made by the authors of [15, 10] are similar to the choices made by the authors of [4, 3] for the analysis of conjugate Midori.

Note that, because of Lemma 9, the specific case $\dim(\mathcal{E}_H) = n-2$ implies that $\dim(\mathcal{E}_{H^{-1}}) = n-2$, and due to Corollary 10, both H and H^{-1} are quadratic.

Examples where both $\mathcal{T} \subseteq \mathrm{AGL}(n, \mathbb{F}_2)$ and $\dim(W^\diamond) = n - 2$ are given in [15, 10]. Another one based on [4] is given in Section 6.4.

6.3.3 About the Weak-Key Space of [4].

As shown in Lemma 7, whenever a key belongs to $W^{\diamond}$, the actual key used does not matter anymore as the behavior is deterministic and independent of the key, and the actual differences. This enables to launch *any kind of differential attack* as it is done in the classical way.

However, contrary to the standard case, the fraction of the keys cannot exceed one quarter of the key space. In the setting where the change of variable is a parallel application of a non-linear change of variables of the size of the S-box, this fraction is in practice way smaller.

But $W^{\diamond}$ is a conservative choice of weak-key space in the sense that it enables any differential attack. On the contrary, if we are instead interested in a *specific* attack taking advantage of some specific transition, we can hope for a bigger set of weak keys. Let $\Delta^{\text{in}}, \Delta^{\text{out}} \in \mathbb{F}_2^n$, and let us consider $\Delta^{\text{in}} \xrightarrow{T_k^H} \Delta^{\text{out}}$. In that case, we actually want to consider the set $W(\Delta^{\text{in}}, \Delta^{\text{out}})$ that is defined as:

$$\begin{aligned} W(\Delta^{\text{in}}, \Delta^{\text{out}}) &:= \{k \in \mathbb{F}_2^n \mid \Pr\left[\Delta^{\text{in}} \xrightarrow{T_k^H} \Delta^{\text{out}}\right] = 1\} \\ &= \{k \in \mathbb{F}_2^n \mid \Delta^{\text{in}} \in \mathcal{E}_{T_k^H},\ \Delta^{\text{out}} = D_{\Delta^{\text{in}}} T_k^H(0)\}. \end{aligned}$$

In other words, we would like to consider linear structures shared by multiple T_k^H, for which the corresponding constant derivatives are equal. A direct corollary of Lemma 7 is that for a given $\Delta \in \mathbb{F}_2^n$ we have:

$$\mathcal{E}_H \subseteq W(\Delta, \Delta).$$

However, in practice $\mathcal{E}_H$ can be a strict subset of $W(\Delta, \Delta)$. An example is given in Section 6.4. Furthermore, while transitions $\Delta \xrightarrow{T_k^H} \Delta$ with probability 1 imitate the standard differential case for the bitwise addition, there might exist $\Delta^{\text{in}} \neq \Delta^{\text{out}}$ such that $W(\Delta^{\text{in}}, \Delta^{\text{out}}) \neq \emptyset$. Therefore, transitions $\Delta^{\text{in}} \xrightarrow{T_k^H} \Delta^{\text{out}}$ with probability 1 can also be considered in a weak-key setting. This is in particular important in the case where T_k^H is affine for any k. In that specific case, $W(\Delta^{\text{in}}, \Delta^{\text{out}})$ becomes

$$W(\Delta^{\text{in}}, \Delta^{\text{out}}) = \{k \in \mathbb{F}_2^n \mid D_{\Delta^{\text{in}}} T_k^H(0) = \Delta^{\text{out}}\};$$

because any derivative of any T_k^H is constant. This also means that for a fixed $\Delta^{\text{in}} \in \mathbb{F}_2^n$, the sets $W(\Delta^{\text{in}}, \Delta^{\text{out}})$ for all $\Delta^{\text{out}} \in \mathbb{F}_2^n$ partition the set of round keys:

$$\bigsqcup_{\Delta^{\text{out}} \in \mathbb{F}_2^n} W(\Delta^{\text{in}}, \Delta^{\text{out}}) = \{k \in \mathbb{F}_2^n\} = \mathbb{F}_2^n.$$

6.4 Complementing Some Examples from [15, 10, 4, 3].

6.4.1 A Conjugate of the Block Cipher Midori.

Let us take a closer look at the analysis of conjugate Midori that is addressed in [4, Appendix A] and in more detail in [3]. As a reminder of these works, there exists a change of variables $H\colon \mathbb{F}_2^4 \to \mathbb{F}_2^4$ for which $\Delta \xrightarrow{S^H} \Delta$ holds with probability 1 for $\Delta = \texttt{0xd}$. This can be easily verified from the look-up tables given in Table 1.

Table 1: A specific change of variables for the S-box of Midori64.

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$H(x)$	0	3	4	7	2	1	6	5	8	a	c	e	b	9	f	d
$S^H(x)$	b	e	f	c	9	5	d	7	8	4	a	0	3	6	1	2
$S^H(x+\Delta)$	6	3	2	1	4	8	0	a	5	9	7	d	e	b	c	f

Furthermore the 16-time parallel application of H, that we denote by $\mathcal{H}\colon (\mathbb{F}_2^4)^{16} \to (\mathbb{F}_2^4)^{16}$, naturally satisfies $\nabla \xrightarrow{\mathcal{S}^{\mathcal{H}}} \nabla$ with probability 1 where $\mathcal{S}\colon (\mathbb{F}_2^4)^{16} \to (\mathbb{F}_2^4)^{16}$ is the full S-box layer, and where $\nabla = (\Delta, \dots, \Delta)$. It can also be verified that $\nabla \xrightarrow{\mathcal{L}} \nabla$ holds with probability 1 for the linear layer of Midori. Let us then take a closer look at the constant addition. The ANF of H and of $T^H\colon (x,k) \mapsto T_k^H(x)$ are given by:

$$H(x) = \begin{pmatrix} x_1x_4 + x_1 + x_3x_4 \\ x_1 + x_3 \\ x_2 \\ x_4 \end{pmatrix}, \quad T_k^H(x) = \begin{pmatrix} x_1 + x_2k_4 + x_4k_1 + x_4k_3 + k_1k_4 + k_1 + k_3k_4 \\ x_2 + k_1 + k_3 \\ x_3 + k_2 \\ x_4 + k_4 \end{pmatrix};$$

where the input bits are listed from the LSB x_1 to the MSB x_4 and the output coordinates are listed top to bottom from the least significant one to the most significant one.

In particular, we are here studying properties of conjugate functions of the form F^H. By Proposition 17, this is equivalent to studying the $\diamond$-differential properties for the law $\diamond$ whose group of translations is $\mathcal{T} = H^{-1}TH$. For this reason, we can also look at all $T_k^{H^{-1}}$. Their ANFs, as well as the ANF of $\diamond$ are given by:

$$\forall x, k \in \mathbb{F}_2^4, \quad x \diamond k := T_{H(k)}^{H^{-1}}(x) = \begin{pmatrix} x_1 + k_1 + (x_1 + x_3)k_4 + x_4(k_1 + k_3) \\ x_2 + k_2 \\ x_3 + k_3 + (x_1 + x_3)k_4 + x_4(k_1 + k_3) \\ x_4 + k_4 \end{pmatrix}. \tag{25}$$

From this ANF, we can easily observe that:

$$W^\diamond = \{k \in \mathbb{F}_2^4 \mid T_k = \mathcal{T}_k\} = \{k \in \mathbb{F}_2^4 \mid k_4 = 0, k_1 = k_3\} = \langle \texttt{0x2}, \texttt{0x5} \rangle .$$

Because of Lemma 8, we can determine $W^\diamond$ without this explicit formula for $\diamond$. Indeed, it suffices to look at the linear structures of H. From the ANF of H, it is clear that:

$$\forall x, \Delta \in \mathbb{F}_2^4, \quad D_\Delta H(x) = \begin{pmatrix} \Delta_1 + x_4(\Delta_1 + \Delta_3) + \Delta_4(x_1 + x_3) + \Delta_4(\Delta_1 + \Delta_3) \\ \Delta_1 + \Delta_3 \\ \Delta_2 \\ \Delta_4 \end{pmatrix}.$$

The derivative $D_\Delta H$ is therefore constant if and only if $\Delta_1 = \Delta_3$ and $\Delta_4 = 0$, and, as expected, the same set $W^\diamond$ is obtained.

However, when focusing on the specific transition $\Delta \xrightarrow{T_k^H} \Delta$ for a given $\Delta \in \mathbb{F}_2^4$, we can compute $W(\Delta, \Delta)$. Let $k \in \mathbb{F}_2^4$. Because T_k^H is affine, its derivative is constant and equal to:

$$\forall\, x \in \mathbb{F}_2^4, \quad D_\Delta T_k^H(x) := \begin{pmatrix} \Delta_1 + \Delta_2 k_4 + \Delta_4 k_1 + \Delta_4 k_3 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \end{pmatrix}.$$

In the specific case where $\Delta = \texttt{0xd} = \texttt{0b1101}$, we obtain:

$$\forall\, x \in \mathbb{F}_2^4, \quad D_\Delta T_k^H(x) = \begin{pmatrix} 1 + k_1 + k_3 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

We then conclude in that case that $D_\Delta T_k^H = \Delta$ if and only if $k_1 + k_3 = 0$ and therefore $W(\Delta, \Delta)$ is equal to:

$$W(\Delta, \Delta) = \{k \in \mathbb{F}_2^4 \mid k_1 + k_3 = 0\} = \langle \texttt{0x2}, \texttt{0x5}, \texttt{0x8} \rangle\,.$$

In particular, $W(\Delta, \Delta)$ is strictly bigger than $W^\diamond$ and the differential trail $\nabla \xrightarrow{(F_k^{(0)})^{\mathcal{H}}} \nabla \to \cdots \xrightarrow{(F_k^{(R-1)})^{\mathcal{H}}} \nabla$ holds with probability 1 if all nibbles of all rounds keys and round constants belong to $W(\Delta, \Delta)$.

Finally in this case, because of Proposition 17, the differential transition $\Delta \xrightarrow{S^H} \Delta$ that holds with probability 1, can equivalently be considered as a probability-1 $\diamond$-differential transition $H^{-1}(\Delta) \to H^{-1}(\Delta)$ for the law $\diamond$ given above, or as probability-1 commutation with the affine function $A := T_\Delta^{H^{-1}}$ where $\Delta = \texttt{0xd}$. Its ANF can easily be deduced from Eq. (25) using $k = H^{-1}(\Delta) = \texttt{0xf} = \texttt{0b1111}$ and is given below:

$$A(x) := \begin{pmatrix} x_3 + 1 \\ x_2 + 1 \\ x_1 + 1 \\ x_4 + 1 \end{pmatrix}. \tag{26}$$

6.4.2 The Toy Cipher of [15, 10].

The toy cipher of [15] In [15], a block cipher with a 15-bit state is proposed to illustrate $\diamond$-differential cryptanalysis. It has an standard SPN structure where the S-box layer is the 5-time parallel application of a single 3-bit S-box $S\colon \mathbb{F}_2^3 \to \mathbb{F}_2^3$.

The look-up table of S is given in Table 2.

Table 2: The S-box used in [15] and a suitable change of variables H.

x	0	1	2	3	4	5	6	7
$S(x)$	0	6	2	1	5	7	4	3

x	0	1	2	3	4	5	6	7
$H(x)$	0	1	2	3	6	7	5	4

In order to study this S-box, the authors introduced the law $\diamond\colon \mathbb{F}_2^3 \times \mathbb{F}_2^3 \to \mathbb{F}_2^3$ that is defined as:

$$\forall x, y \in \mathbb{F}_2^3, \quad x \diamond y := \begin{pmatrix} x_1 + y_1 + x_2 y_3 + x_3 y_2 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}.$$

We can computationally verify that this S-box is APN, however it has a probability-1 $\diamond$-differential transition $\Delta^{\text{in}} \underset{\diamond}{\xrightarrow{S}} \Delta^{\text{out}}$, where $\Delta^{\text{in}} = \texttt{0x6}, \Delta^{\text{out}} = \texttt{0x4}$. This can be equivalently understood as probability-1 commutative property, or as a probability-1 differential of S^H for some H.

We found out by hand that H defined by

$$\forall x \in \mathbb{F}_2^3, \quad H(x) := \begin{pmatrix} x_1 + x_2 x_3 \\ x_2 + x_3 \\ x_3 \end{pmatrix}, \quad H^{-1}(x) = \begin{pmatrix} x_1 + x_3 + x_2 x_3 \\ x_2 + x_3 \\ x_3 \end{pmatrix}$$

satisfies the equality $x \diamond y = T_{H(y)}^{H^{-1}}(x)$ for any x, y. Because of Lemma 6, any isomorphism between $(\mathbb{F}_2^3, \diamond)$ and $(\mathbb{F}_2^3, +)$ actually works. We focus here on this arbitrary case. In particular, we consider A and B that we define as:

$$A := T_{H(\Delta^{\text{in}})}^{H^{-1}} = x \diamond \texttt{0x6} = \begin{pmatrix} x_1 + x_2 + x_3 \\ x_2 + 1 \\ x_3 + 1 \end{pmatrix}, \quad \text{and}$$

$$B := T_{H(\Delta^{\text{out}})}^{H^{-1}} = x \diamond \texttt{0x4} = \begin{pmatrix} x_1 + x_2 \\ x_2 \\ x_3 + 1 \end{pmatrix}.$$

By Proposition 17, it holds that $S \circ A = B \circ S$. This can be verified from the ANF of S that is given below, together with the one of S^H.

$$\forall x \in \mathbb{F}_2^3, \quad S(x) := \begin{pmatrix} x_1 x_2 + x_2 x_3 + x_3 \\ x_1 + x_2 x_3 + x_2 \\ x_1 x_2 + x_1 x_3 + x_1 + x_3 \end{pmatrix}, \quad S^H(x) = \begin{pmatrix} x_1 + x_3 \\ x_1 x_2 + x_2 x_3 + x_2 + x_3 \\ x_1 x_2 + x_1 + x_2 x_3 \end{pmatrix}.$$

We also easily observe that the differential $\texttt{0x5} \xrightarrow{S^H} \texttt{0x6}$ holds with probability 1. This is again due to Proposition 17, because $H(\Delta^{\text{in}}) = \texttt{0x5}$ and $H(\Delta^{\text{out}}) = \texttt{0x6}$.

The toy cipher of [10]. The recent work of Calderini, Civino & Invernizzi [10] deals with the resistance against $\diamond$-differential cryptanalysis of S-boxes which are optimal with respect to standard differential cryptanalysis. They in particular show that such S-boxes have no reason to be optimal for other laws $\diamond$, and among an affine equivalence class, two distinct S-boxes can have two distinct uniformities with respect to $\diamond$.

To illustrate their work, they build a similar SPN than before, this time with a 16-bit block size that is decomposed into 4 cells of four bits. The used 4-bit S-box S is given in Table 3.

Its differential uniformity is 4, but, for a law $\diamond$ built in the same way as before, it has a probability-1 $\diamond$-differential transition $\Delta^{\text{in}} \underset{\diamond}{\xrightarrow{S}} \Delta^{\text{out}}$, where $\Delta^{\text{in}} = \texttt{0x7}, \Delta^{\text{out}} = \texttt{0x6}$. The law is defined[3] as:

[3] A look-up table can be found in the slides of the presentations at WCC 2024.

Table 3: The S-box used in [10].

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$S(x)$	0	e	b	1	7	c	9	6	d	3	4	f	2	8	a	5

$$\forall x, y \in \mathbb{F}_2^4, \quad x \diamond y := \begin{pmatrix} x_1 + y_1 + x_3 y_4 + x_4 y_3 \\ x_2 + y_2 \\ x_3 + y_3 \\ x_4 + y_4 \end{pmatrix} .$$

This corresponds to a commutation with probability 1 from $A = \mathcal{T}_{\texttt{0x7}}$ to $B = \mathcal{T}_{\texttt{0x6}}$, or a probability-1 differential $H(\Delta^{\text{in}}) \xrightarrow{S^H} H(\Delta^{\text{in}})$ for a suitable H. We can for instance use:

$$H := \begin{pmatrix} x_1 + x_3 x_4 + x_4 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} .$$

Acknowledgments. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA – 390781972, by the ERC advanced grant 101097056 (SYMTRUST) and the ERC starting grant 101041545 (ReSCALE), and by ANR (French National Research Agency) grant ANR-20-CE48-0017 (SELECT).

References

[1] T. Ashur and Y. Liu. Rotational cryptanalysis in the presence of constants. *IACR Trans. Symmetric Cryptol.*, 2016(1):57–70, 2016.

[2] D. Bartoli, L. Kölsch, and G. Micheli. Differential biases, c-differential uniformity, and their relation to differential attacks. *CoRR*, abs/2208.03884, 2022.

[3] J. Baudrin. Phd thesis, Sorbonne Université, Dec. 2024. Submitted.

[4] J. Baudrin, P. Felke, G. Leander, P. Neumann, L. Perrin, and L. Stennes. Commutative cryptanalysis made practical. *IACR Trans. Symmetric Cryptol.*, 2023(4):299–329, 2023.

[5] C. Beierle, M. Brinkmann, and G. Leander. Linearly self-equivalent APN permutations in small dimension. *IEEE Trans. Inf. Theory*, 67(7):4863–4875, 2021.

[6] C. Beierle, A. Canteaut, and G. Leander. Nonlinear approximations in cryptanalysis revisited. *IACR Trans. Symmetric Cryptol.*, 2018(4):80–101, 2018.

[7] C. Beierle, P. Felke, G. Leander, P. Neumann, and L. Stennes. On perfect linear approximations and differentials over two-round SPNs. In H. Handschuh and A. Lysyanskaya, editors, *Advances in Cryptology - CRYPTO 2023 - 43rd Annual International Cryptology Conference, CRYPTO*

2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III, volume 14083 of *Lecture Notes in Computer Science*, pages 209–239. Springer, 2023.

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

[9] A. Bors and Q. Wang. Coset-wise affine functions and cycle types of complete mappings. *Finite Fields Their Appl.*, 83:102088, 2022.

[10] M. Calderini, R. Civino, and R. Invernizzi. Optimal s-boxes against alternative operations. *CoRR*, abs/2403.20059, 2024.

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

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

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

[14] C. Carlet, P. Charpin, and V. A. Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Des. Codes Cryptogr.*, 15(2):125–156, 1998.

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

[16] R. Crowell. Graphs of linear transformations over finite fields. *J. Soc. Indust. Appl. Math.*, 10(1):103–112, 1962.

[17] J. Daemen. *Cipher and hash function design strategies based on linear and differential cryptanalysis.* PhD thesis, Doctoral Dissertation, March 1995, KU Leuven, 1995.

[18] J. Daemen and V. Rijmen. *The Design of Rijndael - The Advanced Encryption Standard (AES), Second Edition.* Information Security and Cryptography. Springer, 2020.

[19] J. D. Dixon. Maximal Abelian subgroups of the symmetric groups. *Can. J. Math.*, 23(3):426–438, 1971.

[20] P. Ellingsen, P. Felke, C. Riera, P. Stanica, and A. Tkachenko. C-differentials, multiplicative uniformity, and (almost) perfect c-nonlinearity. *IEEE Trans. Inf. Theory*, 66(9):5781–5789, 2020.

[21] B. Elspas. The theory of autonomous linear sequential networks. *IRE Transactions on Circuit Theory*, 6(1):45–60, 1959.

[22] ETSI/SAGE. Specification of the 3GPP confidentiality and integrity algorithms 128-EEA3 & 128-EIA3. document 4: Design and evaluation report. Technical report, available online at `https://www.gsma.com/about-us/wp-content/uploads/2014/12/EEA3_EIA3_Design_Evaluation_v2_0.pdf`, 2011.

[23] H. Fripertinger. Cycle indices of linear, affine, and projective groups. *Linear Algebra Its Appl.*, 263:133–156, 1997.

[24] F. R. Gantmacher. *The theory of matrices. Vols. 1, 2.* Translated by K. A. Hirsch. Chelsea Publishing Co., New York, 1959.

[25] V. Grosso, G. Leurent, F.-X. Standaert, K. Varici, F. Durvaux, L. Gaspar, and S. Kerckhof. SCREAM & iSCREAM, side-channel resistant authenticated encryption with masking. v1 of a CAESAR submission, available on line: `http://competitions.cr.yp.to/round1/screamv1.pdf`, 2014.

[26] S. Guest, J. Morris, C. E. Praeger, and P. Spiga. Affine transformations of finite vector spaces with large orders or few cycles. *J. Pure Appl. Algebra*, 219(2):308–330, 2015.

[27] D. Khovratovich and I. Nikolic. Rotational cryptanalysis of ARX. In S. Hong and T. Iwata, editors, *Fast Software Encryption, 17th International Workshop, FSE 2010, Seoul, Korea, February 7-10, 2010, Revised Selected Papers*, volume 6147 of *Lecture Notes in Computer Science*, pages 333–346. Springer, 2010.

[28] X. Lai. Additive and linear structures of cryptographic functions. In B. Preneel, editor, *Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings*, volume 1008 of *Lecture Notes in Computer Science*, pages 75–85. Springer, 1994.

[29] X. Lai, J. L. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In D. W. Davies, editor, *Advances in Cryptology - EUROCRYPT '91, Workshop on the Theory and Application of of Cryptographic Techniques, Brighton, UK, April 8-11, 1991, Proceedings*, volume 547 of *Lecture Notes in Computer Science*, pages 17–38. Springer, 1991.

[30] R. Lidl and H. Niederreiter. *Finite Fields.* Cambridge University Press, 1997.

[31] T. Liu, S. Tessaro, and V. Vaikuntanathan. The t-wise independence of substitution-permutation networks. In T. Malkin and C. Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part IV*, volume 12828 of *Lecture Notes in Computer Science*, pages 454–483. Springer, 2021.

[32] T. Liu, S. Tessaro, and V. Vaikuntanathan. The t-wise independence of substitution-permutation networks. *IACR Cryptol. ePrint Arch.*, page 507, 2021.

[33] B. Mennink. Modeling security. In *Symmetric Cryptography 1*, chapter 10, pages 135–146. John Wiley & Sons, Ltd, 2024.

[34] S. Mesnager, B. Mandal, and M. Msahli. Survey on recent trends towards generalized differential and boomerang uniformities. *Cryptogr. Commun.*, 14(4):691–735, 2022.

[35] K. Nyberg. Differentially uniform mappings for cryptography. In T. Helleseth, editor, *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993, Proceedings*, volume 765 of *Lecture Notes in Computer Science*, pages 55–64. Springer, 1993.

[36] A. Pott. Almost perfect and planar functions. *Des. Codes Cryptogr.*, 78(1):141–195, 2016.

[37] D. A. Wagner. Towards a unifying view of block cipher cryptanalysis. In B. K. Roy and W. Meier, editors, *Fast Software Encryption, 11th International Workshop, FSE 2004, Delhi, India, February 5-7, 2004, Revised Papers*, volume 3017 of *Lecture Notes in Computer Science*, pages 16–33. Springer, 2004.

[38] K. Wang. Transition graphs of affine transformation on vector spaces over finite fields. *Journal of the Franklin Institute*, 283(1):55–72, 1967.