Linear Cryptanalysis with Bibrace Characters: Weak-Key Distinguishers for CRAFT

Roberto Civino

Abstract

Linear cryptanalysis measures the correlation of a cipher with the characters of the group used to define differences. If that group is replaced by a second elementary abelian group structure on the same set, here the one coming from a binary bibrace, then the admissible masks are no longer the ordinary scalar products: exactly half of them survive, and the other half are forced to be quadratic. Beyne's geometric approach develops linear cryptanalysis over an arbitrary finite abelian group, providing a natural framework for this setting. We instantiate it on the group of a particular bibrace and apply it to CRAFT.

Over this group the MIDORI/CRAFT S-box has four probability-one relations, forming a small *subgroup* of the dual which the S-box preserves in both directions. A trail that stays inside that subgroup is determined by its first mask and propagates linearly, so the search for the best such trail is a minimum weight codeword problem, which we solve exactly by complete enumeration.

A trail costs correlation, and it restricts the key to a weak-key class. We show that the exact reading, obtained by analysing the diffusion layer and the key addition together rather than separately, gives a class several bits larger than the one obtained cell by cell. One concrete consequence is that CRAFT's round constants, whatever their values, impose no restriction at all.

On CRAFT we obtain weak-key distinguishers up to eighteen rounds. At fourteen rounds a single trail has squared correlation 2^{-44} over a class of 2^{108} keys, against $2^{-62.12}$ for the designers' linear hull, which is the best known linear result on the cipher and holds for all keys. On that class the trail alone therefore exceeds the best linear correlation by eighteen bits at equal round count. Half of the optimal trails begin and end with ordinary linear masks, so these are linear approximations of CRAFT in the usual sense. A contrasting analysis of MIDORI64, which shares the S-box, traces its protection to the round constants and describes two known invariants of that cipher through other bibraces.

Keywords: linear cryptanalysis, CRAFT, bibrace, alternative operations, nonlinear invariants, weak keys

1 Introduction

Context. Differential cryptanalysis, introduced by Biham and Shamir [BS91], studies how a cipher interacts with the group of translations used to define differences; linear cryptanalysis, introduced by Matsui [Mat94], studies instead its correlation with the characters of that group. The two are the standard criteria against which block ciphers are designed, and both usually take the group for granted. On the same message space there are other elementary abelian group structures besides $\oplus$, and Calderini, Civino and Sala [CCS21] described those whose translations are still affine over $\oplus$. Civino, Blondeau

Email: math@robertocivino.it (Roberto Civino)

This work is licensed under a "CC BY 4.0" license.
Date of this document: 2026-11-07.

and Sala [CBS19] put those *alternative* sums to work on the differential side, obtaining a distinguisher on a small-scale network that resists ordinary differential cryptanalysis. Calderini, Civino and Invernizzi [CCI24] extended the construction to *parallel* alternative operations, acting on every S-box of the block at once rather than on a single one, which is the regime we are in. In [CCI26] they classified the diffusion layers that are linear with respect to such operations, and surveyed the differential behaviour of all the classes of optimal 4-bit S-boxes of Leander and Poschmann [LP07]: some of those permutations, optimal in the ordinary sense, turn out to be 16-differentially uniform once the operation is changed, that is, to admit a differential of probability one. The present paper finds the same degeneration on the linear side, where the MIDORI/CRAFT S-box has entries of absolute value 16 in its $\circ$-LAT, that is, 'linear' relations of correlation one.

The alternative operation and the bibrace are the same object under two names, and we use them interchangeably. A bibrace is a set carrying two compatible group structures; presenting an alternative operation $\circ$ alongside $\oplus$ is exactly presenting such a pair, and every statement below about *the bibrace* B is a statement about the operation it defines through (1). The name comes from the braces of Rump [Rum07], introduced to produce set-theoretic solutions of the Yang–Baxter equation in the sense of Etingof, Schedler and Soloviev [ESS99], and generalised to the noncommutative setting by Guarnieri and Vendramin [GV17]; the cryptographic use is unrelated to that motivation, but the algebra is the same. The formalism we follow is that of Civino and Fedele [CF25], whose companion paper [CF26] classifies the small binary bibraces up to dimension eight through the corresponding alternating algebras over $\mathbb{F}_2$, and in particular counts the alternative operations available.

Wagner's commutative diagram cryptanalysis [Wag04] has recently grown into an active line of work: Baudrin et al. [BFL$^+$23] turned it into a concrete tool, exploiting pairs of affine permutations A, B for which $E_k \circ A = B \circ E_k$ holds with high probability over the inputs, for a large class of weak keys. The same authors, joined by Beierle, subsequently allow A and B to be arbitrary maps on a finite abelian group and situate the resulting notion with respect to differential cryptanalysis [BBF$^+$25], obtaining differential, rotational and rotational-differential attacks, as well as c-differentials on S-boxes, as instances of a single framework. The differential side of a bibrace is an instance of this: the $\circ$-translations are affine over $\oplus$, so taking A and B among them is commutative cryptanalysis with the commuting pair drawn from one regular abelian group. Commutative cryptanalysis asks only for a commuting pair, whereas a bibrace gives a whole group operation.

On the linear side, Knudsen and Robshaw [KR96] introduced nonlinear approximations and observed that a key addition does not preserve them in general. Beierle, Canteaut and Leander [BCL18] revisited them by conjugating every keyed instance of the cipher with a fixed permutation G, so that a nonlinear approximation of the cipher becomes a linear approximation of the conjugated one, and called the resulting trails G-shifted linear trails; they leave open how the components of G should be chosen to make the analysis tractable. Beyne introduced correlation matrices as a tool for invariants [Bey18] and then, in the geometric approach [Bey21], recast linear cryptanalysis over an arbitrary finite abelian group, which unifies these variants. Links between the differential and linear worlds were established by Chabaud and Vaudenay [CV95] and developed by Blondeau and Nyberg [BN14] and by Blondeau, Leander and Nyberg [BLN17].

This work. We study here the linear side of a bibrace. The starting point is not new: once the difference group is changed, the masks are forced, being the characters of the new group, and this is exactly the situation the geometric approach of [Bey21] was built for. Moreover the group of a bibrace is isomorphic to the usual one through an explicit quadratic map θ (Section 3), so its characters are the ordinary masks composed with θ, and the trails of this paper are the G-shifted linear trails of [BCL18] with $G = \theta$ on every

cell. The bibrace contributes a choice of G for which the analysis stays tractable, since after conjugation every key addition is still affine, which is what fails for a general G and what makes nonlinear approximations hard to carry across key additions [KR96].

Our target is Craft [BLMR19], a 64-bit tweakable block cipher with a 128-bit key and a 64-bit tweak, designed by Beierle, Leander, Moradi and Rasoolzadeh for efficient protection against differential fault attacks. Its round is built from involutory components, an involutory 4-bit S-box shared with Midori [BBI+15], an involutory binary linear layer and an involutory cell permutation, and its tweakey schedule reuses four round tweakeys with period four. Craft has been examined from several directions, by its designers and then in the (truncated) differential, related-key, zero-correlation, integral and rectangle models [EY19, HSN+19, GSS+20, MA20, HBS21, LR22, MNPN25]. The furthest any of them reaches is the recent work of M'Foukh, Naya-Plasencia and Neumann [MNPN25], whose state-test truncated-differential meet-in-the-middle attacks cover 26 of the 32 rounds in the single-tweak model and for every key. None of these attacks breaks the full cipher in the single-key model, and Section 6.2 places our results against them.

The first contribution of this work is descriptive. We determine the admissible masks of a bibrace completely and in closed form (Section 3). Half of them are ordinary linear masks (*flat* masks) and half are quadratic (*curved* masks), all the quadratic ones share a single quadratic part, and a single vector attached to the bibrace decides which masks are quadratic.

The second is structural, and concerns the Midori/Craft S-box and the Craft round function (Section 4). Over the right bibrace the S-box has four relations of correlation one, and the masks carrying them form a small subgroup of the dual which the S-box preserves in both directions and permutes deterministically. Inside that subgroup the S-box exchanges one flat mask with one curved mask and fixes the other two, so a cell carrying either of the exchanged masks alternates between flat and curved at every S-box, at no cost. A flat cell is unaffected by the key addition and costs nothing at the diffusion layer, so a trail inside the subgroup that starts flat is entirely flat at every other round and pays only at the rounds in between. We also determine exactly what the diffusion layer costs, and why.

The third is the criterion that decides which keys a trail needs (Section 4.3), and it is the part we expect to transfer to other ciphers. We determine exactly which tweakey nibbles a trail constrains, and the constraint turns out to be weaker than the cell-by-cell reading that asks every curved mask to stay in place, which is the one behind the weak keys of [BCL18, Sections 4.2–4.3], so the weak-key class is correspondingly larger. One row of Craft's diffusion layer, the only one that is a sum of an even number of cells, never constrains the key, and that row is exactly where Craft adds its round constants. The constants therefore impose no condition on our trails, whatever their values (Section 4.4).

The fourth is methodological (Section 5). Restricted to the subgroup above, the whole propagation is linear over $\mathbb{F}_2$ and the cost is the weight of a linear image, so finding the best trail *inside that subgroup* is a minimum weight codeword problem. We solve it by complete enumeration, which makes that optimum certified rather than best-effort and, as a by-product, gives the number of optimal masks.

The fifth is the results themselves (Section 6): weak-key distinguishers on up to eighteen rounds of Craft. Half of the optimal trails begin and end with flat masks, so the distinguishers are ordinary linear approximations of Craft whose correlation is computed through a trail of quadratic masks, a situation described in [Bey21, Section 5.3]. The sixth is a contrasting analysis of Midori64 (Section 7), which locates the difference between the two ciphers in the round constants, and shows along the way that the nonlinear invariant Todo, Leander and Sasaki give for Midori [TLS16] is a circle-linear character of a bibrace, so that a known invariant attack, already cast as a shifted linear trail in [BCL18, Section 4.2], becomes an instance of the present framework.

Our distinguishers hold on a class of weak keys, of size 2^{120} at eight rounds down to 2^{108} at fourteen and 2^{110} at eighteen, whereas the linear hull we improve upon holds for every key, so the comparison is not homogeneous and we present it as such; and the optimality of our trails is certified inside the subgroup of masks described above, not over all masks. At fourteen rounds a single trail has squared correlation 2^{-44} on a class of 2^{108} keys, where the best published linear hull gives $2^{-62.12}$ for every key, so on that class our trail alone is eighteen bits above their hull, corresponding to 2^{44} data instead of 2^{62}. Table 5 gives the figures at the other lengths.

The bibrace also has a differential side. The invariant property that Guo et al. [GSS+20] discovered by combinatorial means follows from it, and the same reading explains part of the clustering of CRAFT's $\oplus$ differentials observed by Hadipour et al. [HSN+19]. Neither is a new result, since the facts are those of [GSS+20, HSN+19]; Appendix A gives an algebraic account of the first and a partial one of the second, and answers a question left open in [GSS+20].

2 Preliminaries

This section fixes the two objects of the paper: the alternative group operation attached to a binary bibrace, and the specification of CRAFT. The operation is defined by a pair (ψ, b); the vector b will turn out to decide which masks are quadratic, and the radical of ψ which key additions are harmless.

2.1 The bibrace and its operation

Let $V = \mathbb{F}_2^n$. We use the bibraces of [CF25] in their rank-one case, where a bibrace is the triple $(V, \oplus, \circ)$ where $\circ$ is induced by an alternating bilinear form as described below; Remark 2 explains why this is the relevant case.

Let $\psi : V \times V \to \mathbb{F}_2$ be an alternating bilinear form, that is one with $\psi(x, x) = 0$ for every x, which over $\mathbb{F}_2$ forces $\psi(x, y) = \psi(y, x)$. Its *radical* is

$$\mathrm{rad}(\psi) = \{x \in V : \psi(x, y) = 0 \text{ for every } y \in V\},$$

a subspace of V, and we let $b \in \mathrm{rad}(\psi)$ with $b \neq 0$. Define

$$x \circ y \;=\; x \oplus y \oplus \psi(x, y)\, b. \tag{1}$$

Since b lies in the radical, $(V, \circ)$ is again an elementary abelian group with neutral element 0. We write $\mathrm{Ann} = \mathrm{rad}(\psi)$; on Ann the two operations agree. The name comes from bibrace theory [CF25, CF26], where the product $x \cdot y = (x \circ y) \oplus x \oplus y$ links the two operations and makes $(V, \oplus, \cdot)$ an algebra. Here $x \cdot y = \psi(x, y)\, b$, and the annihilator of this algebra, the set of x with $x \cdot y = 0$ for every y, is exactly $\mathrm{rad}(\psi)$.

Throughout, we let $n = 4$ and we fix

$$\psi(x, y) = (x_0 \oplus x_2)y_3 \oplus x_3(y_0 \oplus y_2), \qquad b = \mathsf{5}, \qquad \mathrm{Ann} = \{\mathsf{0}, \mathsf{2}, \mathsf{5}, \mathsf{7}\},$$

with x_0 the least significant bit. We call this bibrace B. Of the 105 bibraces on $\mathbb{F}_2^4$ as classified in [CF25, CF26] (pairs (ψ, b) with ψ a nonzero alternating form and $0 \neq b \in \mathrm{rad}(\psi)$), B is optimal for our purposes (see Section 4.2).

2.2 CRAFT

CRAFT [BLMR19] is a tweakable block cipher with a 64-bit block, a 128-bit key $K = K_0 \| K_1$ and a 64-bit public tweak T. The state is a 4×4 array of nibbles; we index cells $0, \ldots, 15$ in row-major order, so that cell i lies in row $\lfloor i/4 \rfloor$ and column $i \bmod 4$, and a column j

consists of the cells $\{j, j+4, j+8, j+12\}$ read as rows $0, 1, 2, 3$. The cipher has 32 rounds; for $0 \le t < 31$,

$$\mathcal{R}_t = \mathrm{SB} \circ \mathrm{PN} \circ \mathrm{ATK}_t \circ \mathrm{ARC}_t \circ \mathrm{MC},$$

written as a composition and therefore applied from right to left. The final round omits PN and SB, keeping the three operations that come first. We describe now each of the five operations.

MixColumn (MC). Each column is multiplied by the involutory binary matrix

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{i.e.} \quad \begin{aligned} y[j] &= x[j] \oplus x[j{+}8] \oplus x[j{+}12], & y[j{+}8] &= x[j{+}8], \\ y[j{+}4] &= x[j{+}4] \oplus x[j{+}12], & y[j{+}12] &= x[j{+}12]. \end{aligned} \tag{2}$$

M is a *half-state* matrix, since it copies rows 2 and 3 and mixes only rows 0 and 1.

AddConstants (ARC). Two nibbles are added to cells 4 and 5 only, namely $RC_4[t]$ to cell 4 and $RC_3[t]$ to cell 5, produced by a 4-bit and a 3-bit LFSR. Cells 4 and 5 are the first two cells of *row* 1 (the second from the top), the only row of M that is a sum of an even number of cells, and Section 4.4 shows that for this reason the constants of Craft impose no condition on our trails, whatever their values. Section 7.1 contrasts this with Midori64.

AddTweakey (ATK). Craft is tweakable, meaning that besides the plaintext and the secret key it takes a third input, the tweak T, which is public and freely chosen. For us it is a parameter of the attack and we take $T = 0$ throughout; Section 4.5 records what a different choice would require, and Section 6.1 for how many tweaks the distinguishers survive.

From $K = K_0 \| K_1$ and T, four round tweakeys are derived once and reused with period four:

$$TK_0 = K_0 \oplus T, \quad TK_1 = K_1 \oplus T, \quad TK_2 = K_0 \oplus \mathcal{Q}(T), \quad TK_3 = K_1 \oplus \mathcal{Q}(T),$$

where $\mathcal{Q}$ permutes the nibble positions of the tweak, $\mathcal{Q}(T)[i] = T[\mathcal{Q}(i)]$, with

$$\mathcal{Q} = [12, 10, 15, 5, 14, 8, 9, 2, 11, 3, 7, 4, 6, 0, 1, 13],$$

and $TK_{t \bmod 4}$ is added in round t. So K_0 is used at the even rounds and K_1 at the odd ones, each appearing twice per period but masked by two different images of the tweak. When $T = 0$ the four round tweakeys collapse to K_0, K_1, K_0, K_1, so two conditions on the same nibble at two rounds of the same parity become a single condition; this is what keeps the weak-key class as large as it is (Section 4.5), and it is why we take $T = 0$.

PermuteNibbles (PN). An involutory permutation of the cells,

$$\mathcal{P} = [15, 12, 13, 14, 10, 9, 8, 11, 6, 5, 4, 7, 1, 2, 3, 0].$$

Reading it by rows, $\mathcal{P}$ exchanges rows 0 and 3 and rows 1 and 2, composing each exchange with a permutation of the columns. In particular it exchanges the top half of the state with the bottom half at every round.

Only the top half of the state contributes to the cost of a trail (Theorem 1), and since $\mathcal{P}$ swaps the two halves at every round, no cell stays there for two consecutive rounds.

SubBox (SB). The involutory 4-bit S-box of MIDORI [BBI+15],

$$\mathrm{Sb}_0 = [\mathtt{c}, \mathtt{a}, \mathtt{d}, \mathtt{3}, \mathtt{e}, \mathtt{b}, \mathtt{f}, \mathtt{7}, \mathtt{8}, \mathtt{9}, \mathtt{1}, \mathtt{5}, \mathtt{0}, \mathtt{2}, \mathtt{4}, \mathtt{6}],$$

applied to every cell. It has differential uniformity 4 and linearity 8, so in the ordinary setting an active S-box costs at least one bit of correlation; Table 2 shows that this is not the case over $\circ$.

Reduced versions. For $R < 32$ we call *R-round CRAFT starting at round o* the composition $\mathcal{R}_{o+R-1} \circ \cdots \circ \mathcal{R}_o$ with the genuine constants and tweakeys of those absolute rounds. We keep the starting round as a parameter throughout, since the constants differ from window to window; one of the consequences of Section 4.4 is that in the end it makes no difference.

3 Masks over a bibrace

Changing the group changes what counts as a mask, since the admissible masks are the characters of the new group. Exactly half of them are ordinary scalar products and half are quadratic, all the quadratic ones share a single quadratic part, and the two kinds are told apart by a single bit, the *type*.

3.1 The admissible masks

A *mask* is a Boolean function $\ell : V \to \mathbb{F}_2$, and masks are added pointwise. In ordinary linear cryptanalysis the admissible masks are the characters of the difference group $(V, \oplus)$, namely the scalar products $x \mapsto \langle u, x\rangle$. If the difference group is $(V, \circ)$, the admissible masks are the functions with

$$\ell(x \circ y) = \ell(x) \oplus \ell(y) \qquad \text{for all } x, y \in V, \tag{3}$$

and we write $\widehat{V}$ for the set of solutions, the *dual* of $(V, \circ)$. Some admissible masks are quadratic Boolean functions, but they are characters of $(V, \circ)$ exactly as ordinary masks are characters of $(V, \oplus)$, and correlations, trails and the piling-up of round correlations are the usual ones, read in the dual of the other group.

The two groups are isomorphic, and an isomorphism can be written down explicitly. We call $Q : V \to \mathbb{F}_2$ a quadratic form with associated bilinear form ψ if $Q(x \oplus y) = Q(x) \oplus Q(y) \oplus \psi(x, y)$ for all $x, y \in V$.

Lemma 1. *Let Q be a quadratic form with associated bilinear form ψ and $Q(b) = 0$, and set*

$$\theta(x) = x \oplus Q(x)\, b.$$

Then θ is an isomorphism from $(V, \oplus)$ to $(V, \circ)$, and $\theta \circ \theta$ is the identity. Such a Q always exists.

Proof. Since $b \in \mathrm{rad}(\psi)$, adding a multiple of b to either argument does not change ψ, so $\psi(\theta(x), \theta(y)) = \psi(x, y)$ and

$$\theta(x) \circ \theta(y) = x \oplus y \oplus \big(Q(x) \oplus Q(y) \oplus \psi(x, y)\big)b = x \oplus y \oplus Q(x \oplus y)\, b = \theta(x \oplus y).$$

Moreover $Q(\theta(x)) = Q(x) \oplus Q(x)Q(b) \oplus Q(x)\psi(x, b) = Q(x)$, because $Q(b) = 0$ and $\psi(x, b) = 0$; hence $\theta(\theta(x)) = x$, and in particular θ is bijective. For the existence, write $\psi(x, y) = \bigoplus_{i<j} m_{ij}(x_i y_j \oplus x_j y_i)$, which is possible because ψ is alternating, and put $Q_0(x) = \bigoplus_{i<j} m_{ij} x_i x_j$, whose associated bilinear form is ψ. If $Q_0(b) = 1$, choose c with $\langle c, b\rangle = 1$ and replace Q_0 by $Q_0 \oplus \langle c, \cdot\rangle$, which has the same associated bilinear form and vanishes at b. □

For the bibrace B the construction gives $Q(x) = x_0x_3 \oplus x_2x_3 = (x_0 \oplus x_2)x_3$, which already vanishes at $b = \mathtt{5}$, so that

$$\theta(x) = x \oplus (x_0 \oplus x_2)x_3\, b.$$

The masks are now obtained by composing the ordinary ones with θ.

Corollary 1. *Let Q and θ be as in Lemma 1. The admissible masks are the functions*

$$\ell_u(x) \;=\; \langle u, \theta(x)\rangle \;=\; \langle u, x\rangle \;\oplus\; \langle u, b\rangle\, Q(x), \qquad u \in V, \tag{4}$$

and $u \mapsto \ell_u$ is an isomorphism from $(V, \oplus)$ to $(\widehat{V}, \oplus)$. Moreover:

(i) if $\langle u, b\rangle = 0$ then ℓ_u is the scalar product $\langle u, \cdot\rangle$, and if $\langle u, b\rangle = 1$ it is quadratic; hence a scalar product $\langle u, \cdot\rangle$ is admissible if and only if $\langle u, b\rangle = 0$, and exactly half of the admissible masks are scalar products;

(ii) all the quadratic admissible masks have the same quadratic part, that of Q, and differ only by a linear term;

(iii) for every u and all $x, y \in V$,

$$\ell_u(x \oplus y) = \ell_u(x) \oplus \ell_u(y) \oplus \langle u, b\rangle\, \psi(x, y). \tag{5}$$

Proof. Since θ is an involutive isomorphism from $(V, \oplus)$ to $(V, \circ)$, a function ℓ is a character of $(V, \circ)$ if and only if $\ell \circ \theta$ is a character of $(V, \oplus)$, that is $\ell \circ \theta = \langle u, \cdot\rangle$ for a unique u, that is $\ell = \langle u, \theta(\cdot)\rangle$. The rest follows by expanding θ, since $\langle u, \cdot\rangle \oplus Q$ is not affine, its associated bilinear form being $\psi \neq 0$. □

Once Q is fixed, and when Q has no linear terms as for B, the label of a mask is the vector of its linear part, and the distinction between the two kinds of mask is a linear functional of the label.

Definition 1 (type)**.** The *type* of the mask ℓ_u is the bit

$$q(u) := \langle b, u\rangle.$$

A mask of type 0 is called *flat*; it is the scalar product by a vector orthogonal to b. A mask of type 1 is called *curved*; it is the scalar product by a vector not orthogonal to b, plus Q, hence a genuinely quadratic Boolean function. The type is additive, $q(u \oplus v) = q(u) \oplus q(v)$, and it decides, at every point of the cipher, whether a cell is affected by the key addition (Proposition 1) and whether it costs anything at the diffusion layer (Theorem 1).

Applying Corollary 1 to B gives the sixteen masks of Table 1, indexed by $u \in \{\mathtt{0}, \ldots, \mathtt{f}\}$ so that mask addition is the XOR of indices and the type $q(u) = \langle \mathtt{5}, u\rangle$ is the parity of the bits u_0 and u_2.

With the masks determined, the last ingredient is the quantity measuring how well one of them approximates a map.

Definition 2 ($\circ$-LAT)**.** For $S : V \to V$ and $\alpha, \beta \in V$ set

$$W(\alpha, \beta) = \sum_{x \in V} (-1)^{\ell_\alpha(x) \oplus \ell_\beta(S(x))}, \qquad \mathrm{cor}(\alpha, \beta) = 2^{-n} W(\alpha, \beta),$$

and call the resulting table the *$\circ$-linear approximation table* ($\circ$-LAT) of S.

The sum is over V, which as a set is unchanged; only the admissible ℓ differ from the ordinary case. Since the ℓ_α are the characters of an abelian group, this is an ordinary Walsh table and enjoys all the usual properties [Bey21]; equivalently, by Corollary 1, it is the ordinary LAT of $\theta \circ S \circ \theta$.

Table 1: The sixteen admissible masks of the bibrace B, $\ell_u(x) = \langle u, x\rangle \oplus \langle u, b\rangle (x_0x_3 \oplus x_2x_3)$, with the label u in hexadecimal.

u	$\ell_u(x)$	type	u	$\ell_u(x)$	type
0	0	flat	8	x_3	flat
1	$x_0 + x_0x_3 + x_2x_3$	curved	9	$x_0 + x_3 + x_0x_3 + x_2x_3$	curved
2	x_1	flat	a	$x_1 + x_3$	flat
3	$x_0 + x_1 + x_0x_3 + x_2x_3$	curved	b	$x_0 + x_1 + x_3 + x_0x_3 + x_2x_3$	curved
4	$x_2 + x_0x_3 + x_2x_3$	curved	c	$x_2 + x_3 + x_0x_3 + x_2x_3$	curved
5	$x_0 + x_2$	flat	d	$x_0 + x_2 + x_3$	flat
6	$x_1 + x_2 + x_0x_3 + x_2x_3$	curved	e	$x_1 + x_2 + x_3 + x_0x_3 + x_2x_3$	curved
7	$x_0 + x_1 + x_2$	flat	f	$x_0 + x_1 + x_2 + x_3$	flat

3.2 Key and constant additions

Here curved masks behave differently from flat ones. Knudsen and Robshaw [KR96] observed that a nonlinear approximation is destroyed by key addition, because $Q(x \oplus k)$ is not $Q(x)$ plus a constant. That is true here as well, but the discrepancy $\psi(x, k)$ is itself an admissible mask, so a key addition moves an admissible mask to another element of $\widehat{V}$ rather than out of it.

Proposition 1. *For every admissible mask ℓ and every $k \in V$,*

$$\ell(x \oplus k) = \ell(x) \oplus \ell(k) \oplus q\,\psi(x, k),$$

where q is the type of ℓ. If ℓ is flat the last term is absent. If ℓ is curved, the map $x \mapsto \psi(x, k)$ is itself a flat admissible mask, so the mask is translated inside the dual*, by an element of*

$$\Lambda = \{\psi(\cdot, k) \,:\, k \in V\} \;\cong\; V/\mathrm{Ann}.$$

In particular the magnitude of the correlation is unchanged, the type q is invariant, and the translation is trivial if and only if $k \in \mathrm{Ann}$.

Proof. The identity is (5) with $y = k$. The map $x \mapsto \psi(x, k)$ is linear and vanishes at b because $b \in \mathrm{rad}(\psi)$, so by Corollary 1(i) it is a flat admissible mask, and adding a flat mask does not change the type. Finally $\psi(\cdot, k) \equiv 0$ if and only if $k \in \mathrm{rad}(\psi) = \mathrm{Ann}$, and $\Lambda \cong V/\mathrm{Ann}$ because $k \mapsto \psi(\cdot, k)$ is linear with kernel Ann. Replacing x by $x \oplus k$ permutes V, so the absolute value of the Walsh sum is unchanged. □

Remark 1 (conjugation and correlation matrices). By Corollary 1, a circle-linear approximation of a cipher E is an ordinary linear approximation of the conjugated cipher $\theta \circ E \circ \theta$, with θ applied on every cell, so a circle-linear trail is a G-shifted linear trail in the sense of [BCL18, Section 4.1] with $G = \theta$; in the language of [Bey21], the corollary describes the dual of $(V, \circ)$. Specific to the bibrace is the behaviour of the key addition. In the basis of the characters of $(V, \circ)$, Proposition 1 says that the correlation matrix of $x \mapsto x \oplus k$ is a signed permutation matrix, sending ℓ to $\ell \oplus q\,\psi(\cdot, k)$ with the sign $(-1)^{\ell(k)}$. On the states, the key addition of the conjugated cipher is $x \mapsto x \oplus \psi(x, k)\,b \oplus k \oplus Q(k)\,b$, an affine map for every key. For a general conjugating permutation G it is not, and this is what keeps the shifted trails with $G = \theta$ tractable, since the key permutes the masks up to sign instead of spreading each of them over many.

Besides depending on k only through its class modulo Ann, the translation of Proposition 1 does not depend on the mask. By Corollary 1(ii) all curved masks share the same quadratic part, hence the same bilinear form ψ, so one key translates every curved cell of the state by the same element of Λ. For B there are four possible translations, one per

coset of Ann, the trivial one being $k \in \mathrm{Ann}$. All the weak-key conditions of this paper come from this translation. Asking every curved cell to stay in place, that is $k \in \mathrm{Ann}$ on every curved cell, is sufficient for a trail to survive but not necessary; the necessary and sufficient condition is Theorem 1, which looks at a whole column at once and is strictly weaker. Since Ann is a subspace of dimension two inside a nibble of four bits, a constrained nibble costs *two* bits of key and leaves the other two free, so a trail constraining c nibbles holds on 2^{128-2c} of the 2^{128} keys, which for the c between 4 and 10 that we shall meet is a class of 2^{120} down to 2^{108}.

Remark 2 (bibraces of higher rank). That one key translates every curved mask in the same way depends on the rank of the bibrace. Our bibraces are of rank one: the term added to $x \oplus y$ is $\psi(x,y)\,b$, a scalar $\psi(x,y) \in \mathbb{F}_2$ times a fixed vector b. In general a binary bibrace adds $\beta(x,y) = \psi_1(x,y)\,b_1 \oplus \cdots \oplus \psi_d(x,y)\,b_d$, with $b_1, \ldots, b_d$ linearly independent and each in the radical of every ψ_j, and d is its rank. On $\mathbb{F}_2^4$ every binary bibrace has rank one, so for Craft there is no choice to make. For any rank the key addition still permutes the characters up to sign. For $d \geq 2$, however, the quadratic part of ℓ_u is determined by the vector $t = (\langle u, b_1\rangle, \ldots, \langle u, b_d\rangle) \in \mathbb{F}_2^d$, so only a fraction 2^{-d} of the masks is flat, and the curved ones fall into $2^d - 1$ classes, each with its own bilinear form $\psi_t = \bigoplus_i t_i \psi_i$ and its own condition $k \in \mathrm{rad}(\psi_t)$. A key nibble met by curved masks of different classes must then satisfy all their conditions at once. Rank one is thus the only case with a single group Λ of displacements and a single condition $k \in \mathrm{Ann}$, which is why a nibble constrained several times is paid for once.

4 Circle-linear cryptanalysis of Craft

We now push a mask through Craft, component by component, and find out what each operation costs. The first subsection isolates a subgroup $\mathcal{F}$ of the dual, and the search of this paper is confined to masks lying in $\mathcal{F}$. Part of what follows does not need that restriction, since Proposition 1 and Theorem 1 hold for every admissible mask. What changes outside $\mathcal{F}$ is the S-box, which costs a bit per active cell instead of nothing, and with it the arithmetic of a trail. Section 6.2 carries that arithmetic out and finds the designers' own bound.

Inside $\mathcal{F}$ three components of the round cost nothing, the S-box because its relevant entries are $\pm 2^n$, the cell permutation because it only moves masks around, and the key and constant additions because by Proposition 1 they change which mask one measures but not the magnitude of the correlation. Everything is paid at MixColumn, and Theorem 1 says exactly how much and which key nibbles a trail constrains. Both answers are computed from a single quadratic form attached to the column, the first from its rank and the second from its radical.

4.1 The S-box and the free type flip

The $\circ$-LAT of Sb_0 under B is Table 2. It takes only the values $0, \pm 8, \pm 16$, and among the 225 entries with both masks nontrivial, 174 vanish, 48 have absolute value 8 and 3 have absolute value 16. The ordinary LAT of Sb_0, for comparison, takes the values $0, \pm 4, \pm 8$ with multiplicities $93, 96, 36$. Every usable approximation therefore costs exactly one bit or nothing, whereas in the ordinary table the majority cost two.

The following proposition describes its four entries of absolute value 16.

Proposition 2. *Let $\mathcal{F} = \{\alpha : |W(\alpha,\beta)| = 2^n$ for some $\beta\}$ and, for $\alpha \in \mathcal{F}$, let $\varphi(\alpha)$ be that β. Then*

(i) $\mathcal{F}$ is a subgroup of $\widehat{V}$ and φ is an injective endomorphism of it;

Table 2: The $\circ$-LAT of Sb_0 under the bibrace B, with the labels of Table 1 in hexadecimal. A star marks a curved mask.

$\alpha\backslash\beta$	0	1*	2	3*	4*	5	6*	7	8	9*	a	b*	c*	d	e*	f
0	**+16**	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1*	·	−8	+8	·	·	·	·	·	−8	·	·	−8	·	·	·	·
2	·	+8	·	·	+8	·	·	·	−8	·	·	·	·	+8	·	·
3*	·	·	·	+8	·	−8	·	·	·	·	−8	·	−8	·	·	·
4*	·	·	+8	·	+8	·	·	·	·	·	·	+8	·	−8	·	·
5	·	·	·	−8	·	·	+8	·	·	·	−8	·	·	·	·	−8
6*	·	·	·	·	·	+8	+8	·	·	·	·	·	−8	·	·	+8
7	·	·	·	·	·	·	·	·	·	·	·	·	·	·	**−16**	·
8	·	−8	−8	·	·	·	·	·	−8	·	·	+8	·	·	·	·
9*	·	·	·	·	·	·	·	·	·	**+16**	·	·	·	·	·	·
a	·	·	·	−8	·	−8	·	·	·	·	+8	·	−8	·	·	·
b*	·	−8	·	·	+8	·	·	·	+8	·	·	·	·	+8	·	·
c*	·	·	·	−8	·	·	−8	·	·	·	−8	·	·	·	·	+8
d	·	·	+8	·	−8	·	·	·	·	·	·	+8	·	+8	·	·
e*	·	·	·	·	·	·	·	**−16**	·	·	·	·	·	·	·	·
f	·	·	·	·	·	−8	+8	·	·	·	·	·	+8	·	·	+8

(ii) for $\alpha \in \mathcal{F}$ the row α and the column $\varphi(\alpha)$ contain no other nonzero entry;

(iii) for Sb_0 *under B one has $\mathcal{F} = \{0, \ell_7, \ell_9, \ell_e\}$, with $W(\ell_9, \ell_9) = +16$ and $W(\ell_7, \ell_e) = W(\ell_e, \ell_7) = -16$. Writing $\gamma = \ell_9$ and $\eta = \ell_7$, so that $\gamma \oplus \eta = \ell_e$, the masks γ and $\gamma \oplus \eta$ are curved and η is flat, and φ fixes γ and exchanges η with $\gamma \oplus \eta$, so it is an involution.*

Consequently no transition leads from $\mathcal{F}$ to a mask outside it, nor from outside into $\mathcal{F}$.

Proof. An entry of absolute value 2^n means $\ell_\beta \circ S = \ell_\alpha \oplus c$ identically for some $c \in \mathbb{F}_2$, and β is determined by α since S is a bijection; adding two such identities gives a third, which is (i), and $\varphi(\alpha) = 0$ would make ℓ_α constant. Part (ii) is Parseval: an entry of absolute value 2^n already exhausts the budget $\sum_\beta W(\alpha, \beta)^2 = 2^{2n}$ of its row, and of its column since S is a permutation. Part (iii) is obtained from Table 2. □

Part (i) is general. What is specific to Sb_0 and B is (iii): that $\mathcal{F}$ has order four, and in particular that it contains a flat and a curved mask exchanged by φ. Section 4.2 shows how rare that is.

Corollary 2. *For* Sb_0 *under B one has $\Lambda = \{0, \ell_5, \ell_8, \ell_d\}$, so $\mathcal{F} \cap \Lambda = \{0\}$ and $\widehat{V} = \mathcal{F} \oplus \Lambda$.*

Every admissible mask therefore splits uniquely into a part in $\mathcal{F}$ and a part in Λ. By Proposition 1 a key or constant addition adds an element of Λ and nothing else, so it moves the second component and leaves the first untouched, so *the $\mathcal{F}$-component of a mask is an invariant of the key.* On masks in $\mathcal{F}$ the S-box acts deterministically and at correlation one (Proposition 2). A trail in $\mathcal{F}$ is therefore described by the $\mathcal{F}$-components of its masks, which the key cannot change; the key acts only on the Λ-components, and whether MixColumn can absorb them is the weak-key condition of Theorem 1. This is why the search of Section 5 can be exact, and why the key enters the analysis only as a condition and never as a loss of correlation.

Within $\mathcal{F}$, with γ and η as in Proposition 2(iii), we use the coordinates

$$q\,\gamma \oplus s\,\eta, \qquad (q, s) \in \mathbb{F}_2^2,$$

in which q is the type of Definition 1. By Proposition 2(iii) the S-box acts as

$$\sigma(q, s) = (q \oplus s,\ s), \tag{6}$$

Table 3: Census of the 105 bibraces on $\mathbb{F}_2^4$ against the MIDORI/CRAFT S-box. Multiplicities are over the 225 entries with both masks nontrivial. Only four bibraces provide a free type flip, and only one of them, the B used here, has a $\circ$-LAT plateaued on $\{0, 8, 16\}$.

$\lvert W\rvert = 16$ entries	type flip	$\lvert\mathcal{F}\rvert$	count	multiset of $\lvert W\rvert$
3	yes	4	1	$\{0^{174}, 8^{48}, 16^3\}$
3	yes	4	2	$\{0^{126}, 4^{64}, 8^{32}, 16^3\}$
3	yes	4	1	$\{0^{102}, 4^{96}, 8^{24}, 16^3\}$
1	no	2	10	three profiles, all with one entry at 16
0	no	1	91	ten profiles; 21 of them equal to the ordinary LAT

an $\mathbb{F}_2$-linear involution, so the same map describes the action in both directions and can be used when propagating masks backwards. The four elements of $\mathcal{F}$ and their images under φ are as follows.

mask	u	type	q	s	φ(mask)
0	0	flat	0	0	0
γ	9	curved	1	0	γ
η	7	flat	0	1	$\gamma \oplus \eta$
$\gamma \oplus \eta$	e	curved	1	1	η

A cell with $s = 1$ thus changes type at every S-box at correlation one, and this is the *free type flip*. Since a flat cell is unaffected by the key and costs nothing at MixColumn, a cell that flips is free at every other round, and this is what makes long trails possible.

Outside $\mathcal{F}$ the S-box is no longer free. There every nonzero entry of the $\circ$-LAT has absolute value 2^{n-1}, so an active cell costs one bit, which is also the ordinary bound for Sb_0.

4.2 Choice of the bibrace

The attack rests on two properties of the pair (S-box, bibrace): the existence of correlation-one entries, and the presence inside $\mathcal{F}$ of a flat and a curved mask exchanged by φ, which is the free type flip of Section 4.1. Neither is generic, and we quantify how rare they are.

On $\mathbb{F}_2^4$ a bibrace is a pair (ψ, b) with ψ a nonzero alternating form and $0 \neq b \in \mathrm{rad}(\psi)$; there are 35 forms admitting such a b, each with three choices of b, hence 105 bibraces. This elementary count is all we need in dimension four; the systematic classification, up to dimension eight and up to isomorphism, is that of [CF26]. We computed the $\circ$-LAT of Sb_0 for all of them.

Table 3 summarises the outcome. Four bibraces out of 105 give three correlation-one entries together with a free type flip; ten give a single correlation-one entry and no type flip; on CRAFT they are not usable for long trails, since without the flip a cell cannot alternate between the flat type, which the key does not affect, and the curved type; the remaining ninety-one give none at all, and twenty-one of those have a $\circ$-LAT with the same multiset as the ordinary one.

Among the four, only B is plateaued on $\{0, 8, 16\}$: the other three retain between a quarter and nearly half of their entries at level 4. The plateau makes the count outside $\mathcal{F}$ uniform, so B is the convenient choice for that count. However, the evidence here does not fully let us rank the other three against it, as the difference would only show in an analysis carried out outside $\mathcal{F}$, where their level-4 entries cost, and B is the natural place to start such an analysis.

4.3 MixColumn and the key addition

Evaluating a curved mask on a XOR of several cells produces terms $\psi(x_i, x_j)$ involving two distinct cells, which cannot be written as a sum of functions of single cells. These cross terms are the only source of cost, and they arise because CRAFT's MixColumn is not $\circ$-linear: the diffusion layers that are linear with respect to a parallel alternative operation, and hence free, are classified in [CCI26, CF25], and M is not among them. The theorem below gives the exact cost. The key and constant addition that immediately follows MixColumn acts on those same cross terms, and the two must be analysed *together*: treating them separately gives the correct cost but too many weak-key conditions, as quantified in Section 6.1.

Throughout this subsection we work modulo $\mathrm{Ann} = \mathrm{rad}(\psi)$, in explicit coordinates. For $x \in V$ write $p(x) = x_0 \oplus x_2$ and $t(x) = x_3$, so that

$$\psi(x, y) = p(x)\, t(y) \oplus t(x)\, p(y).$$

The map $x \mapsto (p(x), t(x))$ is linear and onto $\mathbb{F}_2^2$, with kernel Ann, so it identifies V/Ann with $\mathbb{F}_2^2$, and $\psi(u, v)$ depends only on the classes of u and v. On V/Ann the form ψ is nondegenerate, meaning that the only class u with $\psi(u, v) = 0$ for every v is the zero class, since $p(u)t(v) \oplus t(u)p(v)$ vanishes for every v only when $p(u) = t(u) = 0$. A column therefore contributes the 8-dimensional space $U = (V/\mathrm{Ann})^4$, with coordinates $p_i = p(x_i)$ and $t_i = t(x_i)$ for $i = 0, \ldots, 3$, on which the cross terms form a quadratic form and the key contributes a linear form.

Theorem 1. *Let $\ell'_0, \ldots, \ell'_3$ be arbitrary admissible masks on the output of a CRAFT MixColumn, let q'_i be their types, and let $k_0, \ldots, k_3$ be the nibbles added to the four cells of the column immediately afterwards, that is round constant plus round tweakey. With $y_0 = x_0 \oplus x_2 \oplus x_3$, $y_1 = x_1 \oplus x_3$, $y_2 = x_2$, $y_3 = x_3$,*

$$\bigoplus_i \ell'_i(y_i \oplus k_i) \;=\; \bigoplus_i \big(M^\top \ell'\big)_i(x_i) \;\oplus\; E(x) \;\oplus\; D(x) \;\oplus\; \mathrm{const},$$

where

$$E(x) = q'_0\big[\psi(x_0, x_2) \oplus \psi(x_0, x_3) \oplus \psi(x_2, x_3)\big] \oplus q'_1\, \psi(x_1, x_3)$$

and

$$D(x) = q'_0\, \psi(y_0, k_0) \oplus q'_1\, \psi(y_1, k_1) \oplus q'_2\, \psi(x_2, k_2) \oplus q'_3\, \psi(x_3, k_3).$$

Regard E as a quadratic form and D as a linear form on U, let $\mathrm{rank}(E)$ and $\mathrm{rad}(E)$ be the rank and the radical of the bilinear form associated with E, and write $D|_{\mathrm{rad}(E)}$ for the restriction of D to $\mathrm{rad}(E)$. Then $\mathrm{rank}(E) = 4(q'_0 + q'_1)$, and the correlation of the layer from the mask $M^\top \ell'$ to the mask ℓ' equals

$$\pm\, 2^{-2(q'_0+q'_1)} \quad \text{if } D|_{\mathrm{rad}(E)} = 0, \qquad 0 \quad \text{otherwise.}$$

In particular the cost is 0, 2, 2, 4 bits according to (q'_0, q'_1) being $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, and it does not depend on the key.

Proof. We use (5) in the form it takes for a sum of m terms: for $\ell \in \widehat{V}$ of type q,

$$\ell\Big(\bigoplus_{j=1}^{m} u_j\Big) = \bigoplus_j \ell(u_j) \oplus q\left(\bigoplus_{j<j'} \psi(u_j, u_{j'})\right). \tag{7}$$

Applying (7) to $\ell'_i(y_i \oplus k_i)$, viewed as a sum over the cells entering y_i together with k_i, splits the exponent into the cell-wise part $\bigoplus_i (M^\top \ell')_i(x_i)$ with $(M^\top \ell')_0 = \ell'_0$, $(M^\top \ell')_1 = \ell'_1$, $(M^\top \ell')_2 = \ell'_0 \oplus \ell'_2$, $(M^\top \ell')_3 = \ell'_0 \oplus \ell'_1 \oplus \ell'_3$; the cross terms among the x's, which are E;

Table 4: The sixteen type patterns of a column, computed exhaustively. The rank and the radical are those of the quadratic form E of Theorem 1 on $U = (V/\mathrm{Ann})^4 \cong \mathbb{F}_2^8$; the cost is $\mathrm{rank}(E)/2$ bits; the fifth column is the number of independent conditions $D|_{\mathrm{rad}(E)} = 0$ places on the key, and the sixth is what the per-cell reading of Corollary 3 would count. Two identities hold in every row, $\mathrm{rank}(E) = 4(q_0' + q_1')$, which is the cost claim, and $\mathrm{rank}(E) + \dim \mathrm{rad}(E) = 8$. In the three rows marked * even the per-cell reading of Corollary 3 overcounts, because the perturbations from k_0 and from k_2, k_3 act on the same coordinates, y_0 containing x_2 and x_3, and can cancel.

$q_0'q_1'q_2'q_3'$	$\mathrm{rank}(E)$	$\dim \mathrm{rad}(E)$	cost (bits)	indep. conditions	per-cell count
0000	0	8	0	0	0
0001	0	8	0	2	2
0010	0	8	0	2	2
0011	0	8	0	4	4
0100	4	4	2	0	0
0101	4	4	2	0	0
0110	4	4	2	2	2
0111	4	4	2	2	2
1000	4	4	2	2	2
1001	4	4	2	2*	4
1010	4	4	2	2*	4
1011	4	4	2	2*	6
1100	8	0	4	0	0
1101	8	0	4	0	0
1110	8	0	4	0	0
1111	8	0	4	0	0

the cross terms between the x's and k_i, which collect into $\psi(y_i, k_i)$ and give D; and the constants $\ell_i'(k_i)$. Rows 2 and 3 contribute nothing to E because M copies them.

Both E and D factor through $U = (V/\mathrm{Ann})^4$ since ψ does. After normalising by the $|\mathrm{Ann}|^4$ states in each class, the correlation is

$$2^{-8} \sum_{u \in U} (-1)^{E(u) \oplus D(u)},$$

and for a quadratic form over $\mathbb{F}_2$ this Gauss sum vanishes unless $E \oplus D$, which is affine on $\mathrm{rad}(E)$, is constant there. On the radical E is additive, and here it vanishes there, as the description of the radical below shows: in each case E restricted to $\mathrm{rad}(E)$ is a sum of terms $\psi(u, u) = 0$. Hence the sum is nonzero precisely when $D|_{\mathrm{rad}(E)} = 0$, and then it has absolute value $2^{8 - \mathrm{rank}(E)/2}$, giving correlation $2^{-\mathrm{rank}(E)/2}$.

It remains to compute $\mathrm{rank}(E)$. If $q_0' = 0$ and $q_1' = 1$ then $E = \psi(x_1, x_3) = p_1 t_3 \oplus t_1 p_3$, a nondegenerate form in the four coordinates of cells 1 and 3 and zero on the other two cells, so $\mathrm{rank}(E) = 4$. If $q_0' = 1$ and $q_1' = 0$ then $E = \bigoplus_{i<j} \psi(x_i, x_j)$ over $i, j \in \{0, 2, 3\}$, whose associated pairing between the p- and t-coordinates is the matrix $J \oplus I$ of size three, of rank 2; hence $\mathrm{rank}(E) = 4$ again, and $\mathrm{rad}(E)$ consists of the vectors whose classes on cells 0, 2 and 3 coincide, cell 1 being arbitrary. If both are one the two forms share only x_3 and the associated pairing is invertible, so $\mathrm{rank}(E) = 8$ and $\mathrm{rad}(E) = 0$. All four values, and the vanishing criterion, were confirmed by exhaustive summation over the 2^{16} states of a column, for each of the 16 type patterns and each of the 2^8 key classes; Table 4 records the rank and the dimension of the radical pattern by pattern, so that the two identities stated in its caption can be checked without redoing the argument. □

Corollary 3. *Write q'_0, q'_1, q'_2, q'_3 for the types of the masks on the four cells $j, j+4, j+8, j+12$ of a column at the point of the key addition. A sufficient condition for the trail to survive with the stated correlation is that $k_i \in \mathrm{Ann}$ for every cell i in the following list:*

$$\textit{row } 0: \ q'_0 \wedge \neg q'_1, \qquad \textit{row } 1: \ \textit{never}, \qquad \textit{row } 2: \ q'_2 \wedge \neg(q'_0 \wedge q'_1), \qquad \textit{row } 3: \ q'_3 \wedge \neg q'_1.$$

In particular a curved cell of row 1 never constrains the key, and a column with $q'_0 = q'_1 = 1$ constrains no key nibble at all.

Proof. Immediate from Theorem 1 by computing $\mathrm{rad}(E)$ in each of the sixteen cases and testing which single-cell perturbations $\psi(\cdot, k_i)$ vanish on it. □

Remark 3 (MixColumn as a correlation matrix). In the language of [Bey21], the proof of Theorem 1 describes one column of the correlation matrix of MixColumn in the basis of characters. Fix the input mask $M^\top \ell'$ and consider the output masks $\ell' \oplus \lambda$, $\lambda \in \Lambda^4$. A displacement λ adds to the exponent the linear form $D_\lambda(x) = \bigoplus_i \lambda_i(y_i)$ on U, so the correlation is $\pm 2^{-\operatorname{rank}(E)/2}$ if D_λ vanishes on $\mathrm{rad}(E)$ and 0 otherwise. Since M is invertible, $\lambda \mapsto D_\lambda$ is injective, and the nonzero entries are the $2^{\operatorname{rank}(E)}$ masks $\ell' \oplus \lambda$ with λ in a subspace $\Lambda_E \subseteq \Lambda^4$ of dimension $\operatorname{rank}(E)$; their squares sum to one, so by Parseval the column has no other nonzero entry.

The column is therefore a branching into $2^{\operatorname{rank}(E)}$ trails of equal weight, and the cost of Theorem 1 is the price of following one of them. The key decides which. It displaces the mask of cell i by $q'_i\,\psi(\cdot, k_i)$, so the branch it brings back to ℓ' is $\ell' \oplus \lambda(k)$, with $\lambda(k) = (q'_i\,\psi(\cdot, k_i))_i$, and this branch has a nonzero entry exactly when $\lambda(k) \in \Lambda_E$, which is the condition of Theorem 1. A trail in $\mathcal{F}$ follows it. The other branches belong to the same linear hull and reach the next S-box outside $\mathcal{F}$ (Section 4.1), and on a key violating the condition every branch does.

4.4 The even row and the round constants

Row 1 of M is the only row that is a sum of an even number of cells, $y_1 = x_1 \oplus x_3$, and by Corollary 3 a curved cell of row 1 never constrains the nibble added to it, whatever the other three masks of the column. The statement needs no further proof. This subsection highlights the mechanism behind it, and then what it implies for the round constants of Craft, which are added precisely to that row.

Suppose first that only row 1 is curved, so that $\ell'_1 = \ell'$ is curved and the other three masks are flat or zero. Then $(M^\top \ell')_1$ and $(M^\top \ell')_3$ both equal ℓ' up to a flat mask, which we may ignore, since a flat mask distributes over a sum of cells with no residue and under the translation below contributes a constant and nothing else. Using (5) twice,

$$\begin{aligned} \ell'(x_1 \oplus x_3 \oplus k) \oplus \ell'(x_1) \oplus \ell'(x_3) &= \ell'(k) \oplus \psi(x_1, x_3) \oplus \psi(x_1, k) \oplus \psi(x_3, k) \\ &= \ell'(k) \oplus \psi(x_1 \oplus k,\ x_3 \oplus k), \end{aligned}$$

because ψ is symmetric with $\psi(k, k) = 0$. Summing over x_1, x_3 and substituting $u = x_1 \oplus k$, $v = x_3 \oplus k$ returns the sum with $k = 0$. The key is absorbed by a translation of the input, and only the sign $(-1)^{\ell'(k)}$ survives.

The parity of the number of summands decides. Translating every x_j entering $y = \bigoplus_{j=1}^{m} x_j$ by k changes $\bigoplus_{j<j'} \psi(x_j, x_{j'})$ by $(m-1)\,\psi(y, k)$, since each index occurs in $m-1$ pairs. For m even this is exactly the key term $\psi(y, k)$, which cancels; for m odd it is zero, and the translation does not help. Craft's MixColumn has $m = 2$ on row 1, $m = 3$ on row 0 and $m = 1$ on rows 2 and 3, so row 1 is the only even one. The argument needs the masks of rows 0 and 3 to be flat, because x_3 also enters y_0 and y_3, and translating it would produce new key terms there; that row 1 is free in every case is the content of Corollary 3.

The freedom of row 1 may look at odds with Propositions 1 and 2. On a curved cell of row 1 the nibble may lie outside Ann, and then by Proposition 1 it translates the mask of

that cell out of $\mathcal{F}$, where by Proposition 2 no transition through the S-box has correlation one. The two statements coexist because the translation is absorbed by MixColumn. The key addition comes right after MixColumn, so the term $\psi(y_1, k) = \psi(x_1, k) \oplus \psi(x_3, k)$ it produces is a function of the MixColumn input, and Theorem 1 accounts for it together with the cross term $\psi(x_1, x_3)$. In the terms of Remark 3, the trail follows the branch $\ell' \oplus \lambda(k)$ of MixColumn, which the key brings back to ℓ', so the next S-box sees a mask in $\mathcal{F}$. The same holds for every curved cell that Corollary 3 leaves free, on row 1 or elsewhere, so along a trail in $\mathcal{F}$ no translated mask ever reaches an S-box.

The round constants. CRAFT adds its round constants to cells 4 and 5, the first two cells of row 1, and to no other cell. This placement has the following consequence.

Proposition 3. *The round constants of CRAFT impose no condition on any circle-linear trail, whatever the window and whatever the values of the constants. This holds for arbitrary admissible masks, inside $\mathcal{F}$ or not.*

Proof. Every constant is added to a cell of row 1, which by Corollary 3 is never constrained. The corollary depends only on the four types of the column, not on where the masks live. □

The constants do have an effect. Like a key nibble, a constant β contributes a factor $(-1)^{\ell(\beta)}$ and so may flip the sign of the correlation, but that sign is computable, since the constants are public, and the magnitude on which a distinguisher rests is untouched.

Two further consequences follow. The first concerns the search. The constants differ from round to round, so any condition they imposed would depend on where the window starts, and the best trail would have to be searched window by window. By Proposition 3 there is no such condition, and the optimum of Section 5 is the same for every window (Proposition 5). The second concerns the design. The constants of CRAFT land precisely, and only, on the one row of the state where Theorem 1 makes an addition provably harmless, and their LFSR-generated values are irrelevant here, since no value of a row-1 nibble can hurt. In MIDORI64 the constants reach every cell and no row of the linear layer is free, each being a sum of three cells, and Section 7.1 shows that then the masks inside $\mathcal{F}$ that cost nothing all fail the key conditions. Placing round constants is therefore not a neutral choice with respect to this family of attacks.

4.5 The tweakey

Unlike the constants, the tweakey is added to all sixteen cells. A curved cell of row 0, 2 or 3 at round r constrains $TK_{r \bmod 4}[i]$ as in Corollary 3. Conditions coming from $r \equiv 0$ and $r \equiv 2$ constrain the same nibble of K_0 with two tweak offsets and are jointly satisfiable if and only if $T[i] \oplus T[\mathcal{Q}(i)] \in \mathrm{Ann}$. Indeed the two conditions read $K_0[i] \oplus T[i] \in \mathrm{Ann}$ and $K_0[i] \oplus T[\mathcal{Q}(i)] \in \mathrm{Ann}$, and since Ann is a subspace their sum is exactly the stated requirement, which involves the tweak alone; given it, either condition implies the other. The same holds for K_1 at rounds $r \equiv 1, 3$.

A nonzero tweak has two consequences. It does not change the *size* of the weak-key class, only which keys it contains, since the conditions become $K_h[i] \in \mathrm{Ann} \oplus T[i]$, a coset of Ann rather than Ann itself, so the count 2^{128-2c} is unaffected. What it can do is make the conditions incompatible, and only for those nibbles that a trail constrains at two residues of the same parity, since only those see two different tweak offsets. A nibble i constrained alone at both residues requires $T[i] \equiv T[\mathcal{Q}(i)] \pmod{\mathrm{Ann}}$, two bits of the tweak, and since $\mathcal{Q}$ is a single 16-cycle a proper subset of these equalities can never close a cycle, so they are independent. Section 6.1 gives their number for our trails. Since the tweak is public and chosen rather than imposed, this restriction costs nothing, because $T = 0$ satisfies every condition.

Remark 4. Let a trail in $\mathcal{F}$ start from a mask with $q = 0$ or $q = s$, as vectors. Each S-box exchanges the two cases by (6), and P and $M^\top$ preserve them, so the state is entirely flat at alternate rounds. Curved cells, and with them all weak-key conditions, then occur at rounds of one parity only, hence on K_0 alone or on K_1 alone, and the other 64 key bits are unconstrained. Every optimal mask of Section 6 is of this kind, and on all of them we verified that each window constrains a single half of the key.

5 Exact trail search inside $\mathcal{F}$

Inside $\mathcal{F}$ the best trail can be computed exactly, by complete enumeration, rather than bounded with a MILP or SAT solver. This rests on facts already established. The S-box maps $\mathcal{F}$ to itself, deterministically and linearly by (6) (Proposition 2), and so does $M^\top$, while on the weak keys the key translations that would take a mask out of $\mathcal{F}$ are absorbed by MixColumn (Section 4.4); the key and the constants cost nothing (Propositions 1 and 3); and MixColumn costs two bits per curved cell in the top half of a column (Theorem 1). A trail is therefore fixed by its 32-bit starting mask, and the best one is a minimum weight codeword of a binary linear code (Proposition 5).

With every cell mask in $\mathcal{F}$ and the coordinates (q, s) of Section 4.1, a state mask is a pair $v = (q, s)$ of vectors in $\mathbb{F}_2^{16}$, one bit per cell each, that is an element of $\mathbb{F}_2^{32}$. The masks quoted in Section 6 are these words in hexadecimal, with q in the low 16 bits and s in the high 16, and bit i of each half belonging to cell i. On the page the leading four digits are therefore s and the trailing four are q, the reverse of the order in the pair. For instance `10050000` has $s =$ `1005` and $q =$ `0000`, so cells 0, 2 and 12 carry η and all the others are zero.

Proposition 4 (Backward round map)**.** *Let v be a state mask all of whose cells lie in $\mathcal{F}$, written in the coordinates above. Let* PN *act on cell indices and let $M^\top$ act on each column as in Theorem 1. Propagating v backwards through one round of* Craft *inside $\mathcal{F}$ amounts to*

$$v \longmapsto \big(M^\top \circ P \circ \sigma\big)(v), \qquad (Pv)[i] = v[\mathrm{PN}(i)],$$

with σ as in (6), applied cell by cell, and the cost of the round is $2\,\#\{i \in \{0,\dots,7\} : q_i = 1\}$ evaluated on the state reached just after P. All three maps are $\mathbb{F}_2$-linear, so the whole trajectory of a mask is an $\mathbb{F}_2$-linear function of the initial vector v.

Proof. The three components are those of Proposition 2 (SubBox), of the cell permutation, and of Theorem 1 (MixColumn), composed in the order opposite to encryption. The SubBox step is (6), which is linear, and so are P and $M^\top$, which act on q and s separately since masks in $\mathcal{F}$ add as their coordinates. By Theorem 1 the cost of a column depends only on the types of its two top cells, and the top half of the state consists of the cells $0,\dots,7$. Key and constant additions contribute no cost by Proposition 1. Composition of linear maps is linear. □

A trail is therefore determined by the 32 bits of v and their linear evolution under the map of Proposition 4. In particular the cost over R rounds is the number of ones among $8R$ linear functions of v, eight per round, and we now stack them into a matrix.

Fix a window of R rounds starting at round o, and let $\mathcal{T}$ denote the linear map of Proposition 4. For a mask v let $u^{(r)} = P\sigma\mathcal{T}^r(v)$ be the state reached just after P in the r-th backward step, $r = 0,\dots,R-1$, and let $q^{(r)} \in \mathbb{F}_2^{16}$ be its vector of types. Both are linear in v. Define $G \in \mathbb{F}_2^{8R\times 32}$ by stacking, for each r, the eight linear forms $v \mapsto q_i^{(r)}$, $i = 0,\dots,7$: one row per cell of the top half, per round. By Proposition 4 the total cost is

$$\mathrm{cost}(v) \;=\; 2\,\mathrm{wt}(Gv), \tag{8}$$

where wt is the Hamming weight. Each column of G is the entire cost history of one basis mask, and G is built once, by propagating the 32 basis masks through the R rounds.

Proposition 5. *The best circle-linear trail of* Craft *inside* $\mathcal{F}$ *over any window of* R *rounds is a minimum weight nonzero codeword of the code* $\{Gv : v \in \mathbb{F}_2^{32}\} \subseteq \mathbb{F}_2^{8R}$*, a binary linear code of length* $8R$ *and dimension at most* 32*. Since* G *does not depend on the round constants, which impose no condition by Proposition 3, the minimum is the same for every window.*

Since $\mathcal{T}$ is invertible and deterministic, a mask in $\mathcal{F}$ determines the entire trail in $\mathcal{F}$ through it, in both directions. Such trails are exactly the segments of the orbits of $\mathcal{T}$, and two of them concatenate only when they are consecutive segments of one orbit. We enumerate the 2^{32} starting masks with a Gray code, so that consecutive masks differ in one coordinate and the codeword is updated by a single vector XOR.

Algorithm 1 (exact search).
Input: R, o.
Output: the minimum cost, and among all masks attaining it one with the largest weak-key class.

1. For each of the 32 basis masks e_k, propagate R rounds by Proposition 4 and record the column $c_k = Ge_k \in \mathbb{F}_2^{8R}$.
2. Split the 32 coordinates into $32 - w$ high and w low bits, $w = 20$. Distribute the 2^{32-w} high blocks over the available threads.
3. In each block, initialise $c \leftarrow \bigoplus_{k \text{ high}} c_k$ and run a Gray code over the low bits: at step i set $c \leftarrow c \oplus c_{\mathrm{ctz}(i)}$, where $\mathrm{ctz}(i)$ is the position of the least significant set bit of i.
4. Evaluate $2\,\mathrm{wt}(c)$ and update the local record.
5. Merge the thread records; then repeat steps 2–4 keeping only the masks of minimum cost, and among those minimise the number of constrained key nibbles, breaking ties by the smallest mask.

The minimum cost value is unique, but it is attained by several masks, and distinct optimal masks do *not* in general have the same weak-key class: at 8 rounds, 24 masks attain the optimum, and among them the number of constrained key nibbles ranges from 4 to 10, a factor of 2^{12} in the size of the class.

6 Results

This section reports what the search returns, how far the distinguishers reach and on how large a class of keys, and how they compare with the published cryptanalysis of Craft.

6.1 Distinguishers

What we exhibit is a correlation that is large on a class of keys. Membership in the class is decided exactly by the criterion of Theorem 1; for a key outside it the trail contributes nothing, and what remains is the rest of the linear hull of the two masks, which we do not compute. A uniform key lies in the class with probability 2^{-2c}, c being the number of constrained nibbles. At eighteen rounds the data requirement 2^{64} is the whole codebook for a fixed tweak, so that length is the limit of the single-tweak distinguisher.

Table 5 reports the optima inside $\mathcal{F}$ with tweak $T = 0$. At every length we computed, exhaustive enumeration finds the optimal masks split evenly into two kinds. Half have $q = 0$ and are entirely flat, the other half have $q = s$, so that every active cell is $\gamma \oplus \eta$, and one backward round of $\mathcal{T}$ carries each mask of the second kind onto a mask of the

Table 5: Exact circle-linear optima for CRAFT inside $\mathcal{F}$, with tweak $T = 0$. The column # opt. counts the masks attaining the optimum; the input and output masks quoted here are those of one representative of that set, selected by the tie-breaking rule of Algorithm 1. The mask codes are those of Section 5.

Rounds	input mask	output mask	# opt.	correlation	squared corr.	data	weak keys
8	725a0000	01190000	24	2^{-8}	2^{-16}	2^{16}	2^{120}
10	45050000	10050000	16	2^{-12}	2^{-24}	2^{24}	2^{118}
12	2b8e0000	10050000	16	2^{-18}	2^{-36}	2^{36}	2^{116}
14	44260000	00720000	16	2^{-22}	2^{-44}	2^{44}	2^{108}
16	8a0a0000	00670000	32	2^{-26}	2^{-52}	2^{52}	2^{110}
18	174d0000	00670000	32	2^{-32}	2^{-64}	2^{64}	2^{110}

first, bijectively by exhaustive check. A curved optimum and its image are therefore the same orbit read one round apart, which explains the even split. The two kinds differ only in which half of the key they constrain (Remark 4). No optimal mask, at any length we computed, contains a cell carrying γ. Such a cell has $q = 1$ and $s = 0$, so its type never changes and it pays at every round it spends in the top half.

The weak-key classes of Table 5 are 2^{128-2c}, with c the number of nibbles constrained by Corollary 3. On these trails this is also the exact class of Theorem 1, obtained by solving the conditions $D|_{\mathrm{rad}(E)} = 0$ of all columns and rounds as one linear system. The three type patterns of Table 4 in which the per-cell reading overcounts do occur, once or twice per trail from ten rounds on, but the conditions they add fall on nibbles that another column already constrains. The cell-by-cell reading of [BCL18, Sections 4.2–4.3], which keeps every curved cell in place, would instead constrain 6, 7, 9, 14, 14 and 14 nibbles at 8 to 18 rounds, against 4, 5, 6, 10, 9 and 9. The cells it constrains needlessly are those of row 1 (Section 4.4) and, from ten rounds on, cells of row 0 in a column whose row 1 is curved.

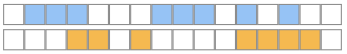
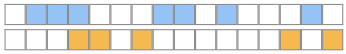
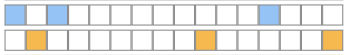
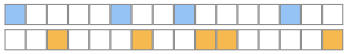
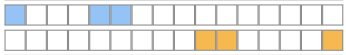
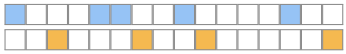
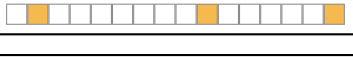
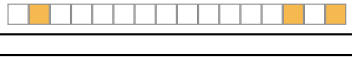
For the same trails the tweak conditions of Section 4.5 concern two nibbles at 8 and 10 rounds, three at 12 and six at 16 and 18. At 14 rounds there are also six, but two of them, nibbles 11 and 15 of K_1, are constrained at one of the two residues only jointly, through a column with three curved cells, and together they give the single condition $T[11] \oplus T[15] \equiv T[\mathcal{Q}(11)] \oplus T[\mathcal{Q}(15)] \pmod{\mathrm{Ann}}$, so the conditions are five. The distinguishers therefore hold for 2^{60} of the 2^{64} tweaks at 8 and 10 rounds, for 2^{58} at 12, for 2^{54} at 14 and for 2^{52} at 16 and 18.

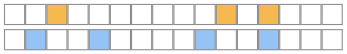
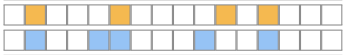
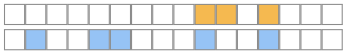
Table 6 shows two optimal 12-round trails, one of each kind, on which the alternation of Remark 4 and the costs and conditions of Theorem 1 can be followed round by round. Appendix B lists all 32 optima at eighteen rounds, the longest length we reach.

6.2 Comparison

Table 7 places our results in context. The best known linear result on CRAFT is the 14-round linear hull of the designers [BLMR19], with expected squared correlation $2^{-62.12}$; Hadipour et al. [HSN+19] report that they could not improve it. At the same 14 rounds the best trail inside $\mathcal{F}$ gives 2^{-44}, and the same method reaches 18 rounds. The comparison is not homogeneous: our result holds on a class of 2^{108} of the 2^{128} keys. Weak-key analyses of CRAFT are however established practice, and they reach further than we do. Guo et al. [GSS+20] give a 20-round weak-key distinguisher with 2^{118} weak keys. Leander and Rasoolzadeh [LR22] exhibit an equivalent Feistel description of CRAFT valid on a sub-class of keys, and derive from it a weak-key single-tweak differential attack on 23 rounds over 2^{112} keys and a related-tweak one on 26 rounds over 2^{108} keys. They also observe that their class weakens CRAFT against linear cryptanalysis, although the attacks they present

Table 6: Two optimal 12-round trails in encryption order, each of cost 18 bits, the flat one with output mask `10050000` at the top and the curved one with output mask `50a150a1` at the bottom. For each round the first grid is the mask before MixColumn and the second the mask at the key point, where cost and key conditions are determined. ☐ inactive, ■ η, ■ $\gamma \oplus \eta$. The mask codes are those of Section 5.

round	mask entering the round	mask at the key point	cost	conditions
0			0	·
1			6	$K_1[3,13,15]$
2			0	·
3			2	$K_1[1,9,15]$
4			0	·
5			4	$K_1[9,15]$
6			0	·
7			0	$K_1[9,10,15]$
8			0	·
9			4	$K_1[9,15]$
10			0	·
11			2	$K_1[1,13,15]$

round	mask entering the round	mask at the key point	cost	conditions
0			6	$K_0[0,12,14]$
1			0	·
2			2	$K_0[2,10,12]$
3			0	·
4			4	$K_0[10,12]$
5			0	·
6			0	$K_0[9,10,12]$
7			0	·
8			4	$K_0[10,12]$
9			0	·
10			2	$K_0[2,12,14]$
11			0	·

are differential. On the linear side they provide a table on the minimum number of active S-boxes, which for Craft is the same for differential and linear patterns because M and $M^\top$ give the same bound. For their weak tweak-key structure it is $2(r-1)$, against $4(r-1)$ for the original cipher. Since an active S-box has absolute correlation at most 2^{-1}, the best *possible* single linear characteristic in their structure has correlation at most $2^{-2(r-1)}$, that is squared correlation 2^{-52} at 14 rounds and 2^{-68} at 18. Our 2^{-44} and 2^{-64} are better at both lengths. The two figures are not of the same kind: theirs are upper bounds derived from an active-S-box count, with no characteristic exhibited, whereas ours are the correlations of explicit trails. What either side gains from the linear hull is unknown, as neither computation has been carried out.

The two weak-key classes are of comparable size, 2^{112} for their single-tweak attack against 2^{108} to 2^{110} for ours, and neither contains the other.

The same comparison can be made with the bound of the designers. Outside $\mathcal{F}$ an active cell costs one bit (Section 4.1), so for a trail lying entirely outside $\mathcal{F}$ their count of active S-boxes applies. Relaxing to the activity pattern alone, sixteen bits, and allowing cancellation at MixColumn whenever two active cells are in the same column gives a lower bound computable by dynamic programming over the 2^{16} patterns. It reproduces the lower bounds on the number of active S-boxes that the designers give for single-tweak trails [BLMR19, Table 5], each active S-box costing at least one bit,

R	8	10	12	14	16	18
entirely outside $\mathcal{F}$, at least	26	36	44	52	60	68
inside $\mathcal{F}$, trail cost	8	12	18	22	26	32

Table 7: Distinguishers and attacks on round-reduced Craft. For linear and circle-linear entries the measure is the expected squared correlation, for differential entries the probability, and for key recoveries the time complexity. ST and RT stand for the single-tweak and the related-tweak model. The circle-linear entries are the best trails inside $\mathcal{F}$ and hold on a class of 2^{108} to 2^{110} of the 2^{128} keys.

type	rounds	measure	keys	source
linear trail (ST)	14	2^{-104}	all	[BLMR19]
linear hull (ST)	14	$2^{-62.12}$	all	[BLMR19]
linear hull (ST)	14	not improved	all	[HSN+19]
circle-linear	14	2^{-44}	2^{108}	this work
circle-linear	16	2^{-52}	2^{110}	this work
circle-linear	18	2^{-64}	2^{110}	this work
differential (ST)	14	$2^{-63.80}$	all	[HSN+19]
truncated diff. (weak key)	20	2^{-63}	2^{118}	[GSS+20]
key recovery (ST)	19	$2^{94.59}$ time	all	[GSS+20]
key recovery (weak key, ST)	23	2^{94} time	2^{112}	[LR22]
key recovery (weak key, RT)	26	2^{105} time	2^{108}	[LR22]
key recovery (ST)	25	$2^{117.58}$ time	all	[MNPN25]
key recovery (ST)	26	2^{118} time	all	[MNPN25]

At 18 rounds an ordinary linear characteristic of the original cipher has correlation at most 2^{-68}, and our trail has 2^{-32}. There is no contradiction, because by Proposition 2 our S-boxes are active but free, and the whole advantage comes from the subgroup $\mathcal{F}$ of order four. The active S-box count, which underlies the security bounds of the designers and of all subsequent analyses, has no purchase on circle-linear trails. Two caveats apply. The upper row prices trails that are outside $\mathcal{F}$ at every cell of every round; a mixed trail, carrying masks in $\mathcal{F}$ on some cells and outside on others, pays nothing for the former and is not covered. We do not treat mixed trails. And the bound counts active S-boxes, which is a statement about single trails. The full $\circ$-LAT differs from the ordinary one even outside $\mathcal{F}$, with larger entries in the $\circ$-one, so a complete linear-hull analysis over the new operation could still diverge from the designers' there. We have not carried it out.

More recently, M'Foukh, Naya-Plasencia and Neumann [MNPN25] applied the state-test technique to truncated-differential meet-in-the-middle cryptanalysis, obtaining single-tweak key recoveries on 24, 25 and 26 of the 32 rounds, valid for every key, the last in time 2^{118} with the full codebook. These are the current state of the art. They are key recoveries rather than distinguishers, and hold for the whole key space, so they are not comparable with the results of this paper in either respect.

Our claim is that on our weak-key class the correlation of the best trail inside $\mathcal{F}$ exceeds the best published linear result for Craft, the all-key linear hull of the designers, by eighteen bits at equal round count. Both concern ordinary linear approximations between fixed masks, since half of the trails we found begin and end flat; ours is the value of a single trail of curved masks, as in [Bey21, Section 5.3], theirs that of a linear hull.

7 Midori and the role of the constants

This section runs the same analysis on Midori64 [BBI+15], a cipher close enough to Craft to share its S-box but on which the masks that cost nothing do not survive the key conditions, and traces the difference to the round constants. Along the way it identifies the nonlinear invariant of [TLS16] as a circle-linear character.

7.1 Contrast with Midori64

Under B, MIDORI64 has the same $\mathcal{F}$, the same type flip and the same Proposition 2 as CRAFT, since it shares the S-box. The two differ in the linear layer and in the constants, and we take the layer first.

Theorem 2 (MIDORI MixColumn)**.** *Let M' be the almost-MDS matrix of MIDORI, $M'_{ij} = 1$ for $i \neq j$ and 0 otherwise, and let $q'_0, \dots, q'_3$ be the types of the masks on the output of a column. The cross term is*

$$E(x) = \bigoplus_{j<k} c_{jk}\, \psi(x_j, x_k), \qquad c_{jk} = \bigoplus_{i \notin \{j,k\}} q'_i,$$

and the exact correlation of the layer is 1 if the column is entirely flat or entirely curved, and 1/4 otherwise. In particular the cost is 0 or 2 bits, never 4.

Proof. As in Theorem 1, expanding $\ell'_i\big(\bigoplus_{j\neq i} x_j\big)$ by (7) produces the cross terms $\psi(x_j, x_k)$ for all pairs $j < k$ with $j, k \neq i$, each weighted by q'_i; summing over i gives the stated coefficient c_{jk}. Because M' has zero diagonal and ones elsewhere, the indices $i \notin \{j, k\}$ are exactly two, so $c_{jk} = q'_{i_1} \oplus q'_{i_2}$ with $\{i_1, i_2\}$ the complement of $\{j, k\}$. If all four types are equal, every c_{jk} vanishes and $E \equiv 0$, whence correlation one; this covers both the all-flat and the all-curved case. Otherwise E is a nonzero sum of copies of ψ on distinct pairs, and evaluating it as in Theorem 1 gives $1/4$ in every remaining case, as confirmed by exhaustive summation over the 2^{16} states of a column. □

Theorem 2 shows that the linear layer of MIDORI is cheaper than that of CRAFT, since a column costs at most two bits against the four of Theorem 1. Combined with (6), the masks carrying η, γ or $\gamma \oplus \eta$ on every cell have a constant type pattern in every column at every round, since the S-box layer exchanges the first and the third and fixes the second, while the cell permutation and M' preserve all three. These three masks therefore cost nothing at any number of rounds, and what protects MIDORI is not the cost but the key conditions.

The key conditions of M' leave no curved cell free. Every row of M' is a sum of three cells, an odd number, so the translation that frees row 1 of CRAFT (Section 4.4) is not available, and an exhaustive check over the 2^8 key classes of a column confirms that every curved cell enters a condition. A column with a single curved cell constrains the nibble added to it, a column with two or three curved cells constrains the XOR of the nibbles added to them, and an all-curved column constrains all four, each condition asking that the nibble, or the XOR, lie in Ann.

MIDORI adds its round constants to every cell, $RK_t = K_{t \bmod 2} \oplus \alpha_t$, where α_t puts one bit on the least significant bit of each cell, so that each constant nibble is `0` or `1`, and $\mathtt{1} \notin$ Ann. Since the same key half returns every two rounds, two rounds t, t' of equal parity with a condition on the same cells constrain the same combination of key nibbles, and the two conditions differ by the XOR of $\alpha_t[i] \oplus \alpha_{t'}[i]$ over those cells, which lies in $\{\mathtt{0}, \mathtt{1}\}$; they are contradictory unless it vanishes. The three masks that cost nothing are curved on every cell at two rounds of equal parity as soon as the window has four rounds, and no two round constants of MIDORI64 coincide, so all three fail. In CRAFT the constants land only on row 1, which is free, and no such contradiction can arise (Proposition 3). What defeats the masks that cost nothing on MIDORI is therefore the combination of rows of odd parity, which here leave no curved cell free, with constants outside Ann that change from round to round. This is the circle-linear counterpart of the design guidance of [BCLR17] on round constants.

The constants do not stop every mask at once. We enumerated all the masks with every cell in $\mathcal{F}$ up to a given cost, starting from round 0. At 6 rounds the cheapest mask satisfying every key condition costs 8 bits, on a class of 2^{96} keys, while at 8 rounds every

mask other than the three above costs at least 16 bits, and we did not enumerate beyond that cost.

Remark 5 (dependence of the protection on the choice of constant). The obstruction just described depends on the bibrace. Since Sb_0 is an involution and has no linear invariant, any bibrace whose correlation-one masks include a curved one has a curved mask g with $g(\mathrm{Sb}_0(x)) = g(x)$, the mask itself if Sb_0 fixes it and the sum of the two masks if Sb_0 exchanges it with a flat one, as in the type flip. A curved mask is a quadratic Boolean function vanishing at 0, so the bibraces that give correlation one can be found among the quadratic invariants of Sb_0. There are fifteen nonzero ones vanishing at 0; eleven have a two-dimensional linear space, and of those exactly seven come from a bibrace, meaning their linear space contains a b with $g(b) = 1$ that generates the group operation. Each of the seven has two such b, and the fourteen bibraces so obtained are those of Table 3 with a correlation-one entry. For all seven the all-curved column crosses the matrix M' of Theorem 2 at correlation *one*, verified by enumeration, so Midori's linear layer blocks none of them and the round constants are the only defence.

Whether they defend depends on where the constant sits, since a constant that takes the values $\mathtt{0}$ and β on a cell, as in Midori, is harmless for an invariant exactly when β lies in its linear space. The seven radicals are

$$\{\mathtt{0},\mathtt{2},\mathtt{5},\mathtt{7}\},\ \{\mathtt{0},\mathtt{5},\mathtt{8},\mathtt{d}\},\ \{\mathtt{0},\mathtt{7},\mathtt{8},\mathtt{f}\},\ \{\mathtt{0},\mathtt{2},\mathtt{8},\mathtt{a}\},\ \{\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{3}\},\ \{\mathtt{0},\mathtt{2},\mathtt{4},\mathtt{6}\},\ \{\mathtt{0},\mathtt{2},\mathtt{d},\mathtt{f}\},$$

and their union is not all of V, since the values $\mathtt{9},\mathtt{b},\mathtt{c},\mathtt{e}$ lie outside every one; a constant taken there would block all seven at once. Midori uses $\mathtt{1}$, which lies in exactly one of them.

With respect to the seven, $\mathtt{1}$ is thus close to the best choice, and with respect to all eleven invariants with a two-dimensional linear space it is among the best. The four that come from no bibrace have radicals $\{\mathtt{0},\mathtt{2},\mathtt{9},\mathtt{b}\}$, $\{\mathtt{0},\mathtt{2},\mathtt{c},\mathtt{e}\}$, $\{\mathtt{0},\mathtt{5},\mathtt{a},\mathtt{f}\}$ and $\{\mathtt{0},\mathtt{7},\mathtt{a},\mathtt{d}\}$, which cover $\mathtt{9},\mathtt{b},\mathtt{c},\mathtt{e}$, so every value of the constant lies in at least one of the eleven radicals. The choice of $\mathtt{1}$ stops the masks of the bibrace of this paper that cost nothing. It does not stop the invariant attached to the one radical it meets, but that invariant is the known one of [TLS16] (Section 7.2). Our $b = \mathtt{5}$ has $\mathtt{2}$ in its radical, so moving the constants by a single bit, from $\mathtt{1}$ to $\mathtt{2}$, would let our masks through. The mask carrying η on every cell, flat at both ends, would then have correlation ± 1 through the whole of Midori64, with its sixteen S-box layers, on the 2^{96} keys $K = K_0 \| K_1$ with every nibble of K_0, the half added at the even rounds, in Ann and K_1 arbitrary. On the differential side the same move gives the probability-one trail of [BFL+23], on a class of the same size. Beyne [Bey18] had found an invariant of the same kind for the radical $\{\mathtt{0},\mathtt{2},\mathtt{8},\mathtt{a}\}$, with constants taken in that radical, again on 2^{96} keys; in our terms it is the type flip of a bibrace with that radical, between the flat mask $\mathtt{5}$ on every cell and a curved one.

7.2 The Todo–Leander–Sasaki invariant as a circle-linear character

The invariant of Remark 5 that the constants of Midori do not stop is the one exhibited by Todo, Leander and Sasaki [TLS16] on Midori64, in the paper that introduced nonlinear invariant attacks. In our terms it is $g(x) = x_2x_3 \oplus x_0 \oplus x_1 \oplus x_2$, whose quadratic part x_2x_3 has as associated bilinear form the form of a bibrace B', distinct from the B used on Craft, namely

$$\psi_{B'}(x,y) = x_2y_3 \oplus x_3y_2, \qquad b = \mathtt{1}, \qquad \mathrm{Ann} = \mathrm{rad}(\psi_{B'}) = \{\mathtt{0},\mathtt{1},\mathtt{2},\mathtt{3}\}.$$

The generator $b = \mathtt{1}$ is the bit the round constant adds, so the constant lies in Ann and the invariant survives it. The invariance $g(\mathrm{Sb}_0(x)) = g(x)$ is a correlation-one relation of B', from the curved mask g to itself, and the invariant of Todo, Leander and Sasaki is

therefore a circle-linear character of B'. Under B' the subgroup of correlation-one masks has order two, $\{0, g\}$, with g curved and fixed by the S-box, so there is no type flip. The same invariant is the shifted linear trail of [BCL18, Section 4.2], whose weak-key set per nibble, $\{0, 1, 2, 3\}$, is the radical of $\psi_{B'}$.

Conclusion

Three features of this work seem to us worth keeping in view. Half of the optimal trails begin and end with flat masks, so the distinguishers are ordinary linear approximations of Craft, and the bibrace serves only to compute their correlation through quadratic masks. Inside $\mathcal{F}$ the S-boxes are active but free, so the count of active S-boxes, on which the security arguments for Craft rest, gives no bound on these trails. And a single language, that of the masks of a bibrace, accounts for the cost of a trail and its exact weak-key class, for the harmlessness of the constants of Craft and the protection given by those of Midori, for the invariants of Midori found by Todo, Leander and Sasaki and by Beyne, and, on the differential side, for the property of Craft found by Guo et al. In that language those invariants are trails of cost zero, each for its own bibrace; the trails on Craft stay in $\mathcal{F}$ as well but pay at MixColumn, and Theorem 1 gives that price and the key condition exactly.

Nonlinear invariants are, in the abstract, a large and unstructured family, and there is no automatic reason for a nonlinear Boolean function to survive an S-box, a linear layer and a key addition together. On the whole state the key addition is the most restrictive of the three. Asking that every key send g to g plus an affine function, as Proposition 1 gives for the masks of a bibrace, means that every derivative $g(x \oplus k) \oplus g(x)$ is affine, which forces g to be at most quadratic; the keys that leave g invariant up to a constant are then those in the radical of its bilinear form, which must be large if the weak-key class is to be large. A bibrace supplies functions of this kind, its curved masks, and in rank one they all share the bilinear form ψ, so a single condition $k \in \mathrm{Ann}$ serves every curved cell (Remark 2).

References

[BBF+25] Jules Baudrin, Christof Beierle, Patrick Felke, Gregor Leander, Patrick Neumann, Léo Perrin, and Lukas Stennes. Commutative cryptanalysis as a generalization of differential cryptanalysis. *Designs, Codes and Cryptography*, 93(8):3243–3281, 2025. doi:10.1007/s10623-025-01625-9.

[BBI+15] Subhadeep Banik, Andrey Bogdanov, Takanori Isobe, Kyoji Shibutani, Harunaga Hiwatari, Toru Akishita, and Francesco Regazzoni. Midori: A block cipher for low energy. In *Advances in Cryptology — ASIACRYPT 2015*, volume 9453 of *Lecture Notes in Computer Science*, pages 411–436. Springer, 2015. doi:10.1007/978-3-662-48800-3_17.

[BCL18] Christof Beierle, Anne Canteaut, and Gregor Leander. Nonlinear approximations in cryptanalysis revisited. *IACR Transactions on Symmetric Cryptology*, 2018(4):80–101, 2018. doi:10.13154/tosc.v2018.i4.80-101.

[BCLR17] Christof Beierle, Anne Canteaut, Gregor Leander, and Yann Rotella. Proving resistance against invariant attacks: How to choose the round constants. In *Advances in Cryptology — CRYPTO 2017*, volume 10402 of *Lecture Notes in Computer Science*, pages 647–678. Springer, 2017. doi:10.1007/978-3-319-63715-0_22.

[Bey18] Tim Beyne. Block cipher invariants as eigenvectors of correlation matrices. In *Advances in Cryptology — ASIACRYPT 2018*, volume 11272 of *Lecture Notes in Computer Science*, pages 3–31. Springer, 2018. doi:10.1007/978-3-030-03326-2_1.

[Bey21] Tim Beyne. A geometric approach to linear cryptanalysis. In *Advances in Cryptology — ASIACRYPT 2021*, volume 13090 of *Lecture Notes in Computer Science*, pages 36–66. Springer, 2021. doi:10.1007/978-3-030-92062-3_2.

[BFL+23] Jules Baudrin, Patrick Felke, Gregor Leander, Patrick Neumann, Léo Perrin, and Lukas Stennes. Commutative cryptanalysis made practical. *IACR Transactions on Symmetric Cryptology*, 2023(4):299–329, 2023. doi:10.46586/tosc.v2023.i4.299-329.

[BLMR19] Christof Beierle, Gregor Leander, Amir Moradi, and Shahram Rasoolzadeh. CRAFT: Lightweight tweakable block cipher with efficient protection against DFA attacks. *IACR Transactions on Symmetric Cryptology*, 2019(1):5–45, 2019. doi:10.46586/tosc.v2019.i1.5-45.

[BLN17] Céline Blondeau, Gregor Leander, and Kaisa Nyberg. Differential-linear cryptanalysis revisited. *Journal of Cryptology*, 30(3):859–888, 2017. doi:10.1007/s00145-016-9237-5.

[BN14] Céline Blondeau and Kaisa Nyberg. Links between truncated differential and multidimensional linear properties of block ciphers and underlying attack complexities. In *Advances in Cryptology — EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 165–182. Springer, 2014. doi:10.1007/978-3-642-55220-5_10.

[BS91] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991. doi:10.1007/BF00630563.

[CBS19] Roberto Civino, Céline Blondeau, and Massimiliano Sala. Differential attacks: using alternative operations. *Designs, Codes and Cryptography*, 87(2–3):225–247, 2019. doi:10.1007/s10623-018-0516-z.

[CCI24] Marco Calderini, Roberto Civino, and Riccardo Invernizzi. Differential experiments using parallel alternative operations. *Journal of Mathematical Cryptology*, 18(1):20230030, 2024. doi:10.1515/jmc-2023-0030.

[CCI26] Marco Calderini, Roberto Civino, and Riccardo Invernizzi. Optimal s-boxes against alternative operations and linear propagation. *Discrete Mathematics*, 349(3):114870, 2026. doi:10.1016/j.disc.2025.114870.

[CCS21] Marco Calderini, Roberto Civino, and Massimiliano Sala. On properties of translation groups in the affine general linear group with applications to cryptography. *Journal of Algebra*, 569:658–680, 2021. doi:10.1016/j.jalgebra.2020.10.034.

[CF25] Roberto Civino and Valerio Fedele. Binary bibraces and applications to cryptography. *Mediterranean Journal of Mathematics*, 22(1):29, 2025. doi:10.1007/s00009-024-02793-z.

[CF26] Roberto Civino and Valerio Fedele. Classification of small binary bibraces via bilinear maps. *Results in Mathematics*, 81(2):53, 2026. doi:10.1007/s00025-026-02612-9.

[CV95] Florent Chabaud and Serge Vaudenay. Links between differential and linear cryptanalysis. In Alfredo De Santis, editor, *Advances in Cryptology — EUROCRYPT '94*, volume 950 of *Lecture Notes in Computer Science*, pages 356–365. Springer, 1995. doi:10.1007/BFb0053450.

[ESS99] Pavel Etingof, Travis Schedler, and Alexandre Soloviev. Set-theoretical solutions to the quantum Yang–Baxter equation. *Duke Mathematical Journal*, 100(2):169–209, 1999. doi:10.1215/S0012-7094-99-10007-X.

[EY19] Muhammad ElSheikh and Amr M. Youssef. Related-key differential cryptanalysis of full round CRAFT. In *Security, Privacy, and Applied Cryptography Engineering (SPACE 2019)*, volume 11947 of *LNCS*, pages 50–66. Springer, 2019. doi:10.1007/978-3-030-35869-3_6.

[GSS+20] Hao Guo, Siwei Sun, Danping Shi, Ling Sun, Yao Sun, Lei Hu, and Meiqin Wang. Differential attacks on CRAFT exploiting the involutory s-boxes and tweak additions. *IACR Transactions on Symmetric Cryptology*, 2020(3):119–151, 2020. doi:10.13154/tosc.v2020.i3.119-151.

[GV17] Leandro Guarnieri and Leandro Vendramin. Skew braces and the Yang–Baxter equation. *Mathematics of Computation*, 86(307):2519–2534, 2017. doi:10.1090/mcom/3161.

[HBS21] Hosein Hadipour, Nasour Bagheri, and Ling Song. Improved rectangle attacks on SKINNY and CRAFT. *IACR Transactions on Symmetric Cryptology*, 2021(2):140–198, 2021. doi:10.46586/tosc.v2021.i2.140-198.

[HSN+19] Hosein Hadipour, Sadegh Sadeghi, Majid M. Niknam, Ling Song, and Nasour Bagheri. Comprehensive security analysis of CRAFT. *IACR Transactions on Symmetric Cryptology*, 2019(4):290–317, 2019. doi:10.13154/tosc.v2019.i4.290-317.

[KR96] Lars R. Knudsen and Matthew J. B. Robshaw. Non-linear approximations in linear cryptanalysis. In *Advances in Cryptology — EUROCRYPT 1996*, volume 1070 of *Lecture Notes in Computer Science*, pages 224–236. Springer, 1996. doi:10.1007/3-540-68339-9_20.

[LP07] Gregor Leander and Axel Poschmann. On the classification of 4 bit s-boxes. In *International workshop on the arithmetic of finite fields*, pages 159–176. Springer, 2007. doi:10.1007/978-3-540-73074-3_13.

[LR22] Gregor Leander and Shahram Rasoolzadeh. Weak tweak-keys for the CRAFT block cipher. *IACR Transactions on Symmetric Cryptology*, 2022(1):38–63, 2022. doi:10.46586/tosc.v2022.i1.38-63.

[MA20] AmirHossein Ebrahimi Moghaddam and Zahra Ahmadian. New automatic search method for truncated-differential characteristics: application to Midori, SKINNY and CRAFT. *The Computer Journal*, 63(12):1813–1825, 2020. doi:10.1093/comjnl/bxaa004.

[Mat94] Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology — EUROCRYPT '93*, volume 765 of *LNCS*, pages 386–397. Springer, 1994. doi:10.1007/3-540-48285-7_33.

[MNPN25] Dounia M'Foukh, María Naya-Plasencia, and Patrick Neumann. The state-test technique on differential attacks: a 26-round attack on CRAFT and other applications. In *Advances in Cryptology — ASIACRYPT 2025*, Lecture Notes

in Computer Science. Springer, 2025. Cryptology ePrint Archive, Report 2025/1597.

[Rum07] Wolfgang Rump. Braces, radical rings, and the quantum Yang–Baxter equation. *Journal of Algebra*, 307(1):153–170, 2007. doi:10.1016/j.jalgebra.2006.03.040.

[TLS16] Yosuke Todo, Gregor Leander, and Yu Sasaki. Nonlinear invariant attack: Practical attack on full SCREAM, iSCREAM and Midori64. In *Advances in Cryptology — ASIACRYPT 2016*, volume 10032 of *Lecture Notes in Computer Science*, pages 3–33. Springer, 2016. doi:10.1007/978-3-662-53890-6_1.

[Wag04] David Wagner. Towards a unifying view of block cipher cryptanalysis. In *Fast Software Encryption — FSE 2004*, volume 3017 of *Lecture Notes in Computer Science*, pages 16–33. Springer, 2004. doi:10.1007/978-3-540-25937-4_2.

A The differential side

This appendix briefly addresses the differential side of the bibrace, where one measures differences in $(V, \circ)$ rather than correlations with its characters. It contains no new differential result. Its purpose is to give an algebraic account of a property of CRAFT found by combinatorial means in [GSS+20], and of part of the clustering observed in [HSN+19].

A.1 The b-fold

Everything below follows from one elementary identity, by which a $\circ$-difference, read as an ordinary XOR-difference, is only defined up to adding b.

Lemma 2 (b-fold)**.** *For every x and every δ,*

$$x \oplus (x \circ \delta) \;=\; \delta \oplus \psi(x, \delta)\, b \;\in\; \{\delta,\ \delta \oplus b\}. \tag{9}$$

As x ranges over V, a fixed $\circ$-difference δ therefore realises the whole set $\{\delta, \delta \oplus b\}$ of XOR-differences, which collapses to the single value δ precisely when $\psi(\cdot, \delta)$ is identically zero, that is when $\delta \in \mathrm{Ann}$. For an individual x the two readings coincide whenever $\psi(x, \delta) = 0$, which happens for other δ as well.

Proof. Immediate from (1). If $\delta \in \mathrm{Ann} = \mathrm{rad}\,\psi$ then $\psi(\cdot, \delta) \equiv 0$; otherwise $\psi(\cdot, \delta)$ is a nonzero linear form, hence balanced, and both values occur equally often. □

Lemma 2 is the differential counterpart of Proposition 1. There a key addition translates a mask inside Λ, here a single $\circ$-difference covers two XOR-differences.

We write $[\delta] = \delta \oplus \mathrm{Ann}$ for the class of δ modulo Ann; the four classes are $[0], [1], [8], [9]$. One has $\mathtt{f} \in [8]$ and $\mathtt{f} \oplus b = \mathtt{f} \oplus 5 = \mathtt{a}$, so $\mathtt{a}$ and $\mathtt{f}$ are the two XOR-representatives of the same $\circ$-difference.

Lemma 3. *Adding a fixed value β to a cell sends the $\circ$-difference $\Delta \mapsto \Delta \oplus \psi(\Delta, \beta)b$. The class $[\Delta]$ is preserved and the shift is deterministic once β is known; it is trivial exactly when $\psi(\Delta, \beta) = 0$, which holds for* every *β if and only if $\Delta \in \mathrm{Ann}$. Consequently, for a uniformly random key nibble, a cell whose difference lies in* Ann *costs nothing and any other cell costs exactly one bit.*

Proof. Let $x \oplus y$ be the XOR-difference of a pair with $\Delta = x \circ y$. Adding β to both leaves $x \oplus y$ unchanged, and recomputing the $\circ$-difference of $x \oplus \beta$ and $y \oplus \beta$ gives the stated shift by bilinearity of ψ. Since $b \in \mathrm{Ann}$ the class is unchanged. For random β the bit $\psi(\Delta, \beta)$ is constant zero when $\Delta \in \mathrm{Ann}$ and balanced otherwise. □

Classes modulo Ann are therefore invariant under AddConstants and AddTweakey, permuted by PermuteNibbles, transformed linearly by MixColumn, and changed only by SubBox.

A.2 The deterministic transition of the S-box

The $\circ$-DDT of Sb_0, whose entry at (δ, δ') counts the x with $\mathrm{Sb}_0(x) \circ \mathrm{Sb}_0(x \circ \delta) = \delta'$, is markedly sparser than the ordinary one, since its nonzero entries take only the values 4, 8 and 16, against the 2 and 4 of the XOR-table outside its trivial row. Besides the trivial row $\delta = 0$, exactly one row has an entry 16, the row $\delta = \mathtt{f}$, so the $\circ$-difference $\mathtt{f}$ is preserved by the S-box with probability one. Nothing of the kind happens over $\oplus$, where Sb_0 is 4-uniform [BBI+15]. This row is the fact behind the properties Guo et al. discovered by other means. We record it, identify which key nibbles let the preservation survive a round, and place it within commutative cryptanalysis.

Proposition 6. Sb_0 *commutes with the* $\circ$*-translation* $T_{\mathtt{f}}$*, that is* $\mathrm{Sb}_0(x \circ \mathtt{f}) = \mathrm{Sb}_0(x) \circ \mathtt{f}$ *for every* $x \in V$*, so the* $\circ$*-differential* $\mathtt{f} \to \mathtt{f}$ *has probability one. Moreover, writing* $\tau_k(x) = \mathrm{Sb}_0(\mathrm{Sb}_0(x) \oplus k)$*,*

$$\tau_k \circ T_{\mathtt{f}} = T_{\mathtt{f}} \circ \tau_k \iff \psi(k, \mathtt{f}) = 0 \iff k \in \{\mathtt{0}, \mathtt{2}, \mathtt{5}, \mathtt{7}, \mathtt{8}, \mathtt{a}, \mathtt{d}, \mathtt{f}\}.$$

Proof. The first statement is verified on the sixteen values of x. The second follows formally. Let $A_k(x) = x \oplus k$. By (1) and bilinearity,

$$T_{\mathtt{f}}(x \oplus k) = x \oplus k \oplus \mathtt{f} \oplus \psi(x, \mathtt{f})b \oplus \psi(k, \mathtt{f})b = A_k\big(T_{\mathtt{f}}(x)\big) \oplus \psi(k, \mathtt{f})b,$$

so $T_{\mathtt{f}}$ commutes with A_k if and only if $\psi(k, \mathtt{f}) = 0$. Since $\tau_k = \mathrm{Sb}_0 \circ A_k \circ \mathrm{Sb}_0$ and Sb_0 commutes with $T_{\mathtt{f}}$ by the first identity, τ_k commutes with $T_{\mathtt{f}}$ if and only if A_k does. Finally $\psi(\cdot, \mathtt{f})$ is a nonzero linear form, so its kernel has eight elements, which are the ones listed. □

Proposition 6 places the differential side of the bibrace precisely inside commutative cryptanalysis in the sense of [Wag04, BFL+23], which looks for affine permutations A, B with $E_k \circ A = B \circ E_k$. A $\circ$-translation is such a map, since by (1),

$$T_\delta(x) = x \circ \delta = \big(x \oplus \psi(x, \delta)b\big) \oplus \delta,$$

and $x \mapsto \psi(x, \delta)b$ is linear because ψ is bilinear, so T_δ is affine over $\oplus$ for every δ. Hence $A = B = T_{\mathtt{f}}$ is an affine commuting pair for Sb_0 holding with probability one, and by the paragraph above it is the only nontrivial translation, for either operation, that commutes with Sb_0.

In this sense, exploiting $(V, \circ)$ in the differential setting is an instance of commutative cryptanalysis, in which the affine maps form an elementary abelian group acting regularly on V, so that one may run an entire differential calculus with them rather than exploit a single commutation. The differential side is developed no further here. We note only that MixColumn copies the $\circ$-differences of rows 2 and 3 and determines those of rows 0 and 1 up to adding b, so that, as on the linear side, only the top half of the state can cost anything.

We also note in passing that the two probability-one entries with $c \neq 0$ recorded in [LR22], namely $c = \alpha = \mathtt{2}$ and $c = \alpha = \mathtt{a}$, are properties of the two-round map τ_c and not of Sb_0 itself, so they do not contradict the above.

A.3 Two published observations

Two facts about CRAFT have been reported in the literature without an algebraic account, an invariant property found by Guo et al. [GSS+20] and the clustering of the cipher's $\oplus$ differentials observed by Hadipour et al. [HSN+19]. The first is a consequence of the b-fold, which also accounts for part of the second; we take them in turn.

The invariant property of Guo et al. Guo et al. [GSS+20] observe that in CRAFT a nibble two rounds apart is computed as $\tau_k(\cdot)$ for a tweakey nibble k, and build weak-tweakey truncated differentials on the fact that τ_k can preserve a difference. Their statements are three.

- *Property 1*. τ_0 preserves every difference.
- *Property 2*. $\tau_{\mathtt{a}}$ preserves the difference $\mathtt{a}$.
- *Lemma 2*. Three *joint* probabilities, each equal to 2^{-2}, in which the S-box maps the difference to $\mathtt{a}$ *and* τ_k maps it to a prescribed value. For $k \in A = \{\mathtt{0},\mathtt{7},\mathtt{a},\mathtt{d}\}$ the $\oplus$-difference $\mathtt{a}$ goes to $\mathtt{a}$, for $k \in B' = \{\mathtt{2},\mathtt{5},\mathtt{8},\mathtt{f}\}$ it goes to $\mathtt{f}$, and for $k \in B'$ the difference $\mathtt{f}$ goes to $\mathtt{a}$.

The key condition of their 20-round weak-key distinguisher, on a class of 2^{118} keys, is

$$K_1[9], K_1[10], K_1[13] \in \{\mathtt{0},\mathtt{a}\}, \qquad K_0[9] \in \{\mathtt{0},\mathtt{2},\mathtt{5},\mathtt{7},\mathtt{8},\mathtt{a},\mathtt{d},\mathtt{f}\}.$$

Proposition 6 accounts for their Lemma 2 and for the condition on $K_0[9]$. Property 1 holds because Sb_0 is an involution, so that τ_0 is the identity, and Property 2 is a direct verification.

Proposition 7. *Let A and B' be the two sets of Lemma 2 of [GSS+20], that is $A = \{\mathtt{0},\mathtt{7},\mathtt{a},\mathtt{d}\}$ and $B' = \{\mathtt{2},\mathtt{5},\mathtt{8},\mathtt{f}\}$. Then A is a subgroup of $(V,\oplus)$, $B' = A \oplus \mathtt{2}$ is its nontrivial coset, and*

$$A \cup B' = \{k : \psi(k,\mathtt{f}) = 0\},$$

which is exactly the commutation set of Proposition 6 and the condition on $K_0[9]$ in the 20-round distinguisher. The splitting of $A \cup B'$ into A and B' corresponds to the two XOR *representatives $\mathtt{a}$ and $\mathtt{f}$ of the single $\circ$-difference $\mathtt{f}$, as given by Lemma 2. It is the coset of k that decides which of the two is reached, which is what their Lemma 2 records as three joint probabilities of 2^{-2}. Furthermore the* XOR*-difference $\mathtt{a}$ is preserved by τ_k deterministically exactly for $k \in \{\mathtt{0},\mathtt{a}\}$, the case $k = \mathtt{a}$ being Property 2 and the case $k = 0$ an instance of Property 1.*

Proof. The subgroup and coset assertions, and the equality with $\{k : \psi(k,\mathtt{f}) = 0\}$, are direct verifications on sixteen elements. The last assertion is again a verification, and so is the middle one, which Lemma 2 explains. Preserving the $\circ$-difference $\mathtt{f}$ means preserving the pair $\{\mathtt{a},\mathtt{f}\}$ setwise, and which of the two is reached is decided by $\psi(\cdot,\mathtt{f})$, that is by $x_0 \oplus x_2$, on the output pair; on the pairs of Lemma 2 of [GSS+20] this value depends only on the coset of k in $A \cup B'$. □

The bibrace supplies a single algebraic statement in place of three case analyses, and it answers the question, left open in [GSS+20], of whether their property can be phrased in the language of invariant attacks [TLS16, BCLR17]: it is the commutation of the S-box with a $\circ$-translation, which is precisely a commutative property in the sense of [Wag04, BFL+23]. Moreover, the bibrace reading does *not* enlarge their weak-key class, since the eight-element condition on $K_0[9]$ is already the one they use, and the two-element condition $\{\mathtt{0},\mathtt{a}\}$ on the nibbles of K_1 is the price of preserving an XOR-difference exactly.

The clustering of the $\oplus$ differentials. Hadipour et al. [HSN+19, Section 4.1] observe that the $\oplus$-differentials of CRAFT cluster strongly, many characteristics of optimal probability sharing the same input and output difference, and that the optimal characteristics of a fixed differential all share one truncated pattern; they report both as observations. To keep their search tractable they restrict the intermediate differences to the set $\{\mathtt{5}, \mathtt{7}, \mathtt{a}, \mathtt{d}, \mathtt{f}\}$, in which every value has a transition of probability 2^{-2} back into the set.

The b-fold accounts for part of this picture. A $\circ$-difference $\delta \notin \mathrm{Ann}$ is the pair of XOR-differences δ and $\delta \oplus b$ (Lemma 2), and since then $\delta \oplus b \neq 0$, two XOR-characteristics that differ by b on cells whose differences lie outside Ann have the same active cells, hence the same truncated pattern. The two heaviest transitions out of $\mathtt{f}$ in the XOR-DDT, $\mathtt{f} \to \mathtt{f}$ and $\mathtt{f} \to \mathtt{a}$ with probability 2^{-2} each, are the two XOR-representatives of the single $\circ$-transition $\mathtt{f} \to \mathtt{f}$ of Proposition 6, since $\mathtt{a} = \mathtt{f} \oplus b$; what has probability one is the $\circ$-transition, not either XOR-branch. We claim nothing quantitative beyond this. How much of the observed clustering these identifications account for is a question for a differential analysis over $\circ$, which we do not carry out, and we leave open whether it can extend the fourteen-round differential of [HSN+19].

B The optima at eighteen rounds

Table 8 lists all 32 masks attaining the optimum at eighteen rounds, by complete enumeration of the 2^{32} masks of $\mathcal{F}$. It is given for reproducibility and because two facts are visible in it that a single representative hides: the number of constrained nibbles is not the same for all optima, sixteen constraining nine and sixteen constraining ten, so one may pick the largest class; and the constrained sets differ from one optimum to another while the cost does not, so each optimum tests a different part of the key.

Table 8: All 32 masks attaining the optimum of 32 bits at 18 rounds. The first sixteen are the flat ones, $q = 0$; the last sixteen have $q = s$. One backward round of $\mathcal{T}$ carries a mask of the second block onto one of the first, bijectively, so the two blocks are the same sixteen orbits read one round apart. Sixteen masks constrain nine key nibbles and sixteen constrain ten: the weak-key class quoted in Table 5 is therefore the largest available at this length and not a figure common to all optima, and the representative quoted there is the first row. The codes are those of Section 5. The 64-bit masks give the label u of each cell (Table 1) as sixteen hexadecimal digits, from cell 0 to cell 15 and grouped by rows of the state, so that `0`, `7`, `9` and `e` stand for 0, η, γ and $\gamma \oplus \eta$. A digit u is the cell mask $\ell_u(x) = \langle u, \theta(x) \rangle$; in the first block every digit is `0` or `7`, flat, so these are ordinary linear masks $\langle u, x \rangle$ of Craft.

code		64-bit mask			
input	output	input	output	cond.	constrained key nibbles
174d0000	00670000	7077 0070 7770 7000	7770 0770 0000 0000	9	$K_1[0, 2, 8, 9, 10, 12, 13, 14, 15]$
8e2b0000	006e0000	7707 0700 0777 0007	0777 0770 0000 0000	9	$K_1[1, 3, 8, 9, 11, 12, 13, 14, 15]$
2b8e0000	009b0000	0777 0007 7707 0700	7707 7007 0000 0000	9	$K_1[1, 3, 9, 10, 11, 12, 13, 14, 15]$
4d170000	009d0000	7770 7000 7077 0070	7077 7007 0000 0000	9	$K_1[0, 2, 8, 10, 11, 12, 13, 14, 15]$
e47c0000	11a50000	0077 7770 0070 0777	7070 0707 7000 7000	10	$K_1[1, 2, 3, 8, 9, 10, 11, 13, 14, 15]$
8e6e0000	152b0000	0777 0770 0777 0007	7707 0700 7070 7000	9	$K_1[1, 3, 8, 9, 11, 12, 13, 14, 15]$
d8bc0000	225a0000	0077 7707 0007 7077	0707 7070 0700 0700	10	$K_1[0, 1, 2, 8, 9, 10, 11, 12, 13, 14]$
4d9d0000	2a170000	7077 7007 7077 0070	7770 7000 0707 0700	9	$K_1[0, 2, 8, 10, 11, 12, 13, 14, 15]$
b1d30000	44a50000	7700 7077 7000 7707	7070 0707 0070 0070	10	$K_1[0, 1, 3, 8, 9, 10, 11, 12, 13, 15]$
2b9b0000	458e0000	7707 7007 7707 0700	0777 0007 7070 0070	9	$K_1[1, 3, 9, 10, 11, 12, 13, 14, 15]$
b1a50000	52d30000	7070 0707 7000 7707	7700 7077 0700 7070	10	$K_1[0, 1, 3, 8, 9, 10, 11, 12, 13, 15]$
e4a50000	587c0000	7070 0707 0070 0777	0077 7770 0007 7070	10	$K_1[1, 2, 3, 8, 9, 10, 11, 13, 14, 15]$
72e30000	885a0000	7700 0777 0700 7770	0707 7070 0007 0007	10	$K_1[0, 2, 3, 8, 9, 10, 11, 12, 14, 15]$
17670000	8a4d0000	7770 0770 7770 7000	7077 0070 0707 0007	9	$K_1[0, 2, 8, 9, 10, 12, 13, 14, 15]$
725a0000	a1e30000	0707 7070 0700 7770	7700 0777 7000 0707	10	$K_1[0, 2, 3, 8, 9, 10, 11, 12, 14, 15]$
d85a0000	a4bc0000	0707 7070 0007 7077	0077 7707 0070 0707	10	$K_1[0, 1, 2, 8, 9, 10, 11, 12, 13, 14]$
9b459b45	552e552e	e0e0 00e0 ee0e e00e	0eee 0e00 e0e0 e0e0	10	$K_0[0, 2, 3, 8, 9, 10, 11, 12, 14, 15]$
6e156e15	558b558b	e0e0 e000 0eee 0ee0	ee0e 000e e0e0 e0e0	10	$K_0[0, 1, 2, 8, 9, 10, 11, 12, 13, 14]$
aa42aa42	671d671d	0e00 00e0 0e0e 0e0e	e0ee e000 eee0 0ee0	10	$K_0[1, 2, 3, 8, 9, 10, 11, 13, 14, 15]$
55245524	6e8b6e8b	00e0 0e00 e0e0 e0e0	ee0e 000e 0eee 0ee0	10	$K_0[0, 1, 2, 8, 9, 10, 11, 12, 13, 14]$
d252d252	73b173b1	0e00 e0e0 0e00 e0ee	e000 ee0e ee00 eee0	9	$K_0[1, 3, 8, 9, 11, 12, 13, 14, 15]$
dc00dc00	78e478e4	0000 0000 00ee e0ee	00e0 0eee 000e eee0	9	$K_0[1, 3, 9, 10, 11, 12, 13, 14, 15]$
55815581	9b2e9b2e	e000 000e e0e0 e0e0	0eee 0e00 ee0e e00e	10	$K_0[0, 2, 3, 8, 9, 10, 11, 12, 14, 15]$
aa18aa18	9d479d47	000e e000 0e0e 0e0e	eee0 00e0 e0ee e00e	10	$K_0[0, 1, 3, 8, 9, 10, 11, 12, 13, 15]$
678a678a	aa1daa1d	0e0e 000e eee0 0ee0	e0ee e000 0e0e 0e0e	10	$K_0[1, 2, 3, 8, 9, 10, 11, 13, 14, 15]$
9d2a9d2a	aa47aa47	0e0e 0e00 e0ee e00e	eee0 00e0 0e0e 0e0e	10	$K_0[0, 1, 3, 8, 9, 10, 11, 12, 13, 15]$
e1a1e1a1	b372b372	e000 0e0e e000 0eee	0e00 eee0 ee00 ee0e	9	$K_0[0, 2, 8, 9, 10, 12, 13, 14, 15]$
ec00ec00	b4d8b4d8	0000 0000 00ee 0eee	000e e0ee 00e0 ee0e	9	$K_0[0, 2, 8, 10, 11, 12, 13, 14, 15]$
73007300	d2b1d2b1	0000 0000 ee00 eee0	e000 ee0e 0e00 e0ee	9	$K_0[1, 3, 8, 9, 11, 12, 13, 14, 15]$
78587858	dce4dce4	000e e0e0 000e eee0	00e0 0eee 00ee e0ee	9	$K_0[1, 3, 9, 10, 11, 12, 13, 14, 15]$
b300b300	e172e172	0000 0000 ee00 ee0e	0e00 eee0 e000 0eee	9	$K_0[0, 2, 8, 9, 10, 12, 13, 14, 15]$
b4a4b4a4	ecd8ecd8	00e0 0e0e 00e0 ee0e	000e e0ee 00ee 0eee	9	$K_0[0, 2, 8, 10, 11, 12, 13, 14, 15]$