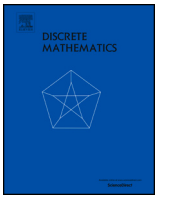
Discrete Mathematics 349 (2026) 114870



Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc



Optimal s-boxes against alternative operations and linear propagation



Marco Calderini [a], Roberto Civino [b,*], Riccardo Invernizzi [c]

[a] University of Trento, Italy
[b] University of L'Aquila, Italy
[c] COSIC, KU Leuven, Belgium

ARTICLE INFO

Article history:
Received 6 March 2025
Received in revised form 9 October 2025
Accepted 27 October 2025
Available online 30 October 2025

Keywords:
Differential cryptanalysis
Alternative operations
4-bit s-boxes
Commutative cryptanalysis

ABSTRACT

Civino et al. (2019) have shown how some diffusion layers can expose a Substitution-Permutation Network to vulnerability from differential cryptanalysis when employing alternative operations coming from groups isomorphic to the translation group on the message space. In this study, we present a classification of diffusion layers that exhibit linearity with respect to certain *parallel* alternative operations, enabling the possibility of an alternative differential attack simultaneously targeting all the s-boxes within the block. Furthermore, we investigate the differential behaviour with respect to alternative operations for all classes of optimal 4-bit s-boxes, as defined by Leander and Poschmann (2007). Our examination reveals that certain classes contain weak permutations w.r.t. alternative differential attacks. Finally, we leverage these vulnerabilities to execute a series of experiments showing the effectiveness of the cryptanalysis performed with a parallel alternative operation compared to the classical one.

© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction and preliminaries

Differential cryptanalysis, originally introduced by Biham and Shamir in the late 1980s [4] and subsequently generalised [6,9,18,23], has become one of the cornerstones for evaluating the robustness of various symmetric primitives. The fundamental premise of differential cryptanalysis is that analysing the differences (differentials) between pairs of plaintexts and the corresponding ciphertexts can unveil undesired biases. While differentials can be calculated with respect to any difference operator, regardless of which operation is responsible for performing the sum with the round key during encryption, it is usual for the two operations to coincide. For this reason, classical differential cryptanalysis of a cipher in which the key is xor-ed to the state is typically performed by studying the distribution of xor-differentials, whose propagation is traditionally prevented by the combined action of the linear diffusion layer and the s-box layer. In particular, s-boxes are pivotal for ensuring the security of almost all contemporary block ciphers, serving as the primary nonlinear component within the cipher, particularly in the case of Substitution-Permutation Networks (SPNs). Equally relevant, the efficiency of a cipher is significantly influenced by the size of the s-boxes. In practical scenarios, s-boxes typically have a size of 4 or 8 bits, with 4 being the most popular choice for ciphers designed to operate on power-constrained devices [2,5,8,21]. It is clear that the selection of appropriate s-boxes is critical to fortify the cipher against various types of attacks. In this sense, Leander

* Corresponding author.
E-mail addresses: marco.calderini@unitn.it (M. Calderini), roberto.civino@univaq.it (R. Civino), riccardo.invernizzi@esat.kuleuven.be (R. Invernizzi).

https://doi.org/10.1016/j.disc.2025.114870
0012-365X/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and Poschmann have classified 4-bit s-boxes which are optimal w.r.t. standard criteria that guarantee poor propagation of xor-differentials [19].

A recent line of research is focused on the study of alternative operations for the differential cryptanalysis of xor-based ciphers [3,10,11,15,22]. These new operations define new differentials to be analysed. Within this approach, a large class of possible alternative operations has been studied, all of which have in common that they are induced by a group of translations isomorphic to the group of translations acting on the message space by means of the xor addition with the key. In the context of an SPN, where the encrypted message is generated by iterating through a sequence of s-box layers, (xor)-linear diffusion, and xor-based key addition layers, altering the differential operator yields a dual impact. On one hand, it is highly probable that differentials cross the s-box layer more effectively, given that its nonlinearity is maximised with respect to xor. On the other hand, differentials do not deterministically propagate through the diffusion layer, as observed in classical scenarios. This pivotal limitation effectively restricts the success of the attack only to cases where the target layer is linear not only concerning xor but also with respect to the operation under consideration for computing differentials. Moreover only a portion of round keys, called *weak*, will let alternative differentials propagate deterministically through the xor-based key-addition layer, while all others will require a probabilistic analysis.

Related works. The approach of differential cryptanalysis with alternative operations can be framed within the broader context of commutative cryptanalysis, introduced by Wagner [24], of which differential cryptanalysis and some of its variations can be seen as particular cases.

A first successful attempt based on the study of the alternative differential properties of an xor-based toy cipher of the SPN family has shown that it is possible to highlight a bias in the distribution of the differences calculated compared to an alternative operation which is instead not detectable by means of the standard xor-differential-based approach [15]. The target cipher featured five 3-bit s-boxes and the operation used to perform the attack acted as the xor on the last four s-boxes, while on the first one matched with one of the alternative sums defined by Calderini et al. [10], coming from another translation group. The advantage of employing an alternative operation in this case was only derived from the benefit induced by a single s-box. In a more recent experimental approach [11], we showed that better results in a similar context can be obtained using an *alternative parallel operation*, in which every s-box can be targeted. In this case, the diffusion layer of the cipher was determined through an algorithm, ensuring that it adheres to the constraint of linearity with respect to both xor and the target operation.

The experiments mentioned above rely on the assumption that the translation group $T_\circ$ defining the new operation $\circ$ consists of functions that are affine with respect to the classical xor operation, that is denoted in this paper simply by $+$. In other words, each translation of the new operation can be expressed as an affine transformation over the usual vector space structure defined by xor. We also assume the dual condition, namely that the classical translations are affine with respect to $\circ$, which ensures that the weak-key space has a controlled and predictable dimension. A similar approach is used by Baudrin et al. [3] to find probability-one distinguishers in both a modified `Midori` [2] with *weakened* round constants and in `Scream`. Their analysis focuses on finding affine maps $A, B \in \mathrm{AGL}(+)$ such that for sufficiently many encryption function holds

$$E_k(A(x)) = B(E_k(x)),$$

where A and B are diagonal matrices with s-box-sized blocks.

Our contribution. This paper establishes a general result for SPNs, characterising xor-linear maps that are simultaneously linear with respect to a fixed parallel alternative operation of affine type (Sec. 2). Specifically, we consider a set of affine maps $\tau_a \subset \mathrm{AGL}(+)$, represented by diagonal matrices with s-box-sized blocks, where a ranges over the message space, and forming a translation group under an operation $\circ$. We derive conditions on the cipher's diffusion layer to ensure that, for each a and b,

$$E_k(\tau_a(x)) = \tau_b(E_k(x)),$$

which implies that

$$E_k(x) \circ E_k(x \circ a) = b$$

with high probability for a set of sufficiently many keys, i.e. weak keys. This finding enables the execution of a differential attack wherein each s-box affected by a nontrivial differential contributes to the final differential probability with increased efficacy compared to the conventional xor differentials. Additionally, differentials propagate deterministically through the linear layer in this scenario. Moreover, we examine all possible alternative operations on 4 bits and investigate the differential properties of the optimal 4-bit s-boxes classes, as defined in the work of Leander and Poschmann [19] (a comparable methodology, albeit in the context of modular addition, was recently employed by Zajac and Jókay [25]). Our analysis shows that each class comprises potentially weak permutations (Sec. 3). When coupled with a diffusion layer as described earlier, these permutations have the potential to render the cipher susceptible to differential attacks with alternative operations. To substantiate our findings, we conclude the paper by presenting experimental results on a family of toy SPNs (Sec. 4).

1.1. Notation

Let V be an n-dimensional vector space over $\mathbb{F}_2$ which represents the message space. We write $V = V_1 \oplus V_2 \oplus \cdots \oplus V_b$, where each V_j is isomorphic to a vector space B such that $\dim(B) = s$ on which every s-box acts. Therefore we have $n = sb$. We denote by $\{e_i\}_{i=1}^n$ the canonical basis of V. If G is any finite group acting on V, for each $g \in G$ and $v \in V$ we denote the action of g on v as vg, i.e. we use postfix notation for every function evaluation. We denote by $\mathrm{Sym}(V)$ the symmetric group acting on V, i.e. the group of all permutations on the message space, by $\mathrm{GL}(V,+)$ the group of linear transformations, and by $\mathrm{AGL}(V,+)$ the group of affine permutations. The identity matrix of size l is denoted by $\mathbb{1}_l$ and the zero matrix of size $l \times h$ is denoted by $\mathbb{0}_{l,h}$, or simply $\mathbb{0}_l$ if $l = h$. We finally denote by T_+ the group of translations on V, i.e.

$$T_+ := \{\sigma_a : x \mapsto x + a \mid a \in V\} < \mathrm{Sym}(V).$$

We remind that the translation σ_k acts on a vector x in the same way the key-addition layer of an SPN acts xor-ing the round key k to the message x, i.e. $x\sigma_k = x + k$.

1.2. Preliminaries on alternative operations

An alternative operation on V can be defined given any 2-elementary abelian regular subgroup $T < \mathrm{AGL}(V,+)$, that we can write as $T = \{\tau_a \mid a \in V\}$, where τ_a is the unique element in T which maps 0 into a. Consequently, for all $a, b \in \mathrm{V}$, we can define $a \circ b := a\tau_b$, resulting in $(V, \circ)$ forming an additive group. The operation $\circ$ induces a vector space structure on V, with the corresponding group of translation being $T_\circ = T$. Additionally, for each $a \in V$, there exists $M_a \in \mathrm{GL}(V,+)$ such that $\tau_a = M_a\sigma_a$, meaning that for every $x \in V$,

$$x \circ a = x\tau_a = xM_a + a.$$

It is also assumed throughout that $T_+ < \mathrm{AGL}(V, \circ)$, where $\mathrm{AGL}(V, \circ)$ is the normaliser in $\mathrm{Sym}(V)$ of $T_\circ$ (i.e. the group of affine permutations w.r.t. $\circ$). This crucial technical assumption renders the key-addition layer an affine operator concerning the new operation, enabling the prediction of how the key addition affects the differentials with a reasonable probability. Further details on this aspect, which may not be directly relevant to the scope of the current paper, can be found in Civino et al. [15]. In this context, we define the *weak keys subspace* as

$$W_\circ := \{a \mid a \in V, \sigma_a = \tau_a\} = \{k \mid k \in V,\ \forall x \in V\ x \circ k = x + k\}.$$

$W_\circ$ is a vector subspace of both $(V,+)$ and $(V,\circ)$. It is known [10,12] that $W_\circ$ is nonempty and that

$$2 - (n \bmod 2) \leq \dim(W_\circ) \leq n - 2. \tag{1}$$

Moreover, up to conjugation, we can always assume $W_\circ$ to be the span of the last d canonical vectors of V [10]. This allows to represent the new sum in a canonical way: for each $a \in V$ there exists a matrix $E_a \in \mathbb{F}_2^{(n-d)\times d}$ such that

$$M_a = \begin{pmatrix} \mathbb{1}_{n-d} & E_a \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix}. \tag{2}$$

Fixing such an operation as above is therefore equivalent to defining the matrices

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

for $1 \leq i \leq n$, where $\mathbf{b}_{i,j} \in \mathbb{F}_2^d$. The assumptions on $T_\circ$ and on $W_\circ$ imply that $E_{e_i} = 0$ for $n - d + 1 \leq i \leq n$, $\mathbf{b}_{i,i} = \mathbf{0}$ and $\mathbf{b}_{i,j} = \mathbf{b}_{j,i}$. In conclusion, the following result characterises the criteria that the vectors $\mathbf{b}_{i,j}$ must adhere to in order to define an alternative operation as previously described.

Theorem 1 *([15]). Let $T_\circ < \mathrm{AGL}(V,+)$ be 2-elementary, abelian, and regular, and let $d \leq n-2$. The operation $\circ$ induced by $T_\circ$ is such that $d = \dim(W_\circ)$, $T_+ < \mathrm{AGL}(V,\circ)$, and $W_\circ = \mathrm{Span}\{e_{n-d+1}, \ldots, e_n\}$ if and only if the matrix $\Theta_\circ \in (\mathbb{F}_{2^d})^{(n-d)\times(n-d)}$ defined as*

$$\Theta_\circ := \begin{pmatrix} \mathbf{b}_{1,1} & \mathbf{b}_{1,2} & \cdots & \mathbf{b}_{n-d,1} \\ \mathbf{b}_{2,1} & \mathbf{b}_{2,2} & \cdots & \mathbf{b}_{n-d,2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_{n-d,1} & \mathbf{b}_{n-d,2} & \cdots & \mathbf{b}_{n-d,d} \end{pmatrix}$$

is zero-diagonal, symmetric and no non-trivial $\mathbb{F}_2$-linear combination of its columns is the zero vector. The matrix $\Theta_\circ$ is also called the defining matrix *for the operation $\circ$.*

This formulation also gives us an efficient algorithm to compute $a \circ b$ [15].
Every operation $\circ$ induces a *product* $\cdot$ defined as follows:

$$\forall a, b \in V \quad a \cdot b := a + b + a \circ b.$$

The product $a \cdot b$ can be interpreted as the error committed when confusing $a \circ b$ with $a + b$. This brings us to the definition of *error space* of an operation $\circ$ as

$$U_\circ := \{a \cdot b : a, b \in V\}.$$

Finally, we recall some key properties of this product, which turns $(V, +, \cdot)$ into an alternating algebra of nilpotency class two. A comprehensive account of the underlying results can be found in Civino et al. [15].

Proposition 1. *Let $\circ$ be an operation as above and let $\cdot$ be the induced product. Then*

- *$\cdot$ is distributive w.r.t. $+$, i.e. $(V, +, \cdot)$ is an $\mathbb{F}_2$-algebra;*
- *$U_\circ \subset W_\circ$;*
- *for each $x, y \in V$ there exists $\epsilon_{x,y} \in U_\circ$ such that*

$$x + y = x \circ y + \epsilon_{x,y},$$

  with $\epsilon_{x,y} = x \cdot y = (0, \dots, 0, (x_1, \dots, x_{n-d})E_y)$;
- *$U_\circ$ is composed of all possible vectors $w \in W_\circ$ whose last d components are $\mathbb{F}_2$-linear combinations of the defining vectors $\boldsymbol{b}_{ij}$;*
- *$x \cdot y \cdot z = 0$, for all $x, y, z \in V$.*

In the subsequent discussion, the term *alternative operation* refers to an additive law $\circ$ on V as defined above.

2. Parallel operations and their automorphism groups

Let $\circ$ be an alternative operation on the block-sized space V. As outlined in the introduction, if $\lambda \in \mathrm{GL}(V, +)$ represents a (xor)-linear diffusion layer, and $\Delta \in V$ is an input difference traversing λ, predicting the output difference with respect to $\circ$, i.e.

$$x\lambda \circ (x \circ \Delta)\lambda,$$

becomes inherently challenging without additional assumptions on λ that ensure a sufficiently high predictive probability. For this reason, the examination of the following object becomes crucial: in cryptographic terms, it contains potential diffusion layers that allow differentials, whether computed with respect to xor or $\circ$, to propagate with a probability of 1.

Definition 1. Let $\circ$ be an alternative operation on V. Let us define

$$H_\circ := \{f \in \mathrm{GL}(V, +) \mid \forall a, b \in V : \ (a \circ b)f = af \circ bf\}$$

to be the subgroup of $\mathrm{GL}(V, +)$ of permutations that are linear w.r.t. the operation $\circ$. More precisely, denoting by $\mathrm{AGL}(V, \circ)$ the normaliser in $\mathrm{Sym}(V)$ of $T_\circ$ and by $\mathrm{GL}(V, \circ)$ the stabiliser of 0 in $\mathrm{AGL}(V, \circ)$, we have $H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$.

Remark 1. Let us note that the mappings in $H_\circ$ are the automorphisms of the algebra $(V, +, \cdot)$ as in Proposition 1. Therefore, checking whether $\lambda \in \mathrm{GL}(V, +)$ is also in $\mathrm{GL}(V, \circ)$ is equivalent to verifying that $(a \cdot b)\lambda = a\lambda \cdot b\lambda$ for any $a, b \in V$.

Lemma 1. *For each $\lambda \in H_\circ$ it holds $W_\circ\lambda = W_\circ$ and $U_\circ\lambda = U_\circ$.*

Proof. Let $\lambda \in H_\circ$, and we begin by proving that $W_\circ\lambda = W_\circ$. Take any $a \in W_\circ$ and $b \in V$. We aim to show that $a\lambda \circ b = a\lambda + b$. Now,

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

and since λ is invertible, it follows that $W_\circ\lambda = W_\circ$. On the other hand, if $a \in U_\circ$, then $a = b \cdot c$ for some $b, c \in V$. Therefore, $a\lambda = (b \cdot c)\lambda = b\lambda \cdot c\lambda$, which implies $a\lambda \in U_\circ$. □

The structure of the group $H_\circ$ in its most general case has not been understood yet. This work addresses this challenge in a specific scenario, guided by assumptions that are deemed reasonable within the context of differential cryptanalysis.

Assumption 1: $\circ$ is a parallel operation. While the operation $\circ$ could, in theory, be defined on the entire message space V, studying the differential properties of the s-box layer, considered as a function with 2^n inputs, is impractical for standard-size ciphers. For this reason, we focus on operations applied in a *parallel* way to each s-box-sized block, i.e., $\circ = (\circ_1, \circ_2, \dots, \circ_b)$, where for each $1 \leq j \leq b$, $\circ_j$ is an operation on V_j. In this scenario, every operation is acting independently on the s-box space B, regardless of the others. This motivates the following definition.

Definition 2. Let $\circ$ be an alternative operation on V. We say that $\circ$ is *parallel* if for each $1 \leq j \leq b$ there exists an alternative operation $\circ_j$ on V_j such that for each $x, y \in V$ we have

$$x \circ y = \begin{pmatrix} x_1 \\ \vdots \\ x_b \end{pmatrix} \circ \begin{pmatrix} y_1 \\ \vdots \\ y_b \end{pmatrix} = \begin{pmatrix} x_1 \circ_1 y_1 \\ \vdots \\ x_b \circ_b y_b \end{pmatrix},$$

where $x = (x_1, x_2, \dots, x_b)$, $y = (y_1, y_2, \dots, y_b)$ and each component belongs to the s-box-sized space, i.e. $x_j, y_j \in V_j \cong B$ for $1 \leq j \leq b$.

In the notation of Sec. 1.2, up to a block matrix conjugation, we can assume that every element $x \in V$ is associated to a translation $\tau_x = M_x \sigma_x$, with

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

where $M_{x_i}^{\circ_i}$ is the matrix associated to the translation τ_{x_i} with respect to the sum $\circ_i$, as defined in Eq. (2).

Assumption 2: $\dim(W_{\circ_j}) = s - 2$. According to Eq. (1), every operation $\circ_j$ defined at the s-box level must satisfy the bound $\dim(W_{\circ_j}) \leq s - 2$, being $s = \dim(B)$. The situation where the (upper) bound is reached holds particular interest for several reasons, as elaborated further in Civino et al. [15]. Notably,

- if the s-box size s is four, the case where $\dim(W_{\circ_j}) = 2$ is the sole possibility;
- imposing $\dim(W_{\circ_j}) = s - 2$ ensures optimal differential diffusion through the key addition layer (see e.g. [15, Theorem 3.19]);
- other possible choices are at the moment less understood, and seem less suitable for our cryptanalytic application; the case $s - 3$ is discussed as an example in Appendix A.

For the reader's convenience, we present the classification result for $H_{\circ_j}$ obtained by Civino et al. in the considered case. Additionally, it is worth recalling that, according to Theorem 1, any $\circ_j$ for which $\dim(W_{\circ_j}) = s - 2$ is determined by a single nonzero vector $\mathbf{b} \in (\mathbb{F}_2)^{s-2}$.

Theorem 2 *([15]). Let $\circ_j$ be an alternative operation such that $d = \dim(W_{\circ_j}) = s - 2$ defined by a vector $\boldsymbol{b} \in (\mathbb{F}_2)^{s-2}$, and let $\lambda \in (\mathbb{F}_2)^{s \times s}$. The following are equivalent:*

- *$\lambda \in H_{\circ_j}$;*
- *there exist $A \in \mathrm{GL}((\mathbb{F}_2)^2, +)$, $D \in \mathrm{GL}((\mathbb{F}_2)^d, +)$, and $B \in (\mathbb{F}_2)^{2 \times d}$ such that*

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

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

We now recall the following result from Calderini et al. [10, Theorem 4.10].

Theorem 3. *Let $V = \mathbb{F}_2^s$, and T and T' be elementary abelian subgroups of $\mathrm{AGL}(V, +)$ defining two operations $\circ$ and $\diamond$ respectively, such that $\dim(W_\circ) = \dim(W_\diamond) = s - 2$. Then, there exists $g \in \mathrm{GL}(V)$ such that $T' = T^g = gTg^{-1}$.*

This theorem can be easily extended to our parallel setting.

Theorem 4. *Let $\circ, \diamond$ be two parallel operation, defined by $\circ_1, \dots, \circ_b$ and $\diamond_1, \dots, \diamond_b$ respectively, such that for all $\circ_i, \diamond_i$ the weak key space has dimension $s - 2$. Then, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$.*

Proof. For each $\circ_i$ and $\diamond_i$, Theorem 3 provides a matrix $g_i \in \mathrm{GL}(V_i)$ such that

$$T_{\diamond_i} = \{M_{x_i}^{\diamond_i}\}_{x_i \in V_i} = g_i T_{\circ_i} g_i^{-1} = g_i \{M_{x_i}^{\circ_i}\}_{x_i \in V_i} g_i^{-1}.$$

Given the structure of the translations τ_x, we can construct a matrix

$$g = \begin{pmatrix} g_1 & \cdots & \mathbb{0} \\ \vdots & \ddots & \vdots \\ \mathbb{0} & \cdots & g_b \end{pmatrix}.$$

By construction, $g \in \mathrm{GL}(V)$. Additionally, Theorem 3 ensures that $T_\diamond$ is the conjugate of $T_\circ$ via g. This completes the proof. □

Lemma 2. *Let $T_\circ$ and $T_\diamond$ be elementary abelian regular subgroups of $AGL(V,+)$ defining two operations $\circ$ and $\diamond$ respectively, and such that $T_\diamond = T_\circ^g$ for $g \in \mathrm{GL}(V)$. Then $H_\diamond = H_\circ^g$.*

Proof. We know that $\mathrm{AGL}(V,\circ)$ and $\mathrm{AGL}(V,\diamond)$ are the normalisers of $T_\circ$ and $T_\diamond$, respectively. Since $T_\diamond = T_\circ^g$, it follows that $\mathrm{AGL}(V,\diamond) = \mathrm{AGL}(V,\circ)^g$. Consequently, we also have $\mathrm{GL}(V,\diamond) = \mathrm{GL}(V,\circ)^g$, where $\mathrm{GL}(V,\circ)$ is the stabiliser of 0 in $\mathrm{AGL}(V,\circ)$. Finally, the intersection with $\mathrm{GL}(V,+)$ remains preserved since $g \in \mathrm{GL}(V,+)$. □

Remark 2. In the case $d = s-2$, Theorem 4 together with Lemma 2 allows us to restrict, up to conjugation, to the case $\circ_1 = \ldots = \circ_b$. For the sake of simplicity, and without loss of generality, we adopt this assumption from now on.

We are now ready to present the characterisation of elements in the group $H_\circ$ for a parallel operation $\circ = (\circ_1, \circ_2, \ldots, \circ_b)$ with components at the s-box level satisfying $\dim(W_{\circ_j}) = s-2$. Notice that this more general result is valid for *all* s-boxes in the relevant case $s = 4$ (see first comment in Assumption 2), but can also be valid for larger s-boxes. Therefore, the approach can be theoretically scaled to any SPN.

Theorem 5. *Let $\circ = (\circ_1, \circ_2, \ldots, \circ_b)$ be a parallel alternative operation on V such that for each $1 \le j \le b$ $\circ_j$ is an alternative operation on V_j. Let us assume that every $\circ_j$ is such that $\dim(W_{\circ_j}) = s-2$ and it is defined by a vector $\boldsymbol{b} \in (\mathbb{F}_2)^{s-2}$. Let $\lambda \in (\mathbb{F}_2)^{n\times n}$. Then, $\lambda \in H_\circ$ if and only if it can be represented in the block form*

$$\lambda = \left(\begin{array}{cc|c|cc} A_{11} & B_{11} & \cdots & A_{1b} & B_{1b} \\ C_{11} & D_{11} & & C_{1b} & D_{1b} \\ \hline \multicolumn{2}{c|}{\vdots} & \ddots & \multicolumn{2}{c}{\vdots} \\ \hline A_{b1} & B_{b1} & \cdots & A_{bb} & B_{bb} \\ C_{b1} & D_{b1} & & C_{bb} & D_{bb} \end{array}\right),$$

where

1. *$A_{ij} \in (\mathbb{F}_2)^{2\times 2}$ such that for each row and each column of blocks there exists one and only one nonzero A_{ij}; moreover, all the nonzero A_{ij} are invertible;*
2. *$B_{ij} \in (\mathbb{F}_2)^{2\times(s-2)}$;*
3. *$C_{ij} = \mathbb{0}_{(s-2)\times 2}$;*
4. *$D_{ij} \in (\mathbb{F}_2)^{(s-2)\times(s-2)}$ such that if A_{ij} is zero, then $\boldsymbol{b}D_{ij} = \boldsymbol{0}$, and if A_{ij} is invertible, then $\boldsymbol{b}D_{ij} = \boldsymbol{b}$. Moreover, the matrix D defined by*

$$D := \begin{pmatrix} D_{11} & \cdots & D_{1b} \\ \vdots & \ddots & \vdots \\ D_{b1} & \cdots & D_{bb} \end{pmatrix}$$

is invertible.

Proof. First, let us introduce some notation to simplify the proof. We denote e_i^j as the vector with a one in the i-th component of the j-th block, which is equivalent to $e_{s(j-1)+i}$. Thus, the i-th strong component (the component spanned by nonweak keys in the i-th block) is $\mathrm{Span}\{e_1^i, e_2^i\}$, and the i-th weak component (also denoted as $W_\circ^i$) is $\mathrm{Span}\{e_3^i, \ldots, e_s^i\}$. We define $W_\circ = W_\circ^1 \oplus \cdots \oplus W_\circ^b$.

($\Rightarrow$) Let $\lambda \in H_\circ$. We need to show that if λ is divided into blocks as described, all the blocks will have the desired properties. Notice that by definition, the block A_{ij} maps the i-th strong component to the j-th strong component, the block B_{ij} maps

the i-th strong component to the j-th weak component, the block C_{ij} maps the i-th weak component to the j-th strong component, and the block D_{ij} maps the i-th weak component to the j-th weak component.

Firstly, by Lemma 1 (which extends naturally to the parallel setting), we have $W_\circ\lambda = W_\circ$. This implies that $C_{ij} = 0$ for all i, j. Consequently, $W_\circ$ is contained in the image of D; since $\dim(W_\circ) = b(s-2)$, D must be invertible.

Now, let us focus on the blocks A_{ij}. If we take a vector x_i with nonzero strong components in V_i, we want to prove that $x_i\lambda = x_j$ where x_j is a vector with nonzero strong components in V_j for some j. We know that $e_1^i \cdot e_2^i \neq 0$; however, $e_{i_1}^j \cdot e_{i_2}^k = 0$ if $j \neq k$, for all combinations of $i_1, i_2 \in \{1, 2\}$.

Consider $\mathrm{Span}\{e_1^i, e_2^i\}\lambda$, and assume it contains an element with more than one nonzero strong component. Without loss of generality, suppose that the strong component of $e_1^i\lambda$ is given by $e_r^j + e_s^k$, for some $k \neq j$ and $r, s \in \{1, 2\}$. Since we started with a 2-dimensional subspace, its image cannot span both strong subspaces of V_j and V_k. If it does not generate V_k, we would need another basis vector e_t^l with $l \neq i$ such that $e_t^l\lambda$ has a nonzero strong component in V_k different from e_s^k. This implies

$$0 = e_1^i \cdot e_t^l = (e_1^i \cdot e_t^l)\lambda = e_1^i\lambda \cdot e_t^l\lambda \neq 0,$$

which is a contradiction. Therefore, each vector with only one nonzero strong component is mapped to a vector with only one nonzero strong component, meaning that for each row (or column) there is exactly one nonzero A_{ij}. Furthermore, the nonzero A_{ij} must span the j-th strong subspace and must be invertible.

Since $d = s - 2$, $e_1^i \cdot e_2^i$ is the vector with $(0, 0, \mathbf{b})$ in V_i and zero elsewhere. Therefore, the component in V_j of $e_1^i \cdot e_2^i$ is $\mathbf{b}D_{ij}$. If $A_{ij} = 0$, it must be that $e_1^i A_{ij} \cdot e_2^i A_{ij} = 0$, which implies $\mathbf{b}D_{ij} = 0$. Conversely, if A_{ij} is invertible, then $\mathbf{b}D_{ij} = \mathbf{b}$ due to Lemma 1.

($\Leftarrow$) We now need to show that if λ has the form described in the statement, then $\lambda \in H_\circ$. Since $\lambda \in \mathrm{GL}(V, +)$ by construction, we only need to verify that $(x \cdot y)\lambda = x\lambda \cdot y\lambda$ for all $x, y \in V$. Moreover, since by Proposition 1 the dot product is distributive with respect to $+$, we can check this equality component-wise.

If either x or y is a weak key, both sides of the equation are zero. If x is a weak key, then $x \cdot y = 0$ by definition of the dot product. On the right-hand side, since all $C_{ij} = 0$, $x\lambda \in W_\circ$, and thus $x\lambda \cdot y\lambda = 0$.

We are left with two cases: strong vectors from different blocks and strong vectors from the same block. For the first case, consider $(e_{i_1}^j \cdot e_{i_2}^k)\lambda = e_{i_1}^j\lambda \cdot e_{i_2}^k\lambda$ for $j \neq k$ and $i_1, i_2 \in \{1, 2\}$. Since $j \neq k$ implies $e_{i_1}^j \cdot e_{i_2}^k = 0$, the condition on the A blocks of λ ensures that the part of $e_{i_1}^j\lambda$ and $e_{i_2}^k\lambda$ that lies outside $W_\circ$ ends up in different blocks. Thus, $e_{i_1}^j\lambda \cdot e_{i_2}^k\lambda = 0$. No restriction is imposed on the components of $e_{i_1}^j\lambda$ and $e_{i_2}^k\lambda$ in $W_\circ$, as the B blocks of λ are unrestricted. However, these components do not affect the dot product.

For the second case, strong vectors from the same block, consider $(e_{i_1}^j \cdot e_{i_2}^j)\lambda = e_{i_1}^j\lambda \cdot e_{i_2}^j\lambda$. If $i_1 = i_2$, both sides are zero. Without loss of generality, assume $i_1 = 1$ and $i_2 = 2$. Then, $e_1^j \cdot e_2^j$ is a vector with $(0, 0, \mathbf{b})$ in the components corresponding to V_j, and zero elsewhere. For a given j, there is exactly one index k such that $\mathbf{b}D_{jk} = \mathbf{b}$, and for all $l \neq k$, $\mathbf{b}D_{jl} = 0$. Thus, the left-hand side is a vector with $(0, 0, \mathbf{b})$ in the components of V_k and zero elsewhere. Since A_{jk} is invertible and $A_{jl} = 0$ for $l \neq k$, $e_1^j\lambda$ and $e_2^j\lambda$ are distinct vectors in the subspace V_k, and their dot product is a vector with $(0, 0, \mathbf{b})$ in the same component, proving the equality. This completes the proof. □

3. Differential properties of optimal s-boxes

In this section, we examine the differential properties of all possible 4-bit permutations, with respect to all possible alternative operations defined as in Sec. 1.2. In particular, we set $s = 4$ and therefore consider $\mathbb{B} := \mathbb{F}_2^4$. We begin by acknowledging that, despite the compact size of the space, the count of alternative operations on $\mathbb{B}$ is considerable:

Proposition 2 *([10]). There exist 105 elementary abelian regular subgroups groups $T_\circ$ in $\mathrm{AGL}(\mathbb{F}_2^4, +)$. Furthermore, each of them satisfies $T_+ < \mathrm{AGL}(\mathbb{F}_2^4, \circ)$ and $\dim W_\circ = s - 2 = 2$.*

We recall that given a permutation $f \in \mathrm{Sym}(\mathbb{B})$ we can define

$$\delta_f(a, b) = \#\{x \in B \mid xf + (x + a)f = b\}.$$

The *differential uniformity* of f is defined as $\delta_f := \max_{a \neq 0} \delta_f(a, b)$ and it represents the primary metric to consider when assessing the resistance of an s-box to differential cryptanalysis [20].

Several cryptographic properties, including differential uniformity, are preserved under affine equivalence for vectorial Boolean functions [13]. Two functions, denoted as f and g, are considered *affine equivalent* if there exist two affine permutations, α and β, in $\mathrm{AGL}(V, +)$ such that $g = \beta f \alpha$.

Leander and Poschmann [19] provided a comprehensive classification (up to affine equivalence) of permutations over $\mathbb{B} = \mathbb{F}_2^4$. They identified 16 classes with *optimal* cryptographic properties. All 16 classes exhibit a classical differential uniformity

Table 1
Optimal 4-bit permutations according to Leander and Poschmann.

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
G_0	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	B_x	C_x	9_x	3_x	E_x	A_x	5_x
G_1	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	B_x	E_x	3_x	5_x	9_x	A_x	C_x
G_2	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	B_x	E_x	3_x	A_x	C_x	5_x	9_x
G_3	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	C_x	5_x	3_x	A_x	E_x	B_x	9_x
G_4	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	C_x	9_x	B_x	A_x	E_x	5_x	3_x
G_5	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	C_x	B_x	9_x	A_x	E_x	3_x	5_x
G_6	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	C_x	B_x	9_x	A_x	E_x	5_x	3_x
G_7	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	C_x	E_x	B_x	A_x	9_x	3_x	5_x
G_8	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	9_x	5_x	A_x	B_x	3_x	C_x
G_9	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	3_x	5_x	9_x	A_x	C_x
G_{10}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	5_x	A_x	9_x	3_x	C_x
G_{11}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	A_x	5_x	9_x	C_x	3_x
G_{12}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	A_x	9_x	3_x	C_x	5_x
G_{13}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	9_x	5_x	B_x	A_x	3_x
G_{14}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	B_x	3_x	9_x	5_x	A_x
G_{15}	0_x	1_x	2_x	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	B_x	9_x	3_x	A_x	5_x

equal to 4, which represents the best possible value for s-boxes in $\mathrm{Sym}(\mathbb{B})$. The representatives of the 16 classes are listed in Table 1, where each vector is interpreted as a binary number, most significant bit first.

3.1. Dealing with affine equivalence

Our goal is to analyse the differential uniformity of each optimal s-box class, with respect to every alternative operation $\circ$ on $\mathbb{B}$. The definitions given above can be generalised in the obvious way setting $\delta_f^\circ(a,b) = \#\{x \in \mathbb{B} \mid xf \circ (x \circ a)f = b\}$ and calling $\circ$-*differential uniformity* of f the value $\delta_f^\circ := \max_{a\neq 0} \delta_f^\circ(a,b)$.

It is noteworthy that, unlike in the case of classic differential uniformity, the value of δ_f° is not invariant under affine equivalence. However, verifying the $\circ$-differential uniformity of $g_2 G_i g_1$ for any optimal class and every pair $g_1, g_2 \in \mathrm{AGL}(\mathbb{B}, +)$ would be impractical. Therefore, a reduction in the number of permutations to be checked is necessary, and for this purpose, we make the following observations. First, similar to the classical case, the $\circ$-differential uniformity is preserved under affine transformations w.r.t. $\circ$.

Proposition 3. *Given $f \in \mathrm{Sym}(\mathbb{B})$ and $g_1, g_2 \in \mathrm{AGL}(\mathbb{B}, \circ)$ we have*

$$\delta^\circ_{g_1 f g_2}(a,b) = \delta^\circ_f(g_2(a), g_1^{-1}(b)).$$

Moreover, Proposition 2 establishes that for any $\circ$ derived from a translation group in $\mathrm{AGL}(\mathbb{B}, +)$, the $+$-translations are affine with respect to $\circ$. This initial observation allows us to narrow down the analysis to $g_2 G_i g_1$ with $g_1, g_2 \in \mathrm{GL}(\mathbb{B}, +)$, which still remains impractical. Furthermore, considering that $H_\circ = \mathrm{GL}(\mathbb{B}, +) \cap \mathrm{GL}(\mathbb{B}, \circ)$, Proposition 3 establishes that left and right multiplication by elements in $H_\circ$ preserves both $\circ$ and $+$-differential uniformity. It is noteworthy that during this process, the rows of the matrix containing all the $\delta_f^\circ(a,b)$ (DDT$^\circ$) are merely shuffled, thereby preserving the highest element of each row. Therefore, the following conclusion can be easily obtained.

Proposition 4. *Let $g_1, g_2 \in \mathrm{GL}(\mathbb{B}, +)$ and $f \in \mathrm{Sym}(\mathbb{B})$. For any $g_1' \in g_1 H_\circ$ and $g_2' \in H_\circ g_2$ we have*

$$\delta^\circ_{g_2 f g_1} = \delta^\circ_{g_2' f g_1'}.$$

Proof. Take $h_1, h_2 \in H_\circ$ such that $g_1' = g_1 h_1$ and $g_2' = h_2 g_2$. Then,[1]

$$x g_2' f g_1' \circ (x \circ a) g_2' f g_1' = x h_2 g_2 f g_1 \circ (x h_2 \circ a h_2 g_2 f g_1) h_1,$$

implying that $\delta^\circ_{g_2' f g_1'}(a,b) = \delta^\circ_{g_2 f g_1}(a h_2, b h_1^{-1})$. So, $\delta^\circ_{g_2' f g_1'} = \delta^\circ_{g_2 f g_1}$. □

The final proposition allows us to focus solely on g_1 and g_2 within the left and right cosets of $H_\circ$. These reductions facilitate the analysis of the potential $\circ$-differential uniformities attainable across all classes of optimal permutations for the 105 conceivable alternative sums defined over $\mathbb{B}$. For each of the 105 alternative operations, we systematically explored each of the 16 classes, following the described procedure, and we recorded the $\circ$-differential uniformity for every candidate.

[1] We remind here that we use postfix notation in this paper, so that xf denotes the image of x under the function f.

Table 2
Avg. number of functions with given $\circ$-differential uniformity.

Class $\downarrow$ // $\delta^\circ \rightarrow$	2	4	6	8	10	12	14	16
G_0	0	780	6695	2956	359	16	0	12
G_1	0	682	6927	2823	374	0	0	12
G_2	0	781	6695	2956	359	16	0	12
G_3	0	896	7566	2210	146	0	0	0
G_4	0	1104	7770	1825	118	0	0	0
G_5	0	822	7994	1790	212	0	0	0
G_6	0	1120	7441	2108	150	0	0	0
G_7	0	898	7628	2139	133	20	0	0
G_8	0	859	6503	3102	296	48	0	12
G_9	0	1123	7062	2457	141	36	0	0
G_{10}	0	1084	7115	2437	147	36	0	0
G_{11}	0	1202	7299	2159	157	0	0	0
G_{12}	0	1099	7275	2291	153	0	0	0
G_{13}	0	916	7749	1965	176	12	0	0
G_{14}	0	1122	7100	2400	149	48	0	0
G_{15}	0	1122	7100	2400	149	48	0	0

To streamline the presentation, we calculated the average across all 105 operations and presented the consolidated results in Table 2.

In our examination, we observe that if, for a given operation $\circ$, certain elements within an affine equivalence class yield a $\circ$-differential uniformity δ, then this value δ is achieved by some element in the entire class for all alternative operations. Our analysis reveals that certain optimal functions may exhibit the highest differential uniformity (16) for alternative operations, specifically the classes G_0 (containing, e.g., the s-box S1 of Serpent [5]), G_1 (containing, e.g., the s-box of PRESENT [8] and the s-box S2 of Serpent), G_2 (containing, e.g., the s-box S0 of Serpent), and G_8 (containing, e.g., the s-box of NOEKEON [17]). Conversely, the classes G_3, G_4, G_5, G_6, G_{11}, and G_{12} demonstrate more favourable behaviour concerning alternative operations.

Remark 3. We observe from Table 2 that we do not obtain maps with $\circ$-differential uniformity equal to 2 or 14. This is because, with respect to the xor operation, such differential uniformities are not present for permutations in 4 bits [19]. The alternative operations $\circ$ that we are considering induce a vector space structure on V isomorphic to $(V,+)$. Specifically, let $\phi : (V,\circ) \rightarrow (V,+)$ be a vector space isomorphism. For any map $f : V \rightarrow V$, the map $\phi f \phi^{-1}$ will have $\circ$-differential uniformity equal to the classical differential uniformity of f. This is due to the following relationship:

$$(x \circ a)\phi f \phi^{-1} \circ (x)\phi f \phi^{-1} = ((x\phi + a\phi)f + (x\phi)f)\phi^{-1}.$$

4. Experiments on a 16-bit block cipher with 4-bit s-boxes

In this concluding section, we aim to apply the results obtained above to a family of (toy) ciphers.[2] These ciphers may exhibit security under classical differential cryptanalysis but reveal vulnerabilities to the alternative differential approach.

In our experiments, we set $V = \mathbb{F}_2^{16}$, $n = 4$, and $s = 4$, defining $\circ$ as the parallel sum by applying the alternative operation defined by the vector $\mathbf{b} = (0,1)$ to each 4-bit block. Moreover, all our ciphers will feature the 4-bit permutation $\gamma : \mathbb{F}_2^4 \rightarrow \mathbb{F}_2^4$ defined by the sequence $(0_x, E_x, B_x, 1_x, 7_x, C_x, 9_x, 6_x, D_x, 3_x, 4_x, F_x, 2_x, 8_x, A_x, 5_x)$ as its s-box. Precisely, four copies of γ will act on the 16-bit block. Notice that the s-box $\gamma \in$ belongs to G_0 and has $\delta_\gamma = 4$ and $\delta_\gamma^\circ = 16$.

In all the experiments described below, we consider the SPN whose i-th round is obtained by the composition of the parallel application of the s-box γ on every 4-bit block, a 'diffusion layer' λ sampled randomly from $H_\circ$, and the xor with the i-th random round key. We study the difference propagation in the cipher in a long-key scenario, i.e., the key-schedule selects a random long key $k \in \mathbb{F}_2^{16r}$ where r is the number of rounds. To avoid potential bias from a specific key choice, we conduct our experiments by averaging over 2^{15} random long-key generations. This approach gives us a reliable estimate of the expected differential probability for the best differentials in this cipher.

In 150 distinct executions, spanning a range of rounds from 3 to 10, we calculated the discrepancy between the most effective $\circ$-trail and $+$-trail. To manage computational resources, our focus was narrowed down to input differences with a Hamming weight of 1.

The results are depicted in Fig. 1, where each dot represents an individual simulation. The x axis corresponds to the negative logarithm of the probability of the best $\circ$ differential, while the y axis represents the difference between that value and the negative logarithm of the probability of the best $+$ differential. Darker dots indicate a higher number of rounds, as explained in the legend. Notably, about half of the dots lie above zero, suggesting that the best $\circ$ differential

[2] The implementation of our simulations is made available here: https://github.com/KULeuven-COSIC/alternative-differential-cryptanalysis.

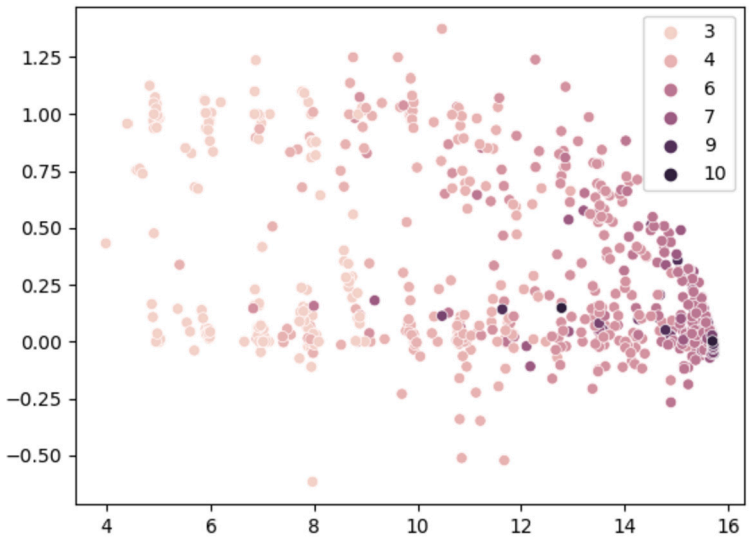


Fig. 1. Comparison of $+$ and $\circ$ trails for random mixing layers.

consistently outperforms the best $+$ differential until they become indistinguishable. Interestingly, this convergence often occurs when the $\circ$ probability is already very close to 2^{-16}, providing potential candidates for our distinguisher attack.

CRediT authorship contribution statement

All authors contributed to writing this manuscript and all authors read and approved the submitted version.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

M. Calderini and R. Civino are members of INdAM-GNSAGA (Italy) and thankfully acknowledge support by MUR-Italy via PRIN 2022RFAZCJ 'Algebraic methods in cryptanalysis'. R. Civino is supported by the Centre of EXcellence on Connected, Geo-Localized and Cybersecure Vehicles (EX-Emerge), funded by Italian Government under CIPE resolution n. 70/2017 (Aug. 7, 2017). R. Invernizzi is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ISOCRYPT - No. 101020788) and by CyberSecurity Research Flanders with reference number VR20192203.

Appendix A. The case $d = s - 3$

A natural next step is to extend the characterization of Theorem 2 to higher-dimensional $W_\circ$. In this section we will consider a new operation $\circ$ such that $d = \dim(W_\circ) = s - 3$. Notice that the following results do not apply in the case $s = 4$ and require larger values of the s-box size. Thanks to Theorem 1, its defining matrix can be written as

$$\Theta_\circ = \begin{pmatrix} \mathbf{0} & \mathbf{b}_{21} & \mathbf{b}_{31} \\ \mathbf{b}_{21} & \mathbf{0} & \mathbf{b}_{32} \\ \mathbf{b}_{31} & \mathbf{b}_{32} & \mathbf{0} \end{pmatrix}$$

where no $\mathbb{F}_2$-linear combination of columns the null vector. The multiplication matrices are

$$M_{e_1} = \left(\begin{array}{c|c} \mathbb{1}_3 & \begin{matrix} \mathbf{0} \\ \mathbf{b}_{21} \\ \mathbf{b}_{31} \end{matrix} \\ \hline \mathbb{0}_{s-3,3} & \mathbb{1}_{s-3} \end{array}\right), \; M_{e_2} = \left(\begin{array}{c|c} \mathbb{1}_3 & \begin{matrix} \mathbf{b}_{21} \\ \mathbf{0} \\ \mathbf{b}_{32} \end{matrix} \\ \hline \mathbb{0}_{s-3,3} & \mathbb{1}_{s-3} \end{array}\right), \; M_{e_3} = \left(\begin{array}{c|c} \mathbb{1}_3 & \begin{matrix} \mathbf{b}_{31} \\ \mathbf{b}_{32} \\ \mathbf{0} \end{matrix} \\ \hline \mathbb{0}_{s-3,3} & \mathbb{1}_{s-3} \end{array}\right),$$

while as always $M_{e_j} = \mathbb{1}_s$ for $j > 3$. In this case we have to deal with multiple error vectors, namely

$$e_i \cdot e_j = e_j \cdot e_i = (0, 0, 0, \mathbf{b}_{ij}).$$

It is hence convenient to introduce the notation $u_{ij} = u_{ji} := e_i \cdot e_j = (0, 0, 0, \mathbf{b}_{ij})$.

Proposition 5. *Let $\circ$ be an operation such that $\dim(W_\circ) = s - 3$. Then $\dim(U_\circ) \in \{2, 3\}$.*

Proof. Thanks to Proposition 1, $U_\circ$ is made of all the vectors $w \in W_\circ$ whose last d components are all the possible $\mathbb{F}_2$-linear combinations of the vectors $\mathbf{b}_{ij}$. By Theorem 1, at least two of these vectors are granted to be independent, hence $\dim(U_\circ) \geq 2$. Moreover, they are all spanned by u_{12}, u_{13} and u_{23}, hence $\dim(U_\circ) \leq 3$. □

Theorem 6. *Let $\circ$ be an operation such that $\dim(W_\circ) = s - 3$, defined by $\Theta_\circ$ as above. Let $\lambda \in (\mathbb{F}_2)^{s\times s}$. Then $\lambda \in H_\circ$ if and only if we can write*

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

for some $A \in \mathrm{GL}((\mathbb{F}_2)^3, +)$, $D \in \mathrm{GL}((\mathbb{F}_2)^d, +)$, and $B \in (\mathbb{F}_2)^{3\times d}$ such that the three compatibility equations

$$\boldsymbol{b}_{ij} D = \sum_{k_1,k_2=1}^{3} (A_{ik_1} A_{jk_2} + A_{ik_2} A_{jk_1}) \boldsymbol{b}_{k_1k_2}$$

for $1 \leq i, j \leq 3$, $i \neq j$ are satisfied, where A_{ij} is the entry of A in the i-th row and j-th column.

Proof. Let us begin with $\lambda \in H_\circ$. By Lemma 1, the first three components of the last d rows must be zero, as they would otherwise map elements from $W_\circ$ outside of $W_\circ$. Furthermore, $\lambda \in \mathrm{GL}((\mathbb{F}_2)^s, +)$ implies that both $A \in \mathrm{GL}((\mathbb{F}_2)^3, +)$ and $D \in \mathrm{GL}((\mathbb{F}_2)^d, +)$ are invertible. For any i, j, we can express $u_{ij}\lambda$ as $e_i\lambda \cdot e_j\lambda$. This gives us:

$$e_i\lambda = \sum_{k=1}^{3} A_{ik} e_k + \sum_{k=1}^{n-3} B_{ik} e_{k+3}$$

and similarly for e_j. Since the dot product is distributive over $+$, we can disregard the second sum, which consists only of weak vectors. Thus, we obtain:

$$e_i\lambda \cdot e_j\lambda = \left(\sum_{k_1=1}^{3} A_{ik_1} e_{k_1}\right) \cdot \left(\sum_{k_2=1}^{3} A_{jk_2} e_{k_2}\right) = \sum_{k_1,k_2=1}^{3} \left(A_{ik_1} A_{jk_2} + A_{ik_2} A_{jk_1}\right) u_{ij}.$$

Note that these equations reduce to zero when $i = j$; since $u_{ii} = 0$, we can focus on the case $i \neq j$. We know that the first three components of these products are zero, so we restrict our attention to the component in $W_\circ$ on both sides. The above expression simplifies to:

$$\mathbf{b}_{ij} D = \sum_{k_1,k_2=1}^{3} (A_{ik_1} A_{jk_2} + A_{ik_2} A_{jk_1}) \mathbf{b}_{k_1k_2}$$

for $1 \leq i, j \leq 3$, $i \neq j$.

Now, suppose λ is in the form prescribed by the theorem; we aim to prove that $\lambda \in H_\circ$. The zero block and the invertibility of A and D are sufficient to show that $\lambda \in \mathrm{GL}(V, +)$. As in Theorem 5, it suffices to show that $(e_i \cdot e_j)\lambda = e_i\lambda \cdot e_j\lambda$. If at least one of $e_i, e_j \in W_\circ$ (say e_i), then the left-hand side is zero. Furthermore, the zero block in the lower left of the matrix ensures that $e_i\lambda \in W_\circ$, which implies the right-hand side is also zero. We are left with the cases where $e_i, e_j \notin W_\circ$. If $i = j$, we have $0 = 0$ by the definition of the dot product. The three cases where $i \neq j$ correspond exactly to the three compatibility equations, as shown in the first part of the proof. □

We will now present two examples of how Theorem 6 allows us to count the number of elements in $H_\circ$. These computations will be important in proving Theorem 7.

Example 1. Let $V = (\mathbb{F}_2)^6$, and let $\circ$ be an operation where $\dim(W_\circ) = \dim(U_\circ) = s - 3 = 3$. By applying Theorem 6, we can determine the size of $H_\circ$. Specifically, A can be any matrix from $\mathrm{GL}((\mathbb{F}_2)^3)$, providing 168 possible choices. Note that A defines the image of e_i for $1 \leq i \leq 3$, and thus determines the image of $U_\circ$. Since in this case $U_\circ = W_\circ$, the compatibility equations uniquely determine D for each selected A (which is not always the case: for each A, any D that fixes $U_\circ$ and permutes the remaining vectors in a basis of $W_\circ$ can be chosen). Additionally, B can be any 3×3 matrix, giving us 512 options. Therefore, the total size of $H_\circ$ is 86016.

Example 2. Let $V = (\mathbb{F}_2)^6$, and let $\circ$ be an operation where $\dim(W_\circ) = 3$ and $\dim(U_\circ) = 2$. In this case, the choice of A is more restricted: the error vectors u_{12}, u_{13}, and u_{23} are not independent, and their relations must be preserved by λ. For some choices of A, the compatibility equations become unsolvable. Let us consider, for example, the operation $\circ$ defined by

$$\Theta_\circ = \begin{pmatrix} 000 & 101 & 110 \\ 101 & 000 & 101 \\ 110 & 101 & 000 \end{pmatrix}.$$

Here, we have $\mathbf{b}_{12} = \mathbf{b}_{23} = (101)$. Now, take the matrix

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

which is invertible but not a valid choice. Let us show why. Assuming $B = 0$ for simplicity, we get

$$e_1\lambda = e_1 + e_2, \quad e_2\lambda = e_1 + e_2 + e_3, \quad e_3\lambda = e_2,$$

which leads to

$$\mathbf{b}_{12}D = e_1\lambda \cdot e_2\lambda = (e_1 \cdot e_3) + (e_2 \cdot e_3) = \mathbf{b}_{13} + \mathbf{b}_{23} = (011),$$
$$\mathbf{b}_{23}D = \mathbf{b}_{12} + \mathbf{b}_{23} = (000).$$

However, this is impossible since we had set $\mathbf{b}_{12} = \mathbf{b}_{23}$.

When $\dim(U_\circ) = 2$, it is more practical to start by choosing D. If we fix a basis for $U_\circ$, we have two vectors whose images must span $U_\circ$ itself; since $U_\circ$ contains 3 pairwise independent vectors, there are 6 ways to choose them. We can then extend this to form a basis for $W_\circ$, and the image of the last vector can be any vector in $W_\circ \setminus U_\circ$, giving us 4 options. Thus, there are 24 possible choices for D.

The choice of D determines the preimage of $U_\circ$ through the compatibility equations, which has dimension 2, leaving us with 4 possible choices for A, one for each way to map the last vector that does not appear in the equations. Combined with the 512 choices for B, we find that $H_\circ$ has a total of 49152 elements.

We will now show that this is a consequence of a deeper fact, i.e. the conjugacy classes of operations with $d = s - 3$ only depend on $\dim(U_\circ)$.

Theorem 7. *Let $\circ$ and $\diamond$ be two operations such that $\dim(W_\circ) = \dim(W_\diamond) = s - 3$, and let $T_\circ$ and $T_\diamond$ be respectively the associated translation groups. Then, there exists $g \in \mathrm{GL}(V)$ such that $T_\diamond = T_\circ^g$ if and only if $\dim(U_\circ) = \dim(U_\diamond)$.*

Proof. Up to conjugation, assume that $W_\circ$ and $W_\diamond$ are generated by $\{e_4, \dots, e_s\}$, and that $U_\circ$ and $U_\diamond$ are generated by $\{e_4, e_5\}$ if $\dim(U_\circ) = 2$, or by $\{e_4, e_5, e_6\}$ if $\dim(U_\circ) = 3$. The multiplication matrices associated with e_i for $1 \le i \le 3$ can be expressed as

$$\begin{pmatrix} M_i^\circ & \mathbb{0}_{6,s-6} \\ \mathbb{0}_{s-6,6} & \mathbb{I}_{s-6} \end{pmatrix}, \quad \begin{pmatrix} M_i^\diamond & \mathbb{0}_{6,s-6} \\ \mathbb{0}_{s-6,6} & \mathbb{I}_{s-6} \end{pmatrix}$$

respectively. Here, M_i° and $M_i^\diamond$ are 6×6 matrices because the rows of the multiplication matrices generate $U_\circ$, as stated in Proposition 1, and are zero in the last $s - 6$ components. We can thus view this sum as consisting of two parallel sums: the first one acting on the first six components and defined by M_i° and $M_i^\diamond$, and the second one acting on the last $s - 6$ components as the standard sum.

Now, let $\bar{T}_\circ$ and $\bar{T}_\diamond$ be the translation groups associated with the alternative operations defined by M_i° and $M_i^\diamond$. Based on the classification in Calderini et al. [10, Table 1], we know that for $s = 6$ and $d = 3$, there are two distinct conjugacy classes of translation groups. Examples 2 and 1 demonstrate that these two classes correspond to $\dim(U_\circ) = 2$ and $\dim(U_\circ) = 3$. By Lemma 2, the conjugation on $T_\circ$ induces conjugation on $H_\circ$, and conjugation must preserve the number of elements. This implies that, if and only if $\dim(\bar{U}_\circ) = \dim(\bar{U}_\diamond)$, there exists $\bar{g} \in \mathrm{GL}_6(V)$ such that $\bar{T}_\diamond = \bar{T}_\circ^{\bar{g}}$. Taking

$$g = \begin{pmatrix} \bar{g} & \mathbb{0}_{6,s-6} \\ \mathbb{0}_{s-6,6} & \mathbb{I}_{s-6} \end{pmatrix}$$

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

Corollary 1. *Given two sum $\circ$ and $\diamond$ with $\dim(W_\circ) = \dim(W_\diamond) = s - 3$, the corresponding groups $H_\circ$ and $H_\diamond$ are conjugated by $g \in \mathrm{GL}(V)$ if and only if $\dim(U_\circ) = \dim(U_\diamond)$.*

Table 3
Distribution of δ°-differential uniformities for $d = 6$ on selected 8-bit s-boxes.

s-box $\downarrow$ // $\delta^\circ \rightarrow$	8	10	12
AES	55	8	·
Camellia	59	4	·
Kuznyechik	·	54	9

Table 4
Distribution of δ°-differential uniformities for $d = 5$ on selected 8-bit s-boxes.

s-box $\downarrow$ // $\delta^\circ \rightarrow$	8	10	12	14	16	18
AES	433	23858	7841	402	14	2
Camellia	470	24087	7494	476	22	1
Kuznyechik	18	18940	12425	1086	80	1

Table 5
Distribution of δ°-differential uniformities for randomly selected operations on the AES s-box for $d = 2, 3, 4$.

$d \downarrow$ // $\delta^\circ \rightarrow$	8	10	12	14	16	18	20	24
4	58	54528	41515	3604	281	12	2	·
3	32	52976	43111	3654	216	11	·	·
2	267	68148	29342	2073	149	20	·	1

Proof. Apply Lemma 2 to Theorem 7. □

Comparing Theorem 7 with its analogous Theorem 3 highlights the increased complexity when transitioning to the case where $d = s - 3$. Higher values of d appear to be even more intricate and are not yet fully understood. We should keep this in mind if we want to use such sums in a parallel setting; for example, we can no longer assume that the same alternative operation can be applied to all components. Consequently, the techniques employed to prove Theorem 5 become invalid. This, along with the considerations discussed in Section 2, clarifies our assumption of $d = s - 2$ for our applications.

A.1. Considerations on 8-bit s-boxes and the case $d = s - 3$

Our analysis naturally extends beyond the 4-bit setting. In particular, when considering 8-bit s-boxes, experiments show that improvements can indeed be observed, especially for small values of the parameter $2 \leq d \leq 6$. This is expected, as d measures the extent to which the alternative operation resembles the classical xor. For larger s-boxes, however, the phenomenon does not reach the extremal levels observed in dimension four, since constant differentials arise only under very specific algebraic coincidences that are unlikely to occur in higher dimensions. Nevertheless, the evidence shows that the study of alternative operations for larger s-boxes is meaningful and can reveal structural weaknesses that remain invisible to classical xor-based analysis.

Extending our analysis to all 8-bit permutations is not feasible: firstly, because the number of such permutations is extremely large, and secondly, because the number of possible alternative operations is also very high (roughly 2^{47} [14]). Nonetheless, we can analyse some common s-boxes for the cases $d = 5$ and $d = 6$ and show that, as expected, the $d = 5$ case exhibits better performance in terms of $\circ$-nonlinearity gain (the operation is more dissimilar from xor than in the case $d = 6$). We tested all alternative operations in canonical form, that is with $W_\circ$ generated by the last d canonical vectors, (63 for $d = 6$ and 32,550 for $d = 5$) on the AES [16], Camellia [1], and Kuznyechik [7] s-boxes, and the results are summarized in Table 3 and Table 4. These data show that in the $s - 3$ setting there is a significant improvement compared to the case $s - 2$, justifying a detailed analysis of the $s - 3$ case. Although absent in 4-bit instances, it becomes highly relevant for 8-bit s-boxes and exhibits behaviours that deserve attention in their own right. Compared to the $s - 2$ case, where the distribution of differential uniformities appears more concentrated, in the $s - 3$ case a greater irregularity emerges. This difference can be traced back to the richer structure of the error space $U_\circ$ in the $s - 3$ situation, as discussed previously.

For the other cases $d = 2, 3, 4$, we have tested 100,000 randomly selected alternative operations on the AES s-box. We remark that, due to the vast number of possible operations, the likelihood of identifying an operation that significantly outperforms the others, if any, is extremely low. The corresponding results are summarized in Table 5.

Data availability

No data was used for the research described in the article.

References

[1] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, T. Tokita, Camellia: a 128-bit block cipher suitable for multiple platforms—design and analysis, in: International Workshop on Selected Areas in Cryptography, Springer, 2000, pp. 39–56.

[2] S. Banik, A. Bogdanov, T. Isobe, K. Shibutani, H. Hiwatari, T. Akishita, F. Regazzoni, Midori: a block cipher for low energy, in: Advances in Cryptology–ASIACRYPT 2015: 21st International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29–December 3, 2015, Proceedings, Part II 21, Springer, 2015, pp. 411–436.

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

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

[5] E. Biham, R. Anderson, L. Knudsen, Serpent: a new block cipher proposal, in: International Workshop on Fast Software Encryption, Springer, 1998, pp. 222–238.

[6] E. Biham, A. Biryukov, A. Shamir, Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials, J. Cryptol. 18 (2005) 291–311.

[7] A. Biryukov, L. Perrin, A. Udovenko, Reverse-engineering the s-box of Streebog, Kuznyechik and STRIBOBr1, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer, 2016, pp. 372–402.

[8] A. Bogdanov, L.R. Knudsen, G. Leander, C. Paar, A. Poschmann, M.J. Robshaw, Y. Seurin, C. Vikkelsoe, Present: an ultra-lightweight block cipher, in: Cryptographic Hardware and Embedded Systems-CHES 2007: 9th International Workshop, Vienna, Austria, September 10-13, 2007. Proceedings 9, Springer, 2007, pp. 450–466.

[9] N. Borisov, M. Chew, R. Johnson, D. Wagner, Multiplicative differentials, in: Fast Software Encryption: 9th International Workshop, FSE 2002 Leuven, Belgium, February 4–6, 2002 Revised Papers 9, Springer, 2002, pp. 17–33.

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

[11] M. Calderini, R. Civino, R. Invernizzi, Differential experiments using parallel alternative operations, J. Math. Cryptol. 18 (1) (2024) 20230030.

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

[13] C. Carlet, Boolean Functions for Cryptography and Coding Theory, Cambridge University Press, 2021.

[14] R. Civino, V. Fedele, Classification of small binary bibraces via bilinear maps, https://arxiv.org/abs/2510.05848, 2025.

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

[16] J. Daemen, V. Rijmen, The Design of Rijndael, vol. 2, Springer, 2002.

[17] J. Daemen, M. Peeters, G. Van Assche, V. Rijmen, Nessie proposal: Noekeon, in: First Open NESSIE Workshop, 2000, pp. 213–230.

[18] L.R. Knudsen, Truncated and higher order differentials, in: Fast Software Encryption: Second International Workshop, Leuven, Belgium, December 14–16, 1994 Proceedings 2, Springer, 1995, pp. 196–211.

[19] G. Leander, A. Poschmann, On the classification of 4 bit s-boxes, in: Arithmetic of Finite Fields: First International Workshop, WAIFI 2007, Madrid, Spain, June 21-22, 2007. Proceedings 1, Springer, 2007, pp. 159–176.

[20] K. Nyberg, Differentially uniform mappings for cryptography, in: Workshop on the Theory and Application of Cryptographic Techniques, Springer, 1993, pp. 55–64.

[21] K. Shibutani, T. Isobe, H. Hiwatari, A. Mitsuda, T. Akishita, T. Shirai, Piccolo: an ultra-lightweight blockcipher, in: Cryptographic Hardware and Embedded Systems–CHES 2011: 13th International Workshop, Nara, Japan, September 28–October 1, 2011. Proceedings 13, Springer, 2011, pp. 342–357.

[22] G. Teşeleanu, The security of quasigroups based substitution permutation networks, in: International Conference on Information Technology and Communications Security, Springer, 2022, pp. 306–319.

[23] D. Wagner, The boomerang attack, in: International Workshop on Fast Software Encryption, Springer, 1999, pp. 156–170.

[24] D. Wagner, Towards a unifying view of block cipher cryptanalysis, in: Fast Software Encryption: 11th International Workshop, FSE 2004, Delhi, India, February 5-7, 2004. Revised Papers 11, Springer, 2004, pp. 16–33.

[25] P. Zajac, M. Jókay, Cryptographic properties of small bijective s-boxes with respect to modular addition, Cryptogr. Commun. 12 (2020) 947–963.