SoK: Design-Level Backdoors in Block Ciphers Constructions, Detection, and Key Recovery

Shah Fahd

Military College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan

Abstract

A block cipher is trusted once it resists every published attack and its S-boxes pass the standard tests. A design-level backdoor turns this trust against its users. Its designer hides a structure that only they can exploit, and that the usual tests miss.

Four such backdoors have been published, by Rijmen and Preneel, Paterson, Bannier and Filiol, and Peyrin and Wang. Each paper shows what the hidden structure gives its designer. Everyone else needs answers to two other questions. Does the structure also weaken the cipher against ordinary cryptanalysis? Can an evaluator who holds only the specification find it? Neither has been answered for all four designs, so an S-box that passes the usual tests is trusted by default.

In this paper we answer both questions. We place the four designs in one taxonomy and in Beyne's correlation-matrix framework. We give an algorithm that recovers a partition trapdoor from the S-box tables, and prove that it never reports a false one and, below an explicit cap, never misses one. We show that inversion and affine equivalence preserve a trapdoor, while extended affine and CCZ equivalence can erase it. We then attack KG-DES (Paterson's S-boxes in DES), BEA-1 and LowMC-M.

The answer to the first question depends on the rest of the cipher. In KG-DES the trapdoor makes ordinary attacks far stronger. The best 16-round differential characteristic has probability $2^{-21.23}$, against $2^{-61.97}$ for DES, and a two-stage attack recovers the 56-bit key in minutes. In BEA-1 the trapdoor opens no useful trail. However, for every key both ciphers act as an exact smaller cipher on the cosets, and fall to known-plaintext distinguishers from about $2^{16.5}$ and $2^{20.5}$ texts. An independent MILP model confirms every optimum it can prove, and shows that single characteristics put the KG-DES boomerang at $2^{-36.79}$ when it returns with probability 1.

The answer to the second question depends on where the backdoor sits. Our algorithm finds Paterson's trapdoor and BEA-1's four in the S-boxes that carry them. Rijmen and Preneel's trapdoor needs unbalanced S-boxes, which linear algebra exposes. LowMC-M's MALICIOUS backdoor sits in the linear layers instead. With an ideal XOF a backdoored instance is distributed as an honest one, so no test of the primitive finds it. Its holder recovered the key from 35 to 409 queries in seven full-size measurements.

A deterministic trapdoor in the S-boxes therefore leaves a trace that a sound algorithm finds. A trapdoor in unexplained linear layers leaves none, and only constants derived verifiably from a public seed rule it out. We recommend both as routine checks. Probabilistic trapdoors and trapdoors in only some S-boxes remain open.

Keywords: SoK, block ciphers, design-level backdoors, trapdoors, differential cryptanalysis, key recovery, imprimitive permutation groups, trapdoor subspaces, KG-DES, BEA-1, subspace trails, MALICIOUS, LowMC-M, correlation matrices, MILP

Email: cryptogeek997@gmail.com (Shah Fahd)

Date of this document: 2026-10-21.

1 Introduction

A block cipher's confusion layer is the first thing an evaluator scrutinizes and the last thing a designer is expected to tamper with. Balance, nonlinearity and differential uniformity are easy to check, and an S-box that passes them is normally trusted. A *design-level trapdoor* exploits that trust. A trapdoored S-box weakens the cipher against no published attack and passes the standard statistical tests, yet it carries a structure that only its designer knows, and from that structure a distinguisher, or worse, can be built. Because the weakness lives in the mathematics and not in an implementation, it survives the black-box evaluation a cipher normally undergoes before standardization.

In earlier work [Fah24] we argued that a malicious designer leaves some trace of the backdoor, and that a careful evaluator can find it. We start from this belief. A companion study [FAI+23] measured the S-boxes of the published backdoored designs against a wide battery of cryptanalytic tests. Here we put the same question to whole ciphers, and to four constructions in turn. What does the trapdoor give the one who knows it, and what can an evaluator who holds only the specification find? The four constructions are Rijmen and Preneel's biased trapdoor, the partition trapdoors of Paterson and of Bannier and Filiol, and the MALICIOUS backdoor of Peyrin and Wang. We collect what is known about them, place them in one taxonomy, give a common algorithm for detecting the structural kind, and put each through the same battery of attacks.

Backdoors in the wild. The problem is not only academic. The best-known case is the Clipper chip. The United States government proposed it in 1993 for encrypted telephony and standardized it the following year as the Escrowed Encryption Standard [Nat94]. Its cipher, Skipjack, was kept secret, but the backdoor was declared, not hidden. Every chip escrowed its device key with two government agencies, and every encrypted call carried a Law Enforcement Access Field (LEAF). The LEAF held the session key, wrapped under the chip's escrowed key, together with an identifier and a 16-bit checksum. A warranted agency could recover the session key from the LEAF and decrypt the call. The scheme was defeated by the protocol around Skipjack, not by an attack on Skipjack itself. Blaze [Bla94] showed that the checksum can be satisfied by trying about 2^{16} candidate fields. A user can thus forge a LEAF that the receiving chip accepts but from which the government recovers nothing. The user keeps Clipper's interoperability and loses the escrow. Clipper was abandoned within a few years. Whatever one thinks of it, its backdoor was overt. It was an escrow field added to the protocol, in plain view and open to attack.

The covert kind is harder to see, and it also has a public instance. NIST standardized the Dual EC DRBG random bit generator in 2006 [Nat06]. It produces its output from two elliptic-curve points, P and Q, and the standard never explained where they came from. Shumow and Ferguson [SF07] observed that whoever had chosen Q as a secret multiple of P could recover the generator's internal state from a few dozen output bytes and then predict everything it would produce afterwards. Documents disclosed in 2013 indicated that the constants had been supplied with such a trapdoor in mind, and the generator was withdrawn in 2014 [BLN16]. Nothing in the specification told the backdoored constants from honest ones. The weakness sat in a number, not in a protocol field, and only a verifiable account of where that number came from could have excluded it.

This paper is about backdoors of that second kind. They are built into the mathematics of a design, announced nowhere, and made to survive the evaluation that would expose an escrow field. We take the view from the other side, that of an evaluator, or an attacker, who pulls structure and in the end keys out of a specification that was meant to look clean.

From biased to exact trapdoors. Rijmen and Preneel [RP97] hid a linear approximation behind S-boxes too wide to scan. Their design was broken the following year [WBDY98],

because the bias that makes the trapdoor usable also enters every published S-box row. The other route is to hide an exact relation, with S-boxes that carry the cosets of one linear subspace onto the cosets of another with certainty, while looking statistically ordinary. Harpes and Massey's partitioning cryptanalysis [HM97] exploits partitions of the message space statistically, and Harpes's thesis [Har96] considers ciphers that map one partition onto another for every key. Paterson [Pat99] reached the exact case from a different direction, the permutation group generated by the round functions. If that group is imprimitive, the cipher preserves a partition of the message space for every key. He built a group-theoretic framework of *imprimitive* ciphers on this, and as a worked instance he gave a 32-round DES-like cipher with eight new S-boxes. We study his S-boxes in standard sixteen-round DES and call the result KG-DES. Two decades later Bannier and Filiol [BF17, Ban17] showed that the construction is not tied to the Feistel structure. They built BEA-1, an AES-like SPN whose bijective 10-bit S-boxes carry 5-dimensional trapdoor subspaces, together with a public/secret dual-publication scheme that shows how such a trapdoor might be deployed and hidden. A third line, Peyrin and Wang's MALICIOUS framework [PW20], instantiated as LowMC-M, puts the trapdoor in the linear layers of a tweakable cipher. The partition trapdoors are the centre of this paper. The Rijmen–Preneel family motivates them, and MALICIOUS shows where the methods developed for them stop working.

What is known and what is new. Much of what follows is prior work by others. This includes the constructions themselves and Paterson's proof that his round functions preserve a coset partition for every key. It also includes the break of the Rijmen–Preneel family, Bannier's proof that BEA-1's round function preserves its partition, his key-recovery attack on the deployed cipher, and the published cryptanalysis of LowMC and LowMC-M. Two parts are our own earlier work. The S-box trapdoor detector and its behaviour under the cipher-design equivalences appeared in [FASA23, Fah24], and this paper corrects them, proves them and lifts them to the round function. The S-box-level connectivity profiles first appeared in [FAI+23]. We attribute each result where we use it, and every case study ends with a subsection that separates what it reproduces from what it adds (Sections 6.11, 7.9 and 9.8). The paper contributes the following.

1. *The Rijmen–Preneel family, revisited* (Section 3). We observe that balanced S-boxes cannot carry a trapdoor of this type. The observation is elementary, but it decides the matter. We also read the trapdoor as a single entry of the S-box correlation matrix. A round function without an expansion passes this entry on by exactly one key-independent trail.

2. *Partition trapdoors in general* (Section 4). We relate the partition condition to the Rijmen–Preneel condition and give an exact Fourier characterization of partition maps. We then show that every partition a cipher preserves is a perfect multidimensional linear approximation [Bey21]. As a consequence, we obtain linear approximations of guaranteed correlation and a known-plaintext distinguisher at the birthday bound on the cosets.

3. *A discovery algorithm and its limits* (Section 5). This is our earlier S-box trapdoor detector and its behaviour under the cipher-design equivalences [FASA23], corrected, proved and extended. We fix the candidate-generation step and make it a factor of 2^n cheaper than the criterion it corrects. We prove the algorithm sound and, below an explicit cap on its candidate set, complete. A spectral variant needs no cap. It is sound, and it returns the trapdoor itself under a condition that it can check. For CCZ equivalence we give an elementary bijectivity criterion for block-triangular maps, and we show with a counting argument that an extended affine, hence CCZ, equivalent of a trapdoored S-box can have no trapdoor pair at all. A general closure

computation carries the pairs through a linear layer, in closed form for selection maps and by Gaussian elimination otherwise. Furthermore, we state what the method cannot detect and compare it with the subspace-trail algorithms of Leander, Tezcan and Wiemer [LTW18].

4. *KG-DES* (Section 6). We recover Paterson's trapdoor blind and restate his invariant subspace in the form of differences, which is the form our trails use. We compute the optimal differential characteristic for every round count, $2^{-21.23}$ over 16 rounds against $2^{-61.97}$ for DES by the same search and $2^{-43.87}$ over the 32 rounds of Paterson's own example, and the best rate of any iterative characteristic, 1.415 bits per round. Two independent methods certify the periodic optima. Building on the optimal characteristic, we give a last-round differential key recovery that takes a different route from the partition-based key recoveries of Paterson and Bannier. From about 2^{25} chosen pairs it recovers the coset-label bits of the round subkey, and a short argument shows that it cannot reach the in-coset bits (Proposition 23). With a trail search for each S-box it recovers the full 56-bit master key of the sixteen-round cipher. A second stage peels the last round without its hidden bits (Proposition 24) and recovers the coset-label bits of the round before. This cuts the final search from 2^{24} to 2^{7}. Beyond the key recovery, we find an optimal linear characteristic of correlation 2^{-16} over the full cipher, against $2^{-22.42}$ for DES, which checking the S-boxes one at a time does not reveal. The DLCT reaches its exact maximum, as we prove from the Walsh spectrum. The remaining maxima come from one quadratic trapdoor component per S-box, and Paterson's own design criteria force this component. The FBCT reaches its bound, a boomerang returns with probability 1 over all sixteen rounds, and the round function contains a type-II Rijmen–Preneel relation built from balanced S-boxes, of the kind Rijmen and Preneel anticipated for ciphers with an expansion. Finally, every key gives an exact 32-bit cipher on the cosets, with a known-plaintext distinguisher from about $2^{16.5}$ texts.

5. *BEA-1* (Section 7). We recover all four trapdoor pairs from the secret S-boxes and none from the deployed ones, and we derive the invariant subspace mechanically from the MixColumns matrix. The optimal differential characteristic costs $2^{-36.07}$ over three rounds and $2^{-67.67}$ over four, and every characteristic over all eleven is bounded by $2^{-170.1}$. In the connectivity tables the FBCT vanishes on the trapdoor directions, and the boomerangs we find are generic and owe nothing to the trapdoor. The optimal linear characteristic has correlation 2^{-76}. However, every key gives an exact 40-bit cipher on the cosets, with linear approximations of correlation at least 2^{-20} and a known-plaintext distinguisher from about $2^{20.5}$ texts.

6. *The two partition trapdoors compared* (Section 8). We compute the branch numbers of both linear layers and explain how each came to preserve its partition. A layer chosen without regard to BEA-1's partition would preserve it at maximal branch number with probability below 2^{-186}. An independent MILP cross-check covers every trail type, and we tabulate the data, time and memory of every attack.

7. *The MALICIOUS backdoor* (Section 9). With the XOF modelled as ideal, we prove that a backdoored LowMC-M instance is distributed as an honest one, within 2^{-72} for LowMC-M-128/3, and that recovering its planted tweak pair costs about 2^{k}. We prove that every detection tool of this paper is blind to the backdoor, and we prove the derivative lemma of the LowMC S-box that lies behind its connectivity tables. On full-size instances of all seven parameter sets we measure how far every trail type reaches with and without the backdoor. In the correlation matrix the backdoor is an identity between the matrices of two tweaks, not a block of either.

Where a computation falls short of a proof, we say so. Four summary tables (Tables 35, 51, 52 and 56) record what was checked and how. Tables 50 and 55 give the data, time and memory of every attack, and the appendix describes the computations and how to reproduce them.

Paper organization. The paper follows one line of argument. The Preliminaries collect the definitions and symbols we use in more than one place. Section 3 begins with the Rijmen–Preneel trapdoor and describes how it is built, what it gives its holder, how it was concealed and why it failed. That failure motivates everything after it. It shows that a biased relation cannot stay hidden, and it leaves open whether an exact one can. Section 4 takes up the question with the partition trapdoors of Paterson and of Bannier and Filiol, compares the mechanisms of all four constructions and fixes the terminology. An exact relation should leave an exact trace, and Section 5 turns that observation into a detection algorithm, proves it correct, extends it to the round function and states what it cannot establish. Sections 6 and 7 then run the algorithm and the full attack battery on two real ciphers. KG-DES comes first because it puts the S-boxes of Paterson's own worked example into the best-understood cipher there is. BEA-1 comes second because it removes the two features that made KG-DES easy, the affine coset maps and the linear layer of low branch number. Section 8 compares the two, since the same trapdoor turns out to give an attacker very different things in them. Section 9 asks what happens when the trapdoor leaves the S-boxes altogether and sits in the linear layers of a tweakable cipher. Section 10 places the four constructions among backdoors in other primitives and collects the detection question into a single table. The Conclusion draws the results together, and the appendix describes the experimental methodology.

2 Preliminaries: Notation and Terminology

In this section we collect every definition and symbol that the paper uses in more than one place. Throughout, an S-box is a map $\mathbb{F}_2^m \to \mathbb{F}_2^n$ with input width m and output width n, and $m = n$ when it is bijective. In generic statements we call it π or S. Table 1 at the end of the section lists the notation.

Linear algebra over $\mathbb{F}_2$.

Definition 1 (The field $\mathbb{F}_2$ and the vector space $\mathbb{F}_2^n$)**.** $\mathbb{F}_2 = \{0, 1\}$ is the finite field with two elements, addition given by XOR ($\oplus$) and multiplication by AND. $\mathbb{F}_2^n$ is the set of n-bit strings, viewed as an n-dimensional vector space over $\mathbb{F}_2$ under coordinatewise XOR. Scalar multiplication is trivial ($1 \cdot x = x$, $0 \cdot x = 0$), so "linear combination" over $\mathbb{F}_2^n$ reduces to "XOR of a subset." Every S-box, round input, round output, and key in this paper is an element of some $\mathbb{F}_2^n$, and every difference $x \oplus x'$ between two such elements is itself an element of the same space. For $x, y \in \mathbb{F}_2^n$, $x \cdot y$ denotes the dot product $\bigoplus_i x_i y_i$.

Definition 2 (Subspace, basis, dimension, span)**.** A (linear) subspace $U \subseteq \mathbb{F}_2^n$ is a subset containing 0 and closed under $\oplus$: $u, u' \in U \implies u \oplus u' \in U$. A basis of U is a minimal generating set $\{b_1, \ldots, b_d\} \subset U$ such that every element of U is a unique XOR-combination $\bigoplus_{i \in I} b_i$ for some $I \subseteq \{1, \ldots, d\}$; $d = \dim(U)$ is the dimension, and $|U| = 2^{\dim(U)}$. The span of a set $C \subseteq \mathbb{F}_2^n$, $\mathrm{span}(C)$, is the smallest subspace containing C, equivalently, the set of all XOR-combinations of elements of C. This is the object Algorithm 1 computes incrementally (the running $\mathbb{F}_2$-linear basis $basis[0..n-1]$ of its candidate-generation step, and the subspace U built by WeakBuild and its fallback). The annihilator of U is $U^\perp := \{\lambda : \lambda \cdot u = 0 \text{ for all } u \in U\}$, a subspace of dimension $n - \dim(U)$ with $(U^\perp)^\perp = U$.

Definition 3 (Coset and the partition $\mathcal{L}(U)$)**.** Given a subspace $U \subseteq \mathbb{F}_2^n$ and $x \in \mathbb{F}_2^n$, the coset $x + U := \{x \oplus u : u \in U\}$, also written $x \oplus U$. Two cosets are either identical or disjoint, and together they partition $\mathbb{F}_2^n$ into exactly $2^{n-\dim(U)}$ blocks, each of size $|U|$. Write $\mathcal{L}(U)$ for this partition: the set of all cosets of U. $\mathcal{L}(U)$ is the object every "maps $\mathcal{L}(U)$ to $\mathcal{L}(W)$" statement in this paper refers to.

Definition 4 (Partition-preserving map)**.** A map $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$, with subspaces $U \subseteq \mathbb{F}_2^m$ and $W \subseteq \mathbb{F}_2^n$ of equal dimension, *maps* $\mathcal{L}(U)$ *to* $\mathcal{L}(W)$ if $\{\pi(C) : C \in \mathcal{L}(U)\} = \mathcal{L}(W)$ as sets of sets, i.e. every coset of U is sent, as a whole set, onto exactly one coset of W (not merely into it, and not split across several). This is the condition Check (Algorithm 2, called from Algorithm 1) tests directly, coset by coset. The pair (U, W) is then called a *trapdoor pair* of π, and U and W its input and output trapdoor subspaces.

Definition 5 (Direct sum of embedded subspaces)**.** Given subspaces $U_1 \subseteq \mathbb{F}_2^{n_1}, \ldots, U_k \subseteq \mathbb{F}_2^{n_k}$, embed each U_i into its own disjoint block of n_i coordinates within $\mathbb{F}_2^{n_1+\cdots+n_k}$. Their direct sum $U_1 \oplus \cdots \oplus U_k$ is the resulting subspace of $\mathbb{F}_2^{n_1+\cdots+n_k}$: the set of all strings whose i-th block lies in U_i for every i, simultaneously. Its dimension is $\sum_i \dim(U_i)$. This is how every per-S-box subspace in this paper (each $\mathbf{U}_i/\mathbf{V}_i/\mathbf{W}_i \subset \mathbb{F}_2^6$, $\mathbb{F}_2^4$, or $\mathbb{F}_2^{10}$) is lifted to a round- or block-level subspace of the full state ($\mathbb{F}_2^{32}$ for KG-DES, $\mathbb{F}_2^{80}$ for BEA-1); e.g. $\mathbf{S} = \{0, 2, 4, 6\}^{\oplus 8}$ of Theorem 2 and $\mathbf{T}_0 = \bigoplus_{i=0}^{7} \mathbf{V}_{i \bmod 4}$ of Proposition 28 are both instances of this construction.

Definition 6 (Linear map, matrix, kernel, preimage)**.** A map $M : \mathbb{F}_2^m \to \mathbb{F}_2^n$ is $\mathbb{F}_2$-linear if $M(x \oplus y) = M(x) \oplus M(y)$ for all x, y; equivalently, fixing bases, M is represented by an $n \times m$ matrix over $\mathbb{F}_2$ and computed by matrix-vector multiplication mod 2. A *selection* map is the special case where every output coordinate copies exactly one input coordinate verbatim, with no XOR combination, as DES's E and P and BEA-1's ShiftRows do. A *general* linear map may combine several input coordinates by XOR into one output coordinate, as BEA-1's MixColumns does. The kernel of M is $\ker(M) = \{x : M(x) = 0\}$, itself a subspace. The preimage of a subspace $W \subseteq \mathbb{F}_2^n$ under M is $M^{-1}(W) = \{x : M(x) \in W\}$, also a subspace, and it is the object that Propositions 15 and 28 construct. It is computed by Gaussian elimination over $\mathbb{F}_2$ in the general case (Remark 20) or by direct table lookup in the selection case (Algorithm 3).

Derivatives and truncated differentials.

Definition 7 (Derivative of a vectorial Boolean function)**.** For $f : \mathbb{F}_2^m \to \mathbb{F}_2^n$ and a direction $a \in \mathbb{F}_2^m$, the *derivative of* f *in direction* a is the function

$$D_a f : \mathbb{F}_2^m \to \mathbb{F}_2^n, \qquad D_a f(x) := f(x) \oplus f(x \oplus a).$$

This is the discrete analogue of a directional derivative. It measures how the output of f changes when the input is shifted by a, and every S-box difference $\pi(x) \oplus \pi(x \oplus \alpha)$ in this paper is one of its values. Three special cases recur throughout:

- a is a *linear structure* of f if $D_a f$ is *constant* (independent of x), and a linear structure of the component function $\lambda \cdot f$ if $\lambda \cdot D_a f$ is constant. The original, uncorrected candidate criterion discussed in Section 5.1 is built on the second notion.
- More generally, a satisfies the *truncated* (subspace-valued) version used throughout this paper if $D_a f(x)$, constant or not, is *confined to a fixed proper subspace* W for every x, i.e. $\operatorname{span}\{D_a f(x) : x \in \mathbb{F}_2^m\} \subsetneq \mathbb{F}_2^n$. This is the necessary condition Lemma 1 proves and Algorithm 1's candidate-generation step tests directly.
- A *probability-1 truncated differential* $U \to W$ is a pair of subspaces with $D_u f(x) \in W$ for every x and every $u \in U$. Lemma 3 and its BEA-1 analogue are of this kind. There $D_u S_i(x) \in \mathbf{W}_i$ for every x, with certainty, even though $D_u S_i(x)$ itself is not constant.

Cryptanalytic tables.

Definition 8 (Difference distribution table)**.** For $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$, an input difference $a \in \mathbb{F}_2^m$ and an output difference $b \in \mathbb{F}_2^n$,

$$\mathrm{DDT}_S(a,b) \;:=\; \big|\{x \in \mathbb{F}_2^m : D_a S(x) = b\}\big|,$$

so that the transition $a \to b$ holds with probability $\mathrm{DDT}_S(a,b)/2^m$ over a uniform input. The *differential uniformity* of S is $\max_{a\neq 0,\, b} \mathrm{DDT}_S(a,b)$.

Definition 9 (Walsh transform, linear approximation table, correlation)**.** For $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$, an input mask $\lambda \in \mathbb{F}_2^m$ and an output mask $\gamma \in \mathbb{F}_2^n$,

$$\widehat{S}(\lambda,\gamma) \;:=\; \sum_{x \in \mathbb{F}_2^m} (-1)^{\lambda\cdot x \,\oplus\, \gamma \cdot S(x)},$$

$$\mathrm{LAT}_S(\lambda,\gamma) \;:=\; \big|\{x : \lambda \cdot x = \gamma \cdot S(x)\}\big| - 2^{m-1} \;=\; \tfrac{1}{2}\,\widehat{S}(\lambda,\gamma).$$

The *correlation* of the linear approximation $\lambda \cdot x = \gamma \cdot S(x)$ is $2^{-m}\widehat{S}(\lambda,\gamma) \in [-1,1]$, and the *linear uniformity* of S is $\max_{\lambda,\,\gamma\neq 0} |\mathrm{LAT}_S(\lambda,\gamma)|$. The entries $\widehat{S}(0,\gamma)$ of the zero input mask are the quantity Rijmen and Preneel plant (Proposition 2). Write $\mathrm{supp}_\gamma(S) := \{\lambda : \widehat{S}(\lambda,\gamma) \neq 0\}$.

Definition 10 (Correlation matrix [DGV95])**.** The correlation matrix of $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$ is the $2^n \times 2^m$ matrix C^F with entries $C^F_{\gamma,\lambda} := 2^{-m}\widehat{F}(\lambda,\gamma)$, the correlation of the approximation with input mask λ and output mask γ. Correlation matrices multiply under composition, $C^{G\circ F} = C^G C^F$, and the correlation matrix of a permutation is orthogonal.

Definition 11 (Differential-Linear Connectivity Table)**.** For $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ and $\Delta \in \mathbb{F}_2^m \setminus \{0\}$, $\gamma \in \mathbb{F}_2^n \setminus \{0\}$,

$$\mathrm{DLCT}_\pi(\Delta,\gamma) \;:=\; \big|\{x \in \mathbb{F}_2^m : \gamma \cdot D_\Delta \pi(x) = 0\}\big| \;-\; 2^{m-1}.$$

DLCT uniformity is $\max_{\Delta,\gamma\neq 0} \mathrm{DLCT}_\pi(\Delta,\gamma)$, upper bounded by 2^{m-1}: the count ranges over the 2^m-element *domain*, so it is the input width m, not the output width n, that sets the bound whenever π is not square. For a function E on $\mathbb{F}_2^N$, typically a whole cipher, the *differential-linear correlation* of an input difference Δ and an output mask λ is

$$c_E(\Delta,\lambda) \;:=\; 2^{-N} \sum_{x \in \mathbb{F}_2^N} (-1)^{\lambda\cdot(E(x)\oplus E(x\oplus\Delta))} \in [-1,1];$$

for a single S-box it equals $\mathrm{DLCT}_\pi(\Delta,\lambda)/2^{m-1}$.

Definition 12 (Boomerang and Feistel Boomerang Connectivity Tables)**.** For $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ and distinct nonzero $\Delta, \nabla \in \mathbb{F}_2^m$, the Feistel Boomerang Connectivity Table [BHL+20] is

$$\mathrm{FBCT}_\pi(\Delta,\nabla) \;:=\; \big|\{x : \pi(x) \oplus \pi(x\oplus\Delta) \oplus \pi(x\oplus\nabla) \oplus \pi(x\oplus\Delta\oplus\nabla) = 0\}\big|,$$

a count of vanishing second-order derivatives that needs no bijectivity. The case $\Delta = \nabla$ is excluded because the defining equation then reads $0 = 0$ for every π. For a bijection π of $\mathbb{F}_2^n$ and nonzero $\Delta, \nabla \in \mathbb{F}_2^n$, the Boomerang Connectivity Table [CHP+18] is

$$\mathrm{BCT}_\pi(\Delta,\nabla) \;:=\; \big|\{x : \pi^{-1}(\pi(x)\oplus\nabla) \oplus \pi^{-1}(\pi(x\oplus\Delta)\oplus\nabla) = \Delta\}\big|.$$

The uniformity of either table, its maximum over the admissible arguments, is at most 2^m.

Characteristics, trails and diffusion.

Definition 13 (Characteristic, weight, active S-box, trail)**.** A *differential characteristic* of an iterated cipher fixes the difference entering and leaving every S-box layer; for a Feistel cipher, the input and output differences of every round function. An S-box is *active* when its input difference is nonzero. With independent round keys the probability of a characteristic is taken, as usual, to be the product of $\mathrm{DDT}(\alpha,\beta)/2^m$ over its active S-boxes, and its *weight* is $-\log_2$ of that probability. A *linear characteristic* fixes masks instead; its correlation is the product of the correlations of its active S-boxes, with a key-dependent sign, and its weight is $-\log_2|c|$. The correlation of a linear *approximation*, whose only fixed masks are the outer ones, is the signed sum over every characteristic sharing those masks, the linear hull. *Trail* is used for any such sequence of intermediate properties, including the boomerang and differential-linear trails of Sections 7.6.2 and 6.9.1. For an S-box with a trapdoor pair (U, W), the *trapdoor block* consists of its differential transitions with input difference in U and its linear approximations with input mask in $U^{\perp}$ and output mask in $W^{\perp}$.

Definition 14 (Branch number)**.** For a linear map M on a state made of words (S-box inputs or bundles), the *differential branch number* is $\min_{x\neq 0}\big(w(x)+w(Mx)\big)$ and the *linear branch number* is the same quantity for $M^{\top}$, where w counts the nonzero words; at bit level, w counts the nonzero bits.

Conventions.

Remark 1 (Symbols with more than one role). A few symbols are used in two senses established in the literature, and context rather than typography separates them. P denotes DES's bit permutation, following the DES specification, and also a plaintext block in Definition 18 and in the plaintext pairs (P_i, P_j) of Corollary 3, following the differential-cryptanalysis convention; the permutation always appears applied to a 32-bit argument, the plaintext never does. E denotes DES's expansion map wherever the KG-DES round function is under discussion, and a whole encryption function in the partition-trapdoor definitions and the correlation-matrix results of Section 4 and in the trail analyses. As a cipher it is either subscripted by a key, E_K, or split into parts, $E = E_1 \circ E_m \circ E_0$; the expansion map is never written either way. We keep both, because renaming either would depart from the conventions of the sources against which the results are stated. C denotes a coset in Definitions 3–4, the candidate set of Algorithm 1, and a ciphertext; a correlation matrix always carries a superscript, C^F. L_r is the left half of a KG-DES state, L the number of S-box layers of a reduced BEA-1, and the partition into cosets is always written $\mathcal{L}(U)$. q is the number of S-boxes in a LowMC-M round, and q, q' are also the quotient maps onto the cosets in Section 4.4. In the differential-linear estimate $p\,r\,q^2$ of Section 6.9.1, r is a DLCT entry and q the correlation of the linear part, and in the boomerang estimate q is the probability of the lower characteristic; elsewhere r counts rounds. R is the number of LowMC-M rounds, and R_r, always subscripted, the right half of a KG-DES state.

Remark 2 (A note on notation across ciphers). The letter used for the "input-side" subspace differs by source. It is $\mathbf{U}$ for KG-DES (Section 6) and $\mathbf{V}$ for BEA-1 (Section 7, following Bannier's published notation [Ban17]). Both denote exactly Definition 4's input subspace. We keep the letters distinct only to avoid suggesting that they are the same object in two unrelated ciphers; the underlying concept is the same. The output-side letter is shared, and $\mathbf{W}_i$ is indexed $i = 1,\ldots,8$ for KG-DES and $i = 0,\ldots,3$ for BEA-1.

Table 1: Main notation.

Symbol	Meaning	Defined in
m, n	input and output width of an S-box	this section
$x \cdot y$, $U^{\perp}$	dot product; annihilator of U	Defs. 1, 2
$x + U$, $\mathcal{L}(U)$	coset of U (also $x \oplus U$); partition into the cosets of U	Def. 3
$D_a f$	derivative of f in direction a	Def. 7
DDT_S	difference distribution table	Def. 8
$\widehat{S}$, LAT_S, $\mathrm{supp}_{\gamma}(S)$	Walsh transform, linear approximation table, Walsh support	Def. 9
C^F	correlation matrix	Def. 10
DLCT, BCT, FBCT; c_E	connectivity tables; differential-linear correlation	Defs. 11, 12
$S_1, \ldots, S_8$; π	KG-DES S-boxes; their layer $(S_1, \ldots, S_8)$	Defs. 17, 18
E, P, IP	DES expansion, bit permutation, initial permutation	Def. 18
L_r, R_r	left and right halves of the KG-DES state after r rounds	Def. 18
$\mathbf{U}$, $\mathbf{U}^*$	KG-DES input trapdoor subspace $\{0, 4, 8, 12\}$ and its nonzero part	Def. 19
$\mathbf{W}_i$	output trapdoor subspace of S_i (KG-DES)	Table 13
γ_i^{q}; $\mathbf{N}_i$, $\mathbf{N}_i^0$, d_i	quadratic trapdoor component of S_i; its directions of constant and of zero derivative, $\mathbf{N}_i^0 = \mathbf{U} \cup (d_i \oplus \mathbf{U})$	Prop. 19
$\mathbf{S}$	invariant subspace of KG-DES, nibbles in $\{0, 2, 4, 6\}$	Thm. 2
β, p_T, c_T	Rijmen–Preneel mask, its probability and correlation	Section 3.1
θ; θ_{in}, θ_{out}	a linear layer; the linear maps before and after the S-box layer of a round function	Rem. 20, Prop. 1
C	candidate set of Algorithm 1	Alg. 1
pr_i, $\mathbf{T}_0$	projection onto S-box i's input; round-level candidate subspace	Prop. 15
$S_0, \ldots, S_3$	BEA-1 S-boxes, secret or public	Def. 20
$\mathbf{V}_i$, $\mathbf{W}_i$	input and output trapdoor subspaces of BEA-1's S_i	Table 36
$\mathbf{T}_0$, $\mathbf{W}$	for BEA-1, $\bigoplus_i \mathbf{V}_{i \bmod 4}$ and $\bigoplus_i \mathbf{W}_{i \bmod 4}$ in $\mathbb{F}_2^{80}$	Prop. 28, Lem. 6
X_r, k_r	BEA-1 state and round keys	Def. 20
B_r, B_r^{L}	optimal weight of an r-round differential or linear characteristic of KG-DES	Tables 23, 30
$A(L)$	minimum number of active S-boxes over L layers of BEA-1	Section 7.4
s, q; $x^{(0)}$, $x^{(1)}$	LowMC-M non-linear size and number of S-boxes, $s = 3q$; non-linear and linear part of a state	Section 9
T, t_i, Δt_i; a, R	tweak, sub-tweaks and their differences; number of embedded characteristics, rounds	Defs. 21, 24
$\mathbf{Z}$, H_d	states with zero non-linear part; derivative hyperplane of the LowMC S-box	Section 9, Lem. 8

3 The Rijmen–Preneel Trapdoor

The word *trapdoor* carries two nearly opposite meanings in cryptography. In public-key cryptography a trapdoor is a published, desirable feature. It is a function that is easy to evaluate and hard to invert unless one holds a secret, and the secret is the private key. In symmetric cryptography the same word denotes something covert, a hidden mathematical weakness, planted by the designer, that lets whoever knows it break a cipher that withstands every public form of analysis. This paper is concerned with the second sense, which the more recent literature calls a *backdoor* to keep the two apart [BF17, Ban17, PW20].

The design problem this poses is sharper than it first appears, and every trapdoor proposal since has been judged against it. The designer must satisfy two requirements that pull in opposite directions. The hidden structure must be *exploitable*, strong enough to yield a key recovery or a distinguisher at practical data and time cost. It must also be *undetectable*, invisible to an evaluator who holds the full specification, including every S-box table, and applies the standard battery of tests. The two conflict because anything strong enough to break the cipher is, by definition, a large deviation from random behaviour, and large deviations are what evaluators look for. Every trapdoor design therefore claims, in effect, to have found a place where such a deviation can hide.

In this section we work through the first serious claim of this kind, that of Rijmen and Preneel [RP97], in enough detail to see both the idea and the reason it failed. The failure matters, because it motivates the entirely different route taken by Paterson and by Bannier and Filiol, and hence the rest of this paper. We also put to it the four questions that every later construction will face. What is the hidden object, what can its holder do with it, how is it concealed, and how does an evaluator find it?

3.1 The Construction

Rijmen and Preneel proposed hiding a *linear approximation*. They plant, inside the S-boxes, a linear relation of the kind that Matsui's linear cryptanalysis [Mat94b] searches for, but they place it where the search cannot reach. Rijmen and Preneel adopt Matsui's classification [Mat94a], and in its terms their primary construction uses a *type II* relation. This is a correlation among the *output* bits of the round function alone, not between input and output bits.

The construction operates on an $m \times n$ S-box, written as its table of 2^m output vectors $S = \{v_0, v_1, \ldots, v_{2^m-1}\}$ with $S(x) = v_x$ and $v_i = \langle v_i[1], \ldots, v_i[n]\rangle$. It proceeds in three steps.

1. Choose a nonzero mask $\beta = \langle \beta[1], \ldots, \beta[n]\rangle \in \mathbb{F}_2^n$ and fix a coordinate j^* with $\beta[j^*] = 1$. This β is the trapdoor.

2. Fill every other output coordinate at random: choose $v_i[j]$ uniformly for all i and all $j \neq j^*$.

3. Set the one remaining bit $v_i[j^*]$ of each row so that

$$\beta[1]v_i[1] \oplus \cdots \oplus \beta[n]v_i[n] \;=\; v_i \cdot \beta \;=\; 0 \tag{1}$$

   holds for a p_T fraction of the rows, where p_T is chosen close to 1.

Step 3 is always solvable, and in exactly one way. Since $\beta[j^*] = 1$, Equation (1) determines the remaining coordinate from those already fixed,

$$v_i[j^*] = \bigoplus_{j \neq j^*} \beta[j]\, v_i[j].$$

The S-box therefore satisfies $\beta \cdot S(x) = 0$ with probability p_T over a uniformly random input, equivalently with correlation

$$c_T \;=\; 2p_T - 1$$

between $\beta \cdot S(x)$ and the constant zero function.

Example 1 (A toy instance of Equation (1))**.** The construction is small enough to carry out by hand. Take $m = 2$, $n = 4$, trapdoor $\beta = 1011$, and $j^* = 1$, so that Equation (1) reads $v_i[1] \oplus v_i[3] \oplus v_i[4] = 0$ and step 3 sets $v_i[1] := v_i[3] \oplus v_i[4]$. Drawing coordinates $2, 3, 4$ at random gives the four-row table below; the last row has been left deliberately unrepaired, so that three rows out of four satisfy the relation and $p_T = 3/4$, $c_T = 1/2$.

Table 2: A 2×4 Rijmen–Preneel S-box built to Equation (1) with $\beta = 1011$. Only coordinate $j^* = 1$ is constrained; coordinates $2, 3, 4$ are uniformly random. The worked instance is our own, constructed to illustrate the recipe of [RP97], not taken from that paper.

x	random part $v_x[2..4]$	v_x	$\beta \cdot v_x$
0	101	1101	0
1	011	0011	0
2	110	1110	0
3	001	0001	1 (left unrepaired)

Two features of Table 2 generalize to full size. First, only one bit per row is touched, and since that bit is an XOR of uniformly random bits it is itself uniformly distributed, so no individual output coordinate looks biased. Second, nothing about the table announces which mask was used, since an evaluator sees four random 4-bit strings.

3.2 What the Trapdoor Gives Its Holder

What makes the relation worth planting is that it survives assembly into a round function for free. Rijmen and Preneel instantiated their family on Feistel ciphers of the CAST and LOKI91 type. A CAST-like round function splits its 32-bit input into bytes and XORs four S-box outputs,

$$F(x) \;=\; S_1(x_1) \oplus S_2(x_2) \oplus S_3(x_3) \oplus S_4(x_4), \qquad x = x_1 \| x_2 \| x_3 \| x_4,$$

with each S_i an 8×32 S-box. If all four S-boxes are built with the *same* β (possibly at different $p_T^{(i)}$) then, because $\beta \cdot (\,\cdot\,)$ is linear,

$$\beta \cdot F(x) \;=\; \beta \cdot S_1(x_1) \oplus \beta \cdot S_2(x_2) \oplus \beta \cdot S_3(x_3) \oplus \beta \cdot S_4(x_4),$$

and by the piling-up lemma the whole round function correlates with the constant zero function at

$$c_F \;=\; \prod_{i=1}^{4} c_T^{(i)}.$$

The designer, knowing β, thus starts from a strong linear approximation of the round function and runs ordinary linear cryptanalysis [Mat94b] to recover key material. No diffusion analysis is needed, because the mask passes through the XOR structure untouched.

The trapdoor in the correlation matrix. In the language of correlation matrices (Definition 10), Rijmen and Preneel plant a single entry of an S-box's correlation matrix, $C^S_{\beta,0} = c_T$. It lies in column 0, the column that records imbalance. Since $C^S_{\gamma,0} = 2^{-m}\widehat{S}(0,\gamma)$, a balanced S-box has $C^S_{\gamma,0} = 0$ for every $\gamma \neq 0$ (Proposition 2 below). Figure 1 shows this column for an 8×8 example, before and after planting. In Beyne's classification the trapdoor is an ordinary one-dimensional linear approximation, from the constant function to the character of β, with correlation $c_T < 1$. This is the opposite extreme from the perfect multidimensional approximations that the partition trapdoors of Section 4 produce (Theorem 1). How such an entry reaches the round function is given by the following formula.

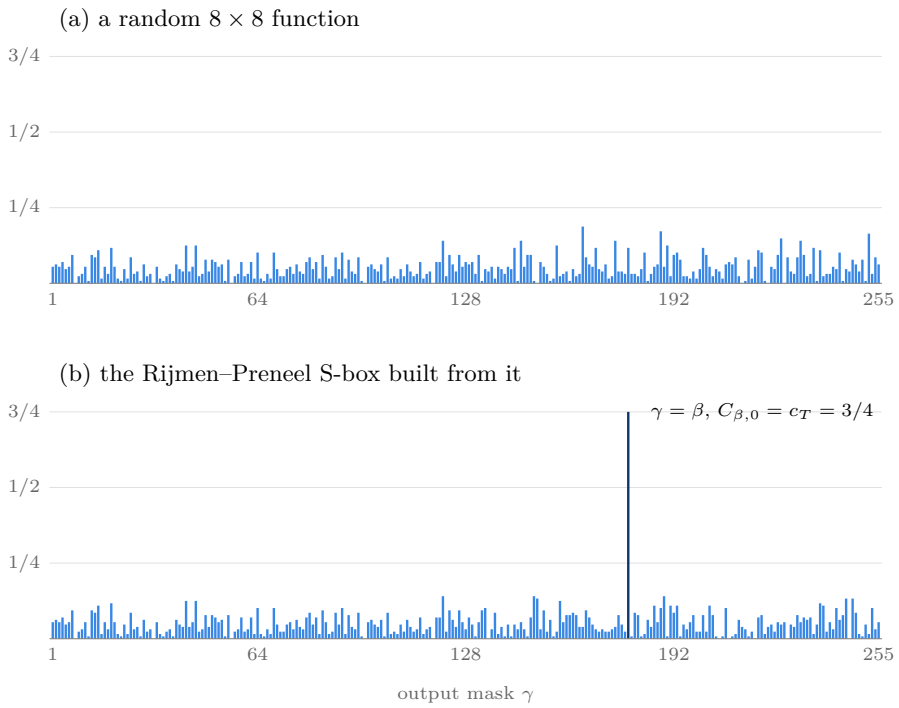

Figure 1: Column 0 of the correlation matrix, $|C^S_{\gamma,0}|$ for the 255 nonzero output masks γ, of (a) a random 8×8 function and (b) the Rijmen–Preneel S-box obtained from it by the recipe of Section 3.1, with $\beta = \texttt{b2}$ and $p_T = 7/8$. Planting raises one entry to $c_T = 3/4$, against at most 0.19 elsewhere in (a) and 0.14 elsewhere in (b). For a permutation, or any balanced S-box such as those of KG-DES and BEA-1, the column is identically 0.

Proposition 1 (Type-II entries of a round function)**.** *Let* $F_K(x) = \theta_{\text{out}}\big(S(\theta_{\text{in}}(x) \oplus K)\big)$*, with* θ_{in} *and* θ_{out} *linear and* S *a layer of parallel S-boxes* S_j*. For every output mask* A*, with* $\gamma = \theta_{\text{out}}^{\top} A$*,*

$$C^{F_K}_{A,0} \;=\; \sum_{\mu\,:\,\theta_{\text{in}}^{\top}\mu = 0} (-1)^{\mu \cdot K} \prod_j C^{S_j}_{\gamma_j,\mu_j}.$$

Proof. C^{F_K} is the product of the correlation matrices of θ_{in}, of the key addition, of S and of θ_{out} (Definition 10). A linear map L has $C^L_{v,u} = 1$ if $u = L^{\top} v$ and 0 otherwise; the key addition $x \mapsto x \oplus K$ is diagonal with entries $(-1)^{u \cdot K}$; and the correlation matrix of a parallel layer is the tensor product of those of its S-boxes, $C^S_{\gamma,\mu} = \prod_j C^{S_j}_{\gamma_j,\mu_j}$. Multiplying out column 0 gives the sum. □

Without an expansion, as in a CAST-type round (θ_{in} the identity, θ_{out} the XOR of the S-box outputs) or an SPN round such as BEA-1's, $\theta_{\text{in}}^{\top}$ has trivial kernel and only $\mu = 0$ remains. Then $C_{A,0}^{F_K} = \prod_j C_{\gamma_j,0}^{S_j}$, a single trail, independent of the key. For Rijmen and Preneel's round, where every $\gamma_j = \beta$, this is $c_F = \prod_i c_T^{(i)}$ exactly, so the piling-up step above is an identity and not an independence assumption. The same holds for any keyed bijection in front of the S-boxes, such as CAST's key-dependent rotations, because the input mask is 0. For any such round with balanced S-boxes, an SPN round such as BEA-1's among them, every product vanishes, and no round function of this shape carries a type-II relation. With an expansion in front of the S-boxes the kernel of $\theta_{\text{in}}^{\top}$ is nontrivial and the sum has further terms, a case Section 6.5.1 meets in KG-DES. Rijmen and Preneel anticipated this case. Their tLOKI variant uses the expansion of LOKI91 to place a trapdoor that is visible only in the round function. They also note that when the key is added after the expansion, as in DES, linear relations on the shared bits give correlations whose magnitude does not depend on the key. They report checking DES for such trapdoors and finding none [RP97, §§3.1 and 4]. Proposition 1 gives the exact correlation.

Rijmen and Preneel's relation is key-independent because it uses $\mu = 0$, and that choice is what forces unbalanced S-boxes, the property Wu et al.'s linear algebra detects (Section 3.4). Iterated over a Feistel cipher, a type-II entry c of the round function gives the two-round iterative characteristic $(\beta, 0) \to (0, \beta) \to (\beta, 0)$ on the masks of (L, R), of correlation c^k over $2k$ rounds. Its key mask is 0, so it yields no key bit directly, and key recovery must guess outer round keys. Figure 2 draws the trail, with one character at every step and a correlation below 1.

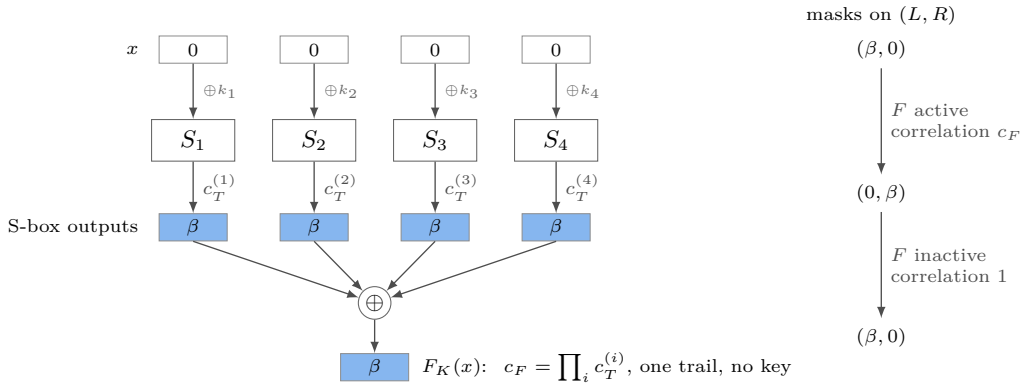

Figure 2: The Rijmen–Preneel trail. Left: the CAST-type round function with its masks written in the cells. The input mask is 0, so the key additions do not enter; every S-box contributes its planted entry $C_{\beta,0}^{S_i} = c_T^{(i)}$, and the XOR of the outputs carries β to each of them, giving $c_F = \prod_i c_T^{(i)}$ by a single trail (Proposition 1). Right: the two-round iteration of the Feistel masks. Each space in the trail is one-dimensional, spanned by a single character, and the correlation is below 1, the opposite extreme from the perfect subspace trails of the partition trapdoors (Section 4.4).

A toy instance confirms the picture. Two 4×8 Rijmen–Preneel S-boxes with the same β and $c_T = 3/4$, in a CAST-type 8-bit round function, give $C_{\beta,0}^{F_K} = 9/16$ for all 256 round keys. On the 16-bit Feistel cipher built from them, the exact correlation of $\beta \cdot L_0 \oplus \beta \cdot L_r$ over all 2^{16} plaintexts equals $9/16$ at $r = 2$ for every key. At $r = 4, 6, 8$ it stays within 0.013 of $(9/16)^{r/2}$ over 40 keys and never changes sign; the other trails, through the 63 further nonzero entries of row β of C^{F_K}, add only small key-dependent terms.

3.3 Concealment

The separation between what is published and what is withheld is the whole of the design.

- *Public:* a conventional Feistel cipher and the complete S-box tables. Every entry is visible. Individual output bits are balanced, and the tables are random apart from one determined coordinate per row.
- *Hidden:* the single mask β, and with it the fact that coordinate j^* of every row is a fixed linear function of the others.

The trapdoor is *key-independent*. It is a property of the tables, so it holds for every key, and the designer needs no per-key setup. This is a feature this family shares with the partition trapdoors studied in the rest of this paper (Theorem 2 is likewise stated for any round key), and it is what makes both families dangerous in practice.

The claimed undetectability rests on one quantitative observation, and the authors phrase it in terms of a specific tool. Verifying a candidate β means measuring the correlation of $\beta \cdot S(x)$. Rijmen and Preneel cost this as one Walsh–Hadamard transform on an m-bit Boolean function, i.e., $O(m\,2^m)$ operations. An evaluator who does not know β must repeat this across $2^n - 1$ candidate masks. From the outset, therefore, detection is posed as a question about the Walsh spectrum. We return to this point in Section 4.3, because it is what makes a precise comparison with the partition family possible.

Their own verdict is not uniform across the parameter sets they discuss, and it is more candid than the construction is sometimes given credit for.

Table 3: Cost of scanning for the planted mask, as Rijmen and Preneel [RP97] themselves assess it, at one Walsh–Hadamard transform per candidate β. They do not claim undetectability at 8×32; the claim is made only for the wider boxes, and the 10×80 setting is the one offered as secure.

Parameter set	$m \times n$	Transforms	Authors' own verdict
CAST-like, 64-bit block	8×32	2^{32}	"feasible and the trapdoor is detectable"
CAST-like, 128-bit block	8×64	2^{64}	"quite hard", 256× a DES key search
Wider S-boxes	10×80	2^{80}	"at present not feasible"

This is a real asymmetry, and it is the insight of the construction. The designer knows where to look and the evaluator does not, and for $n = 64$ or 80 the gap between knowing and searching is enormous. Ordinary S-boxes are narrow (8×8 or smaller) precisely because their tables must be analyzable. Rijmen and Preneel deliberately gave up that analyzability.

3.4 Why It Failed

The construction was broken the following year by Wu, Bao, Deng and Ye [WBDY98], and the reason is instructive well beyond this particular cipher.

The mechanism is that the planted relation is *linear in the unknown*. Equation (1) says $v_i \cdot \beta = 0$, and the v_i are public. Every row of every S-box is therefore a linear equation satisfied by β, and across l S-boxes the evaluator is handed $N \approx l\,2^m$ such equations for free. Recovering β is not a search problem at all but an instance of solving

$$\begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \beta \;=\; \alpha, \qquad \mathrm{wt}(\alpha) \approx N(1 - p_T),$$

a linear system with a small number of errors. Wu et al.'s attack exploits this directly. It picks n rows at random, solves the homogeneous system $v_{i_k} \cdot x = 0$ over $\mathbb{F}_2$, and tests the resulting solution space against the remaining rows. If all n chosen rows happen to be ones that satisfy the relation (probability about p_T^n), then β lies in that space. For the 128-bit CAST variant ($n = 64$, $p_T = 1 - 2^{-5}$) they report $p_T^n \approx 0.1311$, so roughly 32 repetitions recover β with probability about 98.9%. Against the 2^{64} Walsh–Hadamard transforms that Table 3 records for this parameter set, the cost of finding β is negligible.

Remark 3 (What Rijmen and Preneel did and did not anticipate). It would be wrong to present this as an oversight. Rijmen and Preneel set out the linear-algebraic view themselves, in the same section that proposes the construction. They write Equation (1) in the equivalent form $\bigoplus_i \beta[i]\, a^{(i)} = \delta$ with $\mathrm{w}_h(\delta) = 2^m(1 - p_T)$, and observe that "if p_T is very close to one, it is possible to use Gaussian elimination to solve the problem". They identify the general case as *learning a parity function in the presence of noise* and cite the Parity Assumption for its probable NP-hardness. They then analyze the natural attack within that framework, guessing a low-weight δ and solving, and quantify when it beats exhaustive search, concluding for their own 80-bit tCAST parameters that, $\mathrm{w}_h(\delta)$ being 32, "the Gaussian elimination technique to find the trapdoor will not work faster than exhaustive search."

What Wu et al. [WBDY98] contributed was a different strategy inside the same framework, not a different framework. Rather than guessing the error vector δ, they sample n equations and bet that all of them are error-free. This happens with probability p_T^n and needs no knowledge of δ at all. That possibility is what the original analysis does not cover, and it is what makes the parameters chosen for undetectability fail. The very choice $p_T \to 1$ that makes the trapdoor usable also makes an error-free sample likely.

We draw a lesson from this that neither paper states explicitly. A secret that enters the published data *linearly* is constrained by every published entry at once, so its security rests on the noise rate rather than on the size of the space it is drawn from. The partition trapdoors of Section 4 onward differ precisely here. Their defining relation holds with probability 1, so there is no noise to hide behind, and the question becomes whether the structure is visible at all, not whether it can be solved for. Section 4.3 takes that question up directly.

Wu et al. press the point further with a tradeoff argument that rules out repair by reparameterization. Undetectability wants p_T^n small, so that the random-row trick rarely succeeds; exploitability wants $c_T = 2p_T - 1$ large, so that linear cryptanalysis needs few plaintexts. These are the same parameter pulling in opposite directions. They impose a secure round count ($r \geq 8$), an implementable table size ($\frac{n}{m} \cdot n2^m \leq 2^{30}$ bits) and a genuinely hidden trapdoor ($p_T^n \leq 2^{-64}$) at the same time. Under these conditions they compute the data requirement for Matsui's algorithm 2 across m, and they find a minimum on the order of 2^{64} known plaintexts, beyond any practical use. Their conclusion is that the requirements cannot be met together at any parameter setting.

3.5 Balance as an Obstruction

The linear-algebraic break is not the only obstacle. The quantity the construction plants sits in the zero-input-mask row of the linear approximation table, and a single, easily checked property of the S-box rules it out altogether.

Proposition 2 (Balanced S-boxes admit no type-II trapdoor)**.** *Let $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$ have every component function balanced, i.e. $|\{x : \beta \cdot S(x) = 0\}| = 2^{m-1}$ for every $\beta \neq 0$. Then no Rijmen–Preneel type-II relation exists in S. For every nonzero mask β the correlation $c_T = 2p_T - 1$ of Equation (1) is exactly 0. In particular this holds for every DES-style S-box whose four rows are each a permutation of $\{0, \ldots, 15\}$, and hence for all eight KG-DES S-boxes.*

Proof. The quantity Rijmen and Preneel plant is an entry in the zero-input-mask row of the linear approximation table (Definition 9):

$$\mathrm{LAT}_S(0,\beta) \;=\; \big|\{x : \beta \cdot S(x) = 0\}\big| - 2^{m-1},$$

so $p_T = \frac{1}{2} + 2^{-m}\mathrm{LAT}_S(0,\beta)$ and $c_T = 2^{1-m}\mathrm{LAT}_S(0,\beta)$. Balance of the component function $\beta \cdot S$ is precisely the statement $\mathrm{LAT}_S(0,\beta) = 0$, giving $c_T = 0$. For a DES-style S-box each of the four rows is a permutation of $\mathbb{F}_2^4$, so each output value occurs exactly 4 times over the 64 inputs; every component function is therefore balanced and the hypothesis holds. □

Two structural consequences follow, and together they make the Rijmen–Preneel mechanism and the partition trapdoors of the following sections mutually exclusive at the S-box level of a DES-style architecture. The round function is another matter. An expansion can assemble a type-II relation of the round function out of balanced S-boxes, with a key-dependent sign (Proposition 1 and Section 6.5.1).

First, the Rijmen–Preneel mechanism requires an *unbalanced* component function, and the DES-style construction rules this out by design. A designer who wanted both would have to abandon the row-permutation structure, and imbalance is detectable in $O(2^m)$ time by counting.

Second, the concealment argument does not survive the change of scale. Rijmen and Preneel hid β behind an uncomputable linear approximation table, which needs an output width n of 64 or more (Table 3). The tables of the ciphers studied below, 64×16 for KG-DES and 1024×1024 for BEA-1, are computed in milliseconds, so a planted approximation of *either* type, output-only or input–output, would simply be read off. Width is what buys Rijmen–Preneel secrecy, and neither of those ciphers has it.

3.6 What the Family Leaves Open

We record three limitations of the family, because they define the space in which the rest of this paper works.

First, the attack that the trapdoor enables is statistical and partial. It is linear cryptanalysis with a known good approximation, which recovers part of the key at a data cost governed by c_F. Nothing in the construction is exact.

Second, the same authors proposed a second variant based on *type I* relations, correlations between input and output bits of the round function, and Wu et al. report that it falls to the same attack, for the same reason.

Third, and most relevant here, Rijmen and Preneel also suggested hiding *differentials* instead of linear relations, so as to make a cipher weak against differential rather than linear cryptanalysis. They gave no construction, and they say why they did not pursue one: “exploitation of such trapdoors requires chosen rather than known plaintexts, which is much less practical.” Wu et al., writing a year later, record that hiding differentials “remains an open problem.”

Interestingly, the later history reversed that judgement. The partition trapdoors that we study next are differential trapdoors, and their simplest distinguisher needs one chosen plaintext pair (Corollary 3). Moreover, they do not need chosen plaintexts at all. Paterson’s own key-recovery scenario uses only a few known plaintexts [Pat99, §5], and the same partition gives known-plaintext distinguishers at the birthday bound on its cosets (Section 4.4). The practicality objection was therefore a poor guide to which family would last. The probability-1 guarantee of the partition construction also removes the noise on which the Rijmen–Preneel trapdoor depended for its concealment, and which in the end exposed it.

The constructions studied in this paper answer that open problem, and they do so by changing the kind of hidden object rather than its size. Instead of a statistical bias

concealed behind an unsearchable mask space, Paterson [Pat99] and later Bannier and Filiol [BF17] hide an *algebraic invariant*, a partition of the message space that the round function preserves with probability 1, for every key (Definition 4). In BEA-1 this holds for the secret S-boxes; the deployed cipher keeps the partition only probabilistically (Section 7.1). The resulting differential is truncated, not pointwise, and this is what allows it to hold with certainty instead of with a bias. Certainty removes the exploitability half of the Rijmen–Preneel tradeoff entirely. Whether it also removes the detectability half is the question for the rest of the paper. Section 4 states the partition mechanism precisely, and Section 5 gives an algorithm that recovers the hidden partition from the S-box tables. Sections 6 and 7 apply the algorithm to the two published instances. There the answer turns out to be no less negative than Wu et al.'s, although it takes considerably more work to reach. A later answer to the same open problem is Peyrin and Wang's MALICIOUS framework [PW20], which hides point differentials of probability 1 between related tweaks behind an extendable-output function. We take it up in Section 9.

4 Partition-Based Trapdoors

The Rijmen–Preneel trapdoor hides a biased linear approximation and fails because the bias is noise in a system of equations that the published tables hand to the evaluator. The constructions of Paterson and of Bannier and Filiol hide a different kind of object, a partition of the message space that every round function preserves for every key. In this section we state that mechanism precisely and relate it to the Rijmen–Preneel mechanism in the one setting where the two can be compared exactly, the Walsh spectrum of an S-box. We then read it in the correlation-matrix framework, where it becomes a perfect multidimensional linear approximation, and close by fixing the terminology for the rest of the paper. The terminology needs some care. The literature does not use a single vocabulary. The words "trapdoor", "backdoor" and "hidden structure" are all applied to constructions whose mechanisms have little in common, and this makes it easy to attribute to one author a mechanism introduced by another.

Remark 4 (A note on sources). Each account below is taken from the original work, read directly. These are Rijmen and Preneel's paper [RP97], Paterson's paper [Pat99], and, for Bannier et al., Bannier's thesis [Ban17] and the associated report [BBF16]. Where we use a later author's terminology for an older construction, we say so.

4.1 Paterson's Construction: Imprimitivity

Paterson's mechanism is group-theoretic and is a property of the *cipher*, not of an S-box. Let $(F_k)_{k\in\mathbb{F}_2^l}$ be the keyed round permutations of an n-bit iterated block cipher and let

$$G \;=\; \left\langle F_k \;\middle|\; k \in \mathbb{F}_2^l \right\rangle \;\leq\; \mathrm{Sym}(\mathbb{F}_2^n)$$

be the group they generate. Paterson studies the case in which G acts imprimitively on the message space, i.e., preserves a non-trivial partition $\mathcal{B} = \{B_0, \ldots, B_{2^d-1}\}$ with $0 < d < n$, each block of size 2^{n-d}. Following Bannier [Ban17, §3.1], we then call the cipher an *imprimitive backdoor cipher*. The group generated by the encryption functions is a subgroup of G, so the imprimitivity of G is inherited by the cipher as a whole, whatever the number of rounds.

The trapdoor information is the partition $\mathcal{B}$. Knowing it gives a chosen-plaintext attack. The attacker requests one ciphertext c_i for a plaintext p_i drawn from each block B_i, which reveals the induced permutation $\bar{g}$ on block indices. After that, any ciphertext c localizes its plaintext to the block $B_{\bar{g}^{-1}(j)}$, which contains 2^{n-d} candidates. The parameter d trades data against precision. A large d needs many chosen plaintexts and returns a small candidate set, and a small d does the reverse.

Two refinements matter for the taxonomy. First, suppose there are *two* invariant partitions $\mathcal{A}$ and $\mathcal{B}$ whose blocks meet in single points. Paterson's example uses about $2^{n/2}$ chosen plaintexts (2^{32} for a 64-bit block), chosen so that they meet every block of both systems at once, and recovers both induced permutations. This gives an alternative decryption algorithm for every ciphertext without recovering the key, what is usually called a *global deduction*. Second, Paterson describes a key-schedule-dependent attack. If the schedule is such that the induced permutation of $\mathcal{B}$ depends on only part of the key, that part can be found by exhaustive search from a few known plaintexts, and the rest by a second search. For his own example he estimates that a schedule which lets only 40 of the 80 key bits reach the outer S-box inputs would give the whole key in about 2^{41} trial encryptions. He leaves the schedule itself unspecified, however, and a concrete instance with a specified schedule was given later by Bannier et al. [BBF16].

Nothing in this mechanism involves a biased linear approximation, and nothing in it is probabilistic. The defining relation holds for every key and with certainty. This is the sharpest difference from the Rijmen–Preneel family.

4.2 Bannier et al.: From Imprimitivity to Partitions

Bannier et al. generalize in three distinct steps, only the third of which is specific to these authors.

The first step is not theirs. Relaxing the requirement that the input and output partitions coincide is credited by Bannier [Ban17, §3.1.4] to Harpes' thesis [Har96]. This gives Definition 15 below, of which an imprimitive cipher is the special case $\mathcal{A} = \mathcal{B}$.

Definition 15 (Partition-based backdoor cipher, after [Ban17, Def. 3.10])**.** An iterated n-bit block cipher E is a *partition-based backdoor cipher* if there exist partitions $\mathcal{A}$ and $\mathcal{B}$ of $\mathbb{F}_2^n$ such that for every cipher key K,

$$\{\, E_K(A) \mid A \in \mathcal{A} \,\} = \mathcal{B}.$$

The key-independence in Definition 15 is a design requirement. For a fixed key the image $E_K(\mathcal{A})$ is always *some* partition. What the designer needs is that it be the *same* partition for all keys, since otherwise the pair $(\mathcal{A}, E_K(\mathcal{A}))$ on which the attack depends would itself depend on the unknown key.

The second step is a relaxation to the probabilistic setting of Harpes and Massey's partitioning cryptanalysis [HM97]. It gives a cipher whose $(r-1)$-round function sends a significant excess of each part of $\mathcal{A}$ into one part of $\mathcal{B}$, but not all of it. Bannier formalizes this as a probabilistic partition-based backdoor cipher [Ban17, Definition 3.11].

The third step makes the family tractable. It is a chain of reductions from the cipher to a single S-box. It builds on two earlier results. Harpes showed that the only partitions which every key addition maps to one fixed partition are the coset partitions $\mathcal{L}(V)$ of linear subspaces [Har96, Lemma 4.3]. Caranti et al. [CDVSV06] then answered the question Paterson had raised, whether the trapped structure has to be linear. They showed that, for a key-alternating cipher, every block system of the group generated by the round functions consists of the cosets of a linear subspace. They also observed that such a block system gives a probability-1 truncated differential [CDVSV06, Corollary 4.1], following remarks of Paterson and Shamir. Bannier et al.'s own contribution is to extend these results from imprimitive ciphers to pairs of different partitions and to the whole cipher [Ban17, Theorem 3.26], so that the substitution layer must map one linear partition to another, and then to show that, when the diffusion layer is strongly proper in Calderini's sense, *at least one S-box* must do so [Ban17, Theorem 3.64 and §3.4]. This is what licenses the approach of Sections 5–7, which study one S-box at a time. Bannier et al. then derive lower bounds on the differential and linear uniformity of any S-box carrying such a partition, and give a construction algorithm meeting them. Corollary 1 reconstructs bounds of this kind in spectral form.

4.3 The Partition Condition and the Rijmen–Preneel Condition

Bannier et al.'s reduction places the partition condition in a single S-box, which is also where the Rijmen–Preneel trapdoor lives. Proposition 2 showed that balance rules out the latter. This is a negative result about one mechanism, and it should not be mistaken for a claim that the two mechanisms are unrelated. In this subsection we ask what the relationship is, and answer the question in the one setting where it has a precise answer, the Walsh spectrum of a single S-box. First, however, the three constructions have to be separated, because they are routinely conflated and their authors claimed different things.

Table 4: The three constructions, stated as their authors state them. Paterson's is a group-theoretic property of the cipher, not a property of an S-box. Caranti et al. [CDVSV06] showed that its blocks must be cosets of a linear subspace, answering a question Paterson raised, and Bannier et al. reduced the cipher-level condition to a single-S-box condition.

Construction	Hidden object, as its authors state it	Holds	Lives in
Rijmen–Preneel [RP97]	A mask β with $\beta \cdot S(x) = 0$ biased toward 0; a planted *linear approximation*	With probability $p_T < 1$	One point of the Walsh spectrum, $(0, \beta)$
Paterson [Pat99]	The group generated by the round functions acts *imprimitively*: it preserves a non-trivial block system	With probability 1	The group action on the message space
Bannier et al. [BBF16, Ban17]	The cipher maps a *linear partition* of the plaintext space to one of the ciphertext space; reduced to a condition on a single S-box	With probability 1	The DDT support of one S-box

The distinction in the last column matters. Rijmen and Preneel plant a *magnitude* anomaly at one prescribed location of the spectrum. The partition condition, as the next proposition shows, is instead a *support* condition, a statement about where the spectrum is allowed to be nonzero at all. Both are hidden structure in the same transform, and in this sense the families are related. They are, however, different predicates on that transform, and in this sense they are not. The transform is the Walsh transform $\widehat{S}$ of Definition 9, whose zero-input-mask entries $\widehat{S}(0, \gamma)$ are what Rijmen and Preneel plant.

Proposition 3 (Walsh-support characterization of the truncated differential)**.** *Let $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$, let $U \subseteq \mathbb{F}_2^m$ and $W \subseteq \mathbb{F}_2^n$ be subspaces, and fix $\gamma \in W^\perp \setminus \{0\}$. Then*

$$\gamma \cdot \big(S(x) \oplus S(x \oplus u)\big) = 0 \quad \text{for all } x \in \mathbb{F}_2^m,\ u \in U \qquad \Longleftrightarrow \qquad \operatorname{supp}_\gamma(S) \subseteq U^\perp.$$

Consequently S satisfies the probability-1 truncated differential $D_u S(x) \in W$ for all x and all $u \in U$ (the condition of Lemma 3) if and only if $\operatorname{supp}_\gamma(S) \subseteq U^\perp$ for every $\gamma \in W^\perp \setminus \{0\}$.

Proof. Write $f_\gamma := \gamma \cdot S$ and let $r_\gamma(u) := \sum_x (-1)^{f_\gamma(x) \oplus f_\gamma(x \oplus u)}$ be its autocorrelation. Two standard identities are used. First, grouping inputs by their output difference,

$$r_\gamma(u) \;=\; \sum_{b \in \mathbb{F}_2^n} (-1)^{\gamma \cdot b}\, \mathrm{DDT}_S(u, b),$$

so $r_\gamma(u) = 2^m$ exactly when $\gamma \cdot D_u S(x) = 0$ for every x, and $|r_\gamma(u)| \le 2^m$ always. Second, the Wiener–Khinchin identity, obtained by expanding the square of the Walsh transform and summing over λ,

$$r_\gamma(u) \;=\; 2^{-m} \sum_{\lambda \in \mathbb{F}_2^m} \widehat{S}(\lambda, \gamma)^2\, (-1)^{\lambda \cdot u}; \tag{2}$$

by Definition 11, $r_\gamma(u) = 2\,\mathrm{DLCT}_S(u,\gamma)$. Summing the second identity over $u \in U$ and using $\sum_{u\in U}(-1)^{\lambda\cdot u} = |U|$ if $\lambda \in U^\perp$ and 0 otherwise,

$$\sum_{u\in U} r_\gamma(u) \;=\; 2^{-m}\,|U| \sum_{\lambda\in U^\perp} \widehat{S}(\lambda,\gamma)^2. \tag{3}$$

Parseval's relation gives $\sum_{\lambda\in\mathbb{F}_2^m} \widehat{S}(\lambda,\gamma)^2 = 2^{2m}$ for every $\gamma \neq 0$.

($\Rightarrow$) If $\gamma \cdot D_u S \equiv 0$ for all $u \in U$ then every one of the $|U|$ terms on the left of Equation (3) equals 2^m, so the left side is $|U|\,2^m$ and hence $\sum_{\lambda\in U^\perp} \widehat{S}(\lambda,\gamma)^2 = 2^{2m}$. By Parseval the total over all λ is the same value, so $\widehat{S}(\lambda,\gamma) = 0$ for every $\lambda \notin U^\perp$.

($\Leftarrow$) If $\mathrm{supp}_\gamma(S) \subseteq U^\perp$ then $\sum_{\lambda\in U^\perp} \widehat{S}(\lambda,\gamma)^2 = 2^{2m}$ by Parseval, so Equation (3) makes the left side $|U|\,2^m$. As each of the $|U|$ summands is at most 2^m, each must equal 2^m, which is the stated condition.

For the final sentence, $D_u S(x) \in W$ for all x holds if and only if $\gamma \cdot D_u S(x) = 0$ for every $\gamma \in W^\perp$, since $W = (W^\perp)^\perp$. □

Remark 5 (What Proposition 3 does and does not say). The proposition characterizes the truncated-differential condition of Lemma 3, which is the "into" half of Definition 4. It does not capture the surjectivity and per-coset injectivity that Definition 4 additionally demands, and which Check (Algorithm 2) tests directly; for a non-bijective S-box, such as those of KG-DES, those are strictly stronger. They too have a spectral form, the second half of Proposition 4 below. Nor does it say that the Rijmen–Preneel and partition families are the same construction. It says that both are conditions on one transform, and that they are conditions of different types. Rijmen and Preneel require $\widehat{S}(0,\beta) \neq 0$, a single large value, whereas the partition condition requires $\widehat{S}(\lambda,\gamma) = 0$ off a subspace and is silent about $\lambda = 0$, because $0 \in U^\perp$ always. The two are logically independent. Balance kills the first and leaves the second untouched, as Proposition 2 and the screen of KG-DES in Section 6.5 observe.

Two lower bounds follow from the proposition by pigeonhole, and they measure the price of the partition trapdoor. Bannier et al. [BBF16] prove bounds of this kind for S-boxes that map one linear partition to another. Our derivations below reconstruct them from Proposition 3 instead of restating their form, and we claim no priority for them.

Corollary 1 (Linear and differential cost of the trapdoor)**.** *Let S satisfy Lemma 3's condition for subspaces U, W with $\dim U = d$ and $\dim W = e$. Then for every $\gamma \in W^\perp \setminus \{0\}$,*

$$\max_\lambda |\widehat{S}(\lambda,\gamma)| \;\geq\; 2^{(m+d)/2}, \qquad \textit{and} \qquad \max_b \mathrm{DDT}_S(u,b) \;\geq\; 2^{\,m-e} \;\; \textit{for every } u \in U\setminus\{0\}.$$

Proof. For the first, Proposition 3 confines the 2^{2m} of Parseval's mass to the $|U^\perp| = 2^{m-d}$ points of $U^\perp$, so some point carries at least $2^{2m}/2^{m-d} = 2^{m+d}$ of squared mass. For the second, the 2^m inputs distribute over the $|W| = 2^e$ admissible output differences alone. □

For the 6×4 S-boxes of KG-DES ($m = 6$, $d = e = 2$) these read $\max_\lambda |\widehat{S}| \geq 16$ and $\max_b \mathrm{DDT} \geq 16$. Section 6.5 compares them with the measured values.

4.4 Partition Trapdoors in the Correlation-Matrix Framework

Beyne's geometric approach to linear cryptanalysis [Bey21] describes linear cryptanalysis and its variants through correlation matrices (Definition 10), whose entry $C^F_{\gamma,\lambda}$ is the correlation of the linear approximation with input mask λ and output mask γ. In Beyne's terms an approximation is a pair of subspaces of functions, and it is *perfect* when C^F maps the first into the second. Invariant subspaces and nonlinear invariants are special cases [Bey18], and a partition trapdoor is a perfect approximation of the same kind. Read

in this way, the partition gives an exact Fourier characterization of the partition map, linear approximations of the whole cipher that no single characteristic reveals, and a known-plaintext distinguisher. The Rijmen–Preneel trapdoor sits at the opposite extreme (Section 3.2).

Proposition 4 (Fourier characterization of the partition map)**.** *Let $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$, and let $U \subseteq \mathbb{F}_2^m$ and $W \subseteq \mathbb{F}_2^n$ be subspaces with $\dim U = \dim W$. Then S maps every coset of U bijectively onto a coset of W if and only if*

$$\widehat{S}(\lambda, \gamma) = 0 \quad \text{whenever exactly one of } \lambda \in U^\perp \text{ and } \gamma \in W^\perp \text{ holds.}$$

If in addition S is surjective, S maps $\mathcal{L}(U)$ to $\mathcal{L}(W)$ in the sense of Definition 4.

Proof. The half with $\gamma \in W^\perp$ and $\lambda \notin U^\perp$ is Proposition 3. It holds if and only if $D_u S(x) \in W$ for all x and all $u \in U$, in other words if and only if every coset of U is mapped into a single coset of W.

For the other half, let $d = \dim U$ and, for $a \in \mathbb{F}_2^m$, let $g_a(y) = |\{x \in a + U : S(x) = y\}|$. Since $\mathbf{1}_{a+U}(x) = 2^{d-m} \sum_{\lambda \in U^\perp} (-1)^{\lambda \cdot (x \oplus a)}$,

$$\widehat{g}_a(\gamma) \;:=\; \sum_y (-1)^{\gamma \cdot y}\, g_a(y) \;=\; \sum_{x \in a+U} (-1)^{\gamma \cdot S(x)} \;=\; 2^{d-m} \sum_{\lambda \in U^\perp} (-1)^{\lambda \cdot a}\, \widehat{S}(\lambda, \gamma).$$

As a function of a, the right-hand side is a function on $\mathbb{F}_2^m / U$ whose Fourier coefficients are the $\widehat{S}(\lambda, \gamma)$, $\lambda \in U^\perp$. Hence $\widehat{g}_a(\gamma) = 0$ for every a if and only if $\widehat{S}(\lambda, \gamma) = 0$ for every $\lambda \in U^\perp$. A function on $\mathbb{F}_2^n$ is constant on the cosets of W exactly when its Fourier transform vanishes off $W^\perp$. The second half of the condition therefore says that every g_a is constant on the cosets of W.

If S maps $a + U$ bijectively onto $b + W$, then $g_a = \mathbf{1}_{b+W}$, and both halves hold. Conversely, if both hold, then g_a is supported on a single coset $b + W$ by the first half, constant on it by the second, and of total mass $|U| = |W|$, so $g_a = \mathbf{1}_{b+W}$ and S maps $a + U$ bijectively onto $b + W$. If S is surjective, every coset of W is hit, which is the remaining requirement of Definition 4. □

The second half is what Remark 5 noted that Proposition 3 does not capture, i.e., surjectivity onto each coset and injectivity on it. For a bijective S either half implies the other, because C^S is orthogonal. Rows indexed by $W^\perp$ that are supported on the columns $U^\perp$, as many rows as columns, leave every other row zero there. This is Beyne's correspondence between perfect and zero-correlation approximations [Bey21, Theorem 4.2]. For a non-bijective S-box the two halves are independent conditions.

Similarly, the reasoning lifts from one S-box to a whole cipher.

Theorem 1 (A partition trapdoor is a perfect multidimensional linear approximation)**.** *Let E be a permutation of $\mathbb{F}_2^N$ and let $\mathbf{T}, \mathbf{T}' \subseteq \mathbb{F}_2^N$ be subspaces of the same dimension d such that E maps every coset of $\mathbf{T}$ into a coset of $\mathbf{T}'$. Then $C^E_{v,u} = 0$ whenever exactly one of $u \in \mathbf{T}^\perp$ and $v \in \mathbf{T}'^\perp$ holds, and the block of C^E with $u \in \mathbf{T}^\perp$ and $v \in \mathbf{T}'^\perp$ is the correlation matrix of the bijection $\overline{E} : \mathbb{F}_2^N/\mathbf{T} \to \mathbb{F}_2^N/\mathbf{T}'$ that E induces on the 2^{N-d} cosets. In particular that block is orthogonal.*

Proof. Since E is a permutation and the cosets have equal size, each coset of $\mathbf{T}$ is mapped bijectively onto a coset of $\mathbf{T}'$, and Proposition 4 gives the zero pattern. For $u \in \mathbf{T}^\perp$ and $v \in \mathbf{T}'^\perp$, $u \cdot x$ depends only on the coset $x + \mathbf{T}$ and $v \cdot E(x)$ only on the coset $E(x) + \mathbf{T}'$. Grouping the sum defining $C^E_{v,u}$ by cosets therefore gives the correlation of u and v, read as masks on the quotient spaces, over $\overline{E}$. The correlation matrix of a bijection is orthogonal. □

In Beyne's classification this is a perfect approximation of dimension 2^{N-d}, since every principal correlation [Bey21, Definition 4.2] equals 1. In effect the cipher contains, for every key for which the hypothesis holds, an exact $(N-d)$-bit block cipher acting on the cosets, as Figure 3 draws. Beyne's piling-up principle [Bey21, Theorem 5.1] writes the approximation map of a composition as the product of the round maps plus an error term. The error vanishes when every approximation in the trail is perfect, so a round-by-round trail of such subspaces is exact for every key. This is the correlation-matrix form of the inductions that prove the invariant subspaces of KG-DES and BEA-1 (Theorems 2 and 3).

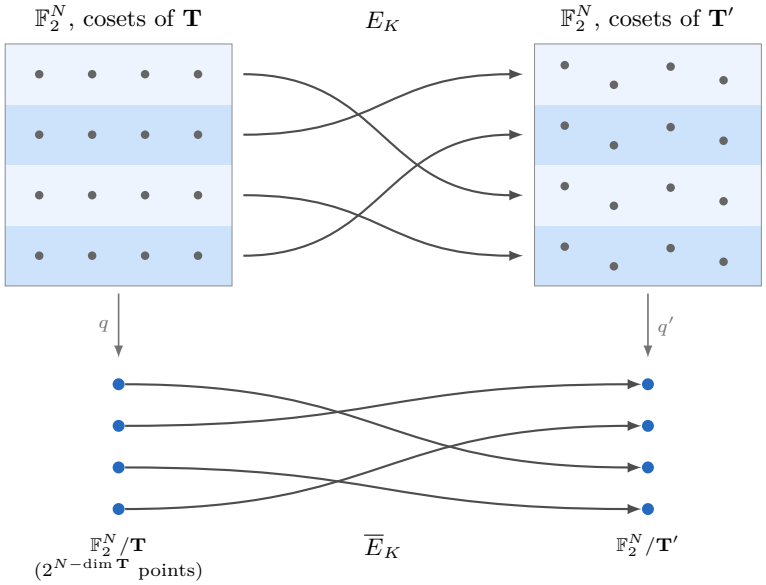

Figure 3: The geometry of Theorem 1, drawn with four cosets. Top: E_K, the cipher under a fixed key, sends each coset of $\mathbf{T}$ onto a whole coset of $\mathbf{T}'$, moving points within cosets in a key-dependent, nonlinear way. Bottom: the quotient maps q, q' collapse each coset to a point, and E_K induces a bijection $\overline{E}_K$ between the quotients. For BEA-1, $\mathbf{T} = \mathbf{T}_0$, $\mathbf{T}' = \mathbf{W}$ and $N = 80$; for KG-DES, $\mathbf{T} = \mathbf{T}' = \mathbf{S} \times \mathbf{S}$ and $N = 64$. A mask $u \in \mathbf{T}^{\perp}$ gives a character that is constant on every coset, i.e., a character of the quotient, so the block of C^{E_K} on $\mathbf{T}^{\perp} \times \mathbf{T}'^{\perp}$ is the correlation matrix of $\overline{E}_K$.

Corollary 2 (Linear approximations from the partition)**.** *Under the hypotheses of Theorem 1, for every nonzero $u \in \mathbf{T}^{\perp}$,*

$$\sum_{v \in \mathbf{T}'^{\perp} \setminus \{0\}} \left(C^{E}_{v,u}\right)^2 \;=\; 1,$$

so some output mask v of the block has $|C^{E}_{v,u}| \geq (2^{N-d}-1)^{-1/2} > 2^{-(N-d)/2}$.

Proof. A column of an orthogonal matrix has unit norm. The entry for $v = 0$ is 0 because $u \neq 0$ and E is a permutation, so the whole norm lies on the $2^{N-d}-1$ nonzero masks of the block, and the largest squared entry is at least their average. □

The corollary bounds a linear *approximation*, the sum over its whole hull, not a characteristic. Which output mask attains the bound depends on the key, and averaged over the nonzero output masks of the block the squared correlation is $(2^{N-d}-1)^{-1}$, against about 2^{-N} for a random permutation. This is the phenomenon Beyne discusses in terms of clustering [Bey21, Section 5.3], after Abdelraheem et al.'s observation that invariant subspaces produce linear approximations of unexpectedly high correlation [AÅBL12].

Beyne's framework includes the projection properties of Harpes and Massey's partitioning cryptanalysis [HM97]. Read as such a property, the partition also gives a known-plaintext distinguisher that needs no correlation estimate at all. Among M random known plaintexts, two in the same coset of $\mathbf{T}$ appear once $M^2/2$ reaches the number 2^{N-d} of cosets, i.e., from about $M = 2^{(N-d+1)/2}$ plaintexts. Their ciphertexts then lie in one coset of $\mathbf{T}'$, which for a random permutation happens with probability about 2^{d-N}. Finding the pairs means reducing each text modulo $\mathbf{T}$ and detecting a repeat, which a hash table does in time and memory linear in M.

4.5 Comparing the Mechanisms

Table 5: The three constructions compared. The rows that most sharply separate them are the mathematical object, the probability, and the data type, not the informal description, which is similar in all three cases.

Axis	Rijmen–Preneel	Paterson	Bannier et al.
Hidden structure lives in	One S-box, one spectral point	The group generated by the round functions	A partition of the message space, reduced to one S-box
Mathematical object	Biased linear form $\beta \cdot S$	Imprimitive group action	Linear partition $\mathcal{L}(V)$
Trapdoor information	The mask β	The invariant partition $\mathcal{B}$	The pair $(\mathcal{A}, \mathcal{B})$, equivalently (V, W)
In the correlation matrix (Section 4.4)	One entry $C_{\beta,0} = c_T < 1$: a one-dimensional approximation	An orthogonal block with zeros around it: a perfect approximation	The same, per S-box and for the cipher
Probability	$p_T < 1$	1	1 (or "$\gg$" in the probabilistic variant)
Key dependence	Key-independent	Key-independent	Key-independent by construction
Attack obtained	Linear cryptanalysis with a known good approximation; partial key	Chosen-plaintext block localization; global deduction; key recovery via the schedule	Truncated-differential / partitioning distinguisher, then key recovery
Data type	Known plaintexts	Chosen plaintexts (basic attack); known plaintexts (key-schedule attack)	Chosen plaintexts; known plaintexts suffice to distinguish
Cipher classes	Feistel with wide S-boxes (CAST, LOKI91)	Iterated ciphers generally	SPNs, and Feistel by the same theory
Concealment rests on	Infeasibility of scanning 2^n masks	Difficulty of finding the partition	Difficulty of finding (V, W)
Status	Broken [WBDY98]	Framework; example given	Framework plus BEA-1

Table 5 suggests two categories. Paterson and Bannier et al. belong to one family. The second is a strict generalization of the first, says so explicitly, and inherits its mechanism and its probability-1 character. Its basic attack uses chosen plaintexts, but Paterson's key-schedule attack needs only known ones [Pat99, §5], and known plaintexts also suffice for a distinguisher (Section 4.4). Rijmen and Preneel are not in that family. Their construction is statistical where the others are exact, and it lives at a single point of the Walsh spectrum where the others constrain its support (Proposition 3). The families share a setting (key-independent structure planted in an S-box, concealed by a search space) and that is enough to make the same analytic tools apply, but not enough to make them one mechanism.

4.6 A Fourth Construction: MALICIOUS

Peyrin and Wang's MALICIOUS framework [PW20] shares neither setting. Its trapdoor is not in an S-box but in the linear layers of a tweakable block cipher. These layers are fitted so that a few secret tweak pairs trigger related-tweak differential characteristics of probability 1. An extendable-output function in the tweak schedule conceals the pairs. The relation is exact, like Paterson's and Bannier's, but it holds between the permutations of two tweaks rather than inside one, so it is neither a biased approximation nor a preserved partition. Its concealment rests on a standard primitive rather than on a search space, and a backdoored instance is distributed as an honest one (Theorem 4). Section 9 places it beside the other three in Table 57.

4.7 Terminology: Backdoor, Trapdoor, Weakness

The looseness of the vocabulary causes real errors, and the literature already contains the resolution. Bannier states the criterion explicitly. A cipher whose round-function group is imprimitive is a backdoor cipher only if "the cipher's designer has to be aware of this property, otherwise we should not talk about backdoor" [Ban17, §3.1.2]. Intentionality is definitional, not incidental, which means it is generally not decidable from the artefact alone.

This suggests four categories, which we keep distinct throughout the paper.

Table 6: Four categories that the single word "backdoor" is often made to cover. Only the first requires intent, and intent is the one attribute no analysis of the primitive can establish.

Category	Criterion	Example from this paper
Backdoor	Structure is exploitable, asymmetric, and *intended*	KG-DES's $(\mathbf{U}, \mathbf{W}_i)$; BEA-1's secret S-boxes
Structural weakness	Exploitable by anyone who finds it; no asymmetry	The invariant subspace attacks on PRINTcipher [LAAZ11]; KG-DES's public differential and linear characteristics, a side effect of its backdoor (Sections 6.6.4, 6.8)
Asymmetric hidden structure	Confers advantage only on the holder, intent unknown	What an evaluator can actually establish about an unfamiliar cipher
Ordinary property	Unusual but confers no advantage	DES S_4's linear structures (Table 9)

The practical consequence is that detection procedures, including those of Section 5, decide membership in the third category, not the first. They establish that a structure is present and that it confers an advantage, but they cannot establish that anyone put it there deliberately. DES's S_4 is a useful reminder. It carries nine linear structures on a 2-dimensional direction set (Table 17), which is anomalous against a random baseline of 0 in 3,000. Nevertheless it is not a trapdoor, because the structure does not survive the expansion and confers no advantage. An anomaly is evidence of structure. Only exploitability makes it a weakness, and only intent makes it a backdoor.

We fix four further terms for the rest of the paper. A *backdoor* is a property of a cipher, in the first category of Table 6. The *trapdoor* is the hidden object that makes a backdoor usable, together with the information its holder needs. This is the mask β for Rijmen and Preneel, a pair of subspaces for a partition trapdoor, and the malicious tweak pairs for MALICIOUS. The trapdoor pair, trapdoor subspaces and trapdoor block of Definitions 4 and 13 take their names from it. A *partition trapdoor* is one whose hidden

object is a partition preserved by the cipher for every key (Definition 15). An *imprimitive* trapdoor is Paterson's special case, in which the input and output partitions coincide and the group generated by the round functions is imprimitive. By Bannier et al.'s reduction the partitions are coset partitions $\mathcal{L}(U)$ of linear subspaces, so in the ciphers studied here the two terms describe the same objects. In this paper, an *invariant subspace* of a cipher is a subspace $\mathbf{T}$ of differences that the cipher preserves with probability 1. Every coset of $\mathbf{T}$ is mapped into a coset of $\mathbf{T}$, or, for BEA-1, whose last round differs, into a coset of a second subspace. These are not the invariant subspaces of Leander et al. [LAAZ11], cosets mapped onto themselves under weak keys, which we call invariant cosets. Finally, a *structural weakness* is exploitable by anyone who finds it. KG-DES's differential and linear characteristics are weaknesses that its backdoor causes (Section 6).

4.8 Chronology

The development is not a single line of descent, and the literature is explicit about which influences are real. Harpes's work on partitions [HM97, Har96] and Rijmen and Preneel's trapdoor family [RP97] appear independently in the same period. Both exploit statistics, the first through partitions of the message space, the second through a biased linear approximation, and Harpes's thesis also considers ciphers that map one partition onto another exactly. Wu et al. [WBDY98] break the second the following year. Paterson [Pat99] introduces the imprimitivity framework, Caranti et al. [CDVSV06] show that its blocks must be cosets of a linear subspace and connect imprimitivity to truncated differentials, and Bannier et al. [BBF16, Ban17] unify the partition and imprimitivity lines, reduce them to a single-S-box condition and instantiate the result as BEA-1. They describe BEA-1 as inspired by Paterson's and Harpes' work. Alongside this line, and not as backdoor research at all, the invariant-subspace attacks [LAAZ11, LMR15] and subspace-trail cryptanalysis [GRR16] arrive at closely related objects from the cryptanalytic side. Bannier observes that the backdoors of his thesis are closely related to constant-dimensional subspace trails. Peyrin and Wang [PW20] then move the trapdoor out of the S-boxes altogether, into the linear layers of a tweakable cipher, and the attacks that followed on LowMC-M [BL20, LIM21, LSW+22] use the same tweak freedom from the outside.

Nearly all of this history is about how to build a backdoor. Very little of it tells an evaluator how to find one. However, the partition constructions rely on an exact relation, and an exact relation ought to leave an exact trace. This is the problem we take up next.

5 Discovering Partition Trapdoors from S-Box Tables

Bannier et al.'s reduction (Section 4.2) places the partition condition of a whole cipher in a single S-box, and Section 4.3 showed that the condition is a statement about where the Walsh spectrum vanishes. An evaluator who suspects a partition trapdoor therefore faces a concrete problem, and we solve it in this section before looking at any cipher.

Problem statement. Given an S-box $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ as a lookup table, find subspaces $U \neq \{0\}$ of $\mathbb{F}_2^m$ and $W \neq \mathbb{F}_2^n$ of $\mathbb{F}_2^n$, of equal dimension, such that π maps $\mathcal{L}(U)$ to $\mathcal{L}(W)$ (Definition 4), with U of maximal dimension; or establish that no such pair exists. The search space is the set of subspaces of $\mathbb{F}_2^m$; for $m = 10$ it holds about $2^{26.7}$ five-dimensional subspaces alone, far too many to test one by one, so the search must be pruned by a condition on single directions. A second problem is posed at the level of the cipher. Given trapdoor pairs of the S-boxes and the linear layer, decide whether their direct sum is preserved by the round function.

We first give the algorithm (Section 5.1) and prove it sound and, below an explicit cap on the candidate set, complete (Section 5.2). A spectral variant needs no cap, but it is

exact only under a condition that it can check (Section 5.3). We then ask how trapdoor pairs behave under the equivalences a designer might use to disguise an S-box (Section 5.4) and lift the S-box result to the round function (Section 5.5). The section closes with an assessment of what the method establishes and a comparison with prior subspace-trail algorithms. Sections 6 and 7 then apply it to KG-DES and BEA-1.

5.1 The Algorithm

Algorithm 1 corrects a candidate-search procedure that we published earlier [FASA23] and used in the thesis [Fah24, Chapter 3]. For every linear combination λ of the output bits (every component function $\lambda \cdot \pi$) and every direction α, the earlier version tested whether $\lambda \cdot \pi(x) \oplus \lambda \cdot \pi(x \oplus \alpha)$ stayed constant across all x. It then accepted α as a candidate only if this held for every λ that admitted *some* such direction at all. On the two ciphers studied later it recovers the correct trapdoors (Sections 6.2 and 7.2), so the reason for replacing it is not a wrong answer there. The reason is that it carries no guarantee, since we only checked, and never proved, that its output coincides with the true trapdoor. Algorithm 1 instead tests a condition that is provably necessary (Lemma 1). A direction α qualifies if and only if the $\mathbb{F}_2$-span of its difference set $\{\pi(x) \oplus \pi(x \oplus \alpha) : \text{all } x\}$ has dimension strictly less than the output width n. The new scan is also cheaper by a factor of 2^n (Remark 8 gives the cost of both).

Neither test, however, is sufficient on its own. A direction that passes a per-α criterion need not combine with the other survivors into a real trapdoor subspace, and a greedy build can take in a spurious element. On the KG-DES S-boxes it does (Remark 34). The algorithm therefore checks the definition itself (Check) before returning anything, and if the result of the greedy build fails that check, it falls back to an exhaustive search over subsets of the candidate set C. The two procedures it calls, Check and Finalize, are given separately in Algorithm 2, and the remarks that follow explain each step.

Algorithm 1 Searching for a trapdoor pair of an S-box (corrected)

procedure SEARCHTRAPDOOR(π, m, n)
    `Input: S-Box π, S-Box Size m × n`
    `Output: Subspace U, Subspace W`
    $C \leftarrow \emptyset$ ▷ candidate generation, difference-span dimension
    for $\alpha \leftarrow 1$ `to` $2^m - 1$ **do**
        $basis[0..n-1] \leftarrow 0$, $dim \leftarrow 0$ ▷ $\mathbb{F}_2$-linear basis, by highest set bit
        for $x \leftarrow 0$ `to` $2^m - 1$ **do**
            $v \leftarrow \pi(x) \oplus \pi(x \oplus \alpha)$
            if $v \neq 0$ **and** inserting v into $basis$ increases its rank **then** ▷ $O(n)$ per call
                $dim \leftarrow dim + 1$
                if $dim \geq n$ **then break** ▷ safe early exit, rank is monotone (Remark 8)
                end if
            end if
        end for
        if $dim < n$ **then** ▷ confined to a proper subspace = necessary condition
            $C \leftarrow C \cup \{\alpha\}$
        end if
    end for
    $U \leftarrow \{0\}$, $basis_U \leftarrow \emptyset$ ▷ WEAKBUILD: fast greedy proxy, not yet trusted
    for each c `in` sorted(C) **do**
        if $c \in U$ **then continue**
        end if
        $new \leftarrow \emptyset$, $all_in \leftarrow$ `true`
        for each $u \in U$ **do**
            $v \leftarrow u \oplus c$
            if $v \notin C \cup \{0\}$ **then** ▷ $O(1)$, indicator array over C
                $all_in \leftarrow$ `false` **break**
            end if
            if $v \notin U$ **then** $new \leftarrow new \cup \{v\}$
            end if ▷ likewise $O(1)$ over U
        end for
        if all_in **then**
            $U \leftarrow U \cup new \cup \{c\}$, $basis_U \leftarrow basis_U \cup \{c\}$
        end if
    end for
    $valid, W \leftarrow$ CHECK(U) ▷ trust nothing until verified
    if $valid$ **then**
        return FINALIZE(π, m, n, U, W)
    end if
    ▷ fallback: WEAKBUILD's basis may miss a true generator (Remark 6); search subsets of C itself, not of $basis_U$
    if $|C| >$ cap **then** ▷ exponential in $|C|$; cap e.g. 24
        report "search incomplete: $|C|$ exceeds cap" ▷ explicit failure
        return $\{0\}, \perp$
    end if
    $best \leftarrow \{0\}$, $best_W \leftarrow \perp$
    for each subset $Q \subseteq C$ **do** ▷ pruned in practice; provably complete (Prop. 5)
        $U' \leftarrow \text{span}(Q)$
        if $|U'| > |best|$ **then**
            $valid, W \leftarrow$ CHECK(U')
            if $valid$ **then**
                $best, best_W \leftarrow U', W$
            end if
        end if
    end for
    return FINALIZE($\pi, m, n, best, best_W$)
end procedure

Algorithm 2 Helper procedures for Algorithm 1: CHECK (the arbiter every candidate subspace must pass) and FINALIZE (the optional duality cross-check)

procedure CHECK(U) ▷ shared by the full build and every pruned candidate
  $W \leftarrow \bot$
  compute W from $shift = 0$ first; the remaining $2^m - 1$ shifts are then mutually independent, checked against that fixed W
  for $shift \leftarrow 0$ `to` $2^m - 1$ **do**
    $Img \leftarrow \{\, \pi(u \oplus shift) : u \in U \,\}$
    if $|Img| \neq |U|$ **then return** `false`, $\bot$ ▷ must be injective on this coset
    end if
    $rep \leftarrow$ any element of Img
    $W_{cand} \leftarrow \{\, v \oplus rep : v \in Img \,\}$ ▷ translate image to contain 0
    if $shift = 0$ **then**
      $W \leftarrow W_{cand}$
      if W is not closed under $\oplus$ **then return** `false`, $\bot$ ▷ W must be a subspace
      end if
    else if $W_{cand} \neq W$ **then** ▷ every coset must map onto a coset of the SAME W
      return `false`, $\bot$
    end if
  end for
  return `true`, W
end procedure

procedure FINALIZE(π, m, n, U, W) ▷ optional; not required for soundness
  if π **is bijective and** $|U| > 1$ **then**
    $\pi^{-1} \leftarrow$ the inverse permutation of π
    $U', W' \leftarrow$ SEARCHTRAPDOOR(π^{-1}, m, n) ▷ re-run the full algorithm on π^{-1}
    report whether $U' = W$ **and** $W' = U$ ▷ holds if the valid subspace is unique (Prop. 7)
  end if
  return U, W
end procedure

5.2 Necessity and Completeness of the Search

Lemma 1 (Necessary condition)**.** *If $u \neq 0$ belongs to a genuine valid subspace U (i.e. π maps $\mathcal{L}(U)$ to some $\mathcal{L}(W)$ with $W \neq \mathbb{F}_2^n$), then $u \in C$, the candidate set produced at the top of Algorithm 1.*

Proof. For every x, the inputs x and $x \oplus u$ lie in the same coset of U, and π maps that coset into a single coset of W. The values $\pi(x)$ and $\pi(x \oplus u)$ therefore lie in the same coset of W, and $\pi(x) \oplus \pi(x \oplus u) \in W$ for every x. Hence $\dim(\mathrm{span}\{\pi(x) \oplus \pi(x \oplus u) : \text{all } x\}) \leq \dim(W) < n$, which is the membership test for C. □

Remark 6 (Soundness versus completeness: a proven counterexample, and a fix)*.* Algorithm 1 is *sound*. It never returns anything that has not passed CHECK, so a reported $(\mathbf{U}, \mathbf{W})$ pair can always be trusted.

Throughout this subsection, $\widehat{\mathbf{U}}$ denotes a genuine planted subspace (one that really does satisfy Definition 4) as distinct from $\mathbf{U}^* = \mathbf{U} \setminus \{0\}$ of Definition 19; note $0 \in \widehat{\mathbf{U}}$.

An earlier version of the fallback step pruned subsets of WEAKBUILD's own basis $basis_U$ rather than of C itself, and an explicit counterexample shows that it is incomplete. We constructed the 64-element bijective S-box π of Table 7 for the purpose. It carries a real trapdoor $\widehat{\mathbf{U}} = \{0, 8, 16, 24\}$, and CHECK verifies that it maps onto $\widehat{\mathbf{W}} = \{0, 1, 2, 3\}$ across all 16 of its cosets.

However, the candidate set here is $C = \{1, 8, 16, 24\}$, and 1, though individually a candidate, is not part of $\widehat{\mathbf{U}}$ at all. Because $1 < 8$, WEAKBUILD accepts 1 first. When it then considers 8, the required check $8 \oplus 1 = 9 \in C$ fails, so 8 is rejected, and 16 and 24 fail in the same way against the contaminated span. The resulting $basis_U = \{1\}$ cannot generate $\widehat{\mathbf{U}}$ by any pruning, since 8, 16 and 24 were never in it. On this S-box that version of the algorithm returns $\mathbf{U} = \{0\}$ and misses a real trapdoor completely.

Algorithm 1 as stated above avoids this by searching subsets of C directly in the fallback, not subsets of $basis_U$. The search is still exponential, but it is provably complete, subject only to a feasibility cap on $|C|$:

Table 7: The constructed 64-element bijective S-box π of Remark 6, as a 4×16 lookup grid: row r (left header, 0–3) and column c (top header, 0–15) give $\pi(16r + c)$; e.g. row 2, column 0 is $\pi(32) = 38$.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	40	55	50	18	10	34	15	1	41	52	51	17	11	35	12	2
1	42	53	49	19	9	33	13	3	43	54	48	16	8	32	14	0
2	38	27	21	4	62	30	45	56	37	26	22	5	61	29	46	57
3	36	25	20	7	63	31	44	59	39	24	23	6	60	28	47	58

Proposition 5 (Completeness of the capped candidate-subset search)**.** *Let C be the candidate set produced by the first step of Algorithm 1, and suppose $|C| \leq$ cap. If $\widehat{\mathbf{U}}$ is any genuine valid subspace, then the fallback's search over subsets of C returns a subspace that passes* Check *and is of maximum size among all valid subspaces, hence at least as large as $\widehat{\mathbf{U}}$. It need not be $\widehat{\mathbf{U}}$ itself, since another valid subspace of the same or larger size may be returned instead.*

Proof. By Lemma 1, $\widehat{\mathbf{U}} \setminus \{0\} \subseteq C$. The fallback enumerates every subset of C, so the specific subset $Q = \widehat{\mathbf{U}} \setminus \{0\}$ is among those tried. Since $\widehat{\mathbf{U}}$ is closed under $\oplus$, $\mathrm{span}(Q) = \widehat{\mathbf{U}}$. Check is not a heuristic. It tests Definition 4 directly, by checking every coset, so Check($\widehat{\mathbf{U}}$) returns `valid` if and only if $\widehat{\mathbf{U}}$ satisfies the definition, which it does by hypothesis. The fallback's size guard (`if` $|U'| > |best|$) skips Check only when $|best| \geq |U'|$, and *best* is only ever assigned a subspace that has passed Check. Hence, when Q is reached, either $\widehat{\mathbf{U}}$ is checked and *best* becomes a valid subspace of size $|\widehat{\mathbf{U}}|$, or *best* is already a valid subspace of size at least $|\widehat{\mathbf{U}}|$, and $|best|$ never decreases afterwards. As $\widehat{\mathbf{U}}$ was an arbitrary valid subspace, the returned *best* is valid and of maximum size. Which maximum-size subspace it is depends on the enumeration order. □

The proposition is about the fallback, which runs only when WeakBuild's result fails Check. For the algorithm as a whole it gives the following guarantee when $|C| \leq$ cap. If a non-trivial valid subspace exists, a non-trivial valid subspace is returned. Since $C \neq \emptyset$ by Lemma 1, WeakBuild accepts at least the smallest candidate, and its result is either returned after passing Check or replaced by the fallback's. When WeakBuild's result passes, it need not be the largest valid subspace. Running the fallback unconditionally returns one of maximum size, at a cost exponential in $|C|$.

With this correction the algorithm recovers the planted trapdoor of the counterexample exactly. The run is short enough to give in full. The candidate scan accepts four directions. The difference set of 1 spans a 5-dimensional subspace, and those of 8, 16 and 24 span 2-dimensional ones. Hence $C = \{1, 8, 16, 24\}$. WeakBuild takes 1, rejects the other three as above and returns $\{0, 1\}$. Check rejects this subspace, because the coset $\{0, 1\}$ goes to $\{40, 55\}$, a translate of $\{0, 31\}$, while the coset $\{2, 3\}$ goes to $\{50, 18\}$, a translate of $\{0, 32\}$. The fallback then searches the subsets of C in order of size (Table 8). It calls Check on 12 of the 15 subsets, and only $\{8, 16\}$ passes. Its span $\{0, 8, 16, 24\}$ is sent onto the cosets of $\{0, 1, 2, 3\}$, and the duality cross-check of Proposition 7 succeeds on the returned pair.

On the 12 S-boxes where the earlier version was correct, the 8 of KG-DES (Table 13) and the 4 secret S-boxes of BEA-1 (Section 7.2), it returns the same answers as before.

Remark 7 (The remaining limitation). Proposition 5 is conditional on $|C| \leq$ cap, and this cannot be avoided, since searching every subset of C is exponential in $|C|$. When $|C|$ exceeds the cap, Algorithm 1 says so. It reports *search incomplete* instead of silently returning a trivial answer that may be wrong. The cap matters only when the fallback actually runs. When WeakBuild's own result passes Check, as for all four secret S-boxes

Table 8: The fallback of Algorithm 1 on the counterexample S-box of Table 7, with $C = \{1, 8, 16, 24\}$; q is the element of Q in the first row and its element other than 1 in the second. A subset whose span is no larger than the best valid subspace found so far is not checked.

Subsets Q of C	span(Q)	Check
$\{1\}, \{8\}, \{16\}, \{24\}$	$\{0, q\}$	fails
$\{1, 8\}, \{1, 16\}, \{1, 24\}$	$\{0, 1, q, q \oplus 1\}$	fails
$\{8, 16\}$	$\{0, 8, 16, 24\}$	passes, $\mathbf{W} = \{0, 1, 2, 3\}$
$\{8, 24\}, \{16, 24\}, \{8, 16, 24\}$	$\{0, 8, 16, 24\}$	not called (size guard)
$\{1, 8, 16\}, \{1, 8, 24\}, \{1, 16, 24\}, \{1, 8, 16, 24\}$	$\{0, 1, 8, 9, 16, 17, 24, 25\}$	fails

of BEA-1, where $|C| = 31$, the search stops at once and $|C|$ is irrelevant. In every case in this paper where the fallback *was* needed, $|C|$ stayed at 7 or below, well inside the cap of 24 that we use. These cases are the counterexample above, with $|C| = 4$, and the eight KG-DES S-boxes of Table 13, with $|C| = 7$ each.

Remark 8 (Early exit, and the cost of the scan). The rank *dim* in the candidate-generation loop never decreases as x is scanned, so once it reaches n, the direction α is disqualified whatever the remaining values of x. Leaving the inner loop at that point is therefore exact, not an approximation, and on the 12 S-boxes of this paper it changes no candidate decision. It does save work. On BEA-1's secret S_0 the scan reads on average 52 values of x per direction instead of all 1024. For the 992 directions that the exit disqualifies it reads only 22, while the 31 candidates are scanned in full. About 95% of the inner-loop work therefore disappears.

We use this scan as the reference point for every cost comparison in the paper. For an $m \times n$ S-box, the candidate scan loops over $2^m - 1$ directions α and 2^m inputs x, at $O(n)$ per basis insertion:

$$O\big(2^{2m} n\big), \qquad \text{i.e. } O(2^{2n} n) \text{ for a square S-box.}$$

The alternative was to reuse our earlier backdoor-detection construction [FASA23], [Fah24, Chapter 3], which builds a $2^n \times 2^m$ table $LS[\lambda, \alpha] = 1$ iff $\lambda \cdot D_\alpha \pi(x)$ is independent of x, i.e., iff α is a linear structure of the component function $\lambda \cdot \pi$. A nonzero row of LS at α therefore says that flipping the bits of α changes some linear combination of output bits in a way that does not depend on the input. Built as the original pseudocode describes, with a triple loop over λ, α and x and an $O(n)$ dot product per cell, the table costs $O(n \cdot 2^{n+2m})$, a factor of 2^n *more* than the scan above. A Walsh–Hadamard reformulation removes that factor. For each λ, two transforms of length 2^m give the autocorrelation of $\lambda \cdot \pi$ at every α at once, for a total of $O(m\, 2^{m+n})$. This is the scan's order for a square S-box, and below it by a factor of about $(n/m)\, 2^{m-n}$ when $n < m$. For KG-DES's 6×4 S-boxes, the scan takes $4 \cdot 2^{12}$ steps against $6 \cdot 2^{10}$ for the transforms. The transforms always cost their full amount, however, whereas the early exit keeps the scan far below its worst case. Recast as a per-α test (does *some* nonzero λ annihilate the difference set of α, i.e. is the autocorrelation of $\lambda \cdot \pi$ at α equal to $+2^m$?), either form leads to the same necessary condition. The choice between them is therefore one of cost and simplicity, not of correctness.

5.3 A Spectral Variant of the Search

Proposition 3 is constructive, so it can serve as a detection procedure as well as an explanation. Algorithm 1 scans input differences and builds a candidate set. The spectral route instead reads U and W directly off the geometry of the support:

1. For each $\gamma \neq 0$, compute $\widehat{S}(\cdot, \gamma)$ and the span T_γ of its support.

2. Collect $\Gamma := \{\gamma \neq 0 : \dim T_\gamma < m\}$, the masks whose spectrum is confined to a proper subspace.

3. Return $\mathbf{W} := \operatorname{span}(\Gamma)^\perp$ and $\mathbf{U} := \big(\operatorname{span} \bigcup_{\gamma\in\Gamma} T_\gamma\big)^\perp$.

We ran it on all eight KG-DES S-boxes with no prior knowledge of either subspace. It returns $\Gamma = \mathbf{W}_i^\perp \setminus \{0\}$, $\mathbf{U} = \{0, 4, 8, 12\}$ and $\mathbf{W} = \mathbf{W}_i$ exactly, in every case, matching Definition 19 and Table 13.

The more demanding test is BEA-1, whose secret S-boxes are 10×10 bijections carrying a dimension-5 trapdoor. They differ from KG-DES's in width and arity, and their trapdoor is two dimensions larger. The detector transfers without modification. For each of the four secret S-boxes it returns $|\Gamma| = 31$, $\dim \mathbf{U} = \dim \mathbf{W} = 5$, in about one second per S-box, reproducing Table 36. In addition, we checked separately that the recovered $\mathbf{U}$ and $\mathbf{W}$ are closed under $\oplus$ and satisfy $D_u S(x) \in \mathbf{W}$ on sampled (u, x). Against the four publicly deployed S-boxes it returns $\Gamma = \emptyset$, no trapdoor, agreeing with the trivial subspaces of Table 36.

The controls behave as they should. Seven of the eight genuine DES S-boxes produce $\Gamma = \emptyset$ and are not flagged at all. The eighth, S_4, whose nine linear structures the screen of Section 6.5 also finds, is flagged but yields the *trivial* $\mathbf{U} = \{0\}$. The detector thus correctly reports a spectral peculiarity that is not a partition trapdoor. No random DES-style S-box in 2,000 was flagged.

Table 9: The spectral detector of Proposition 3 run blind against both ciphers and two controls. Flagging DES S_4 while returning trivial $\mathbf{U}$ for it is the desired behaviour, not a false positive, since the box has a degenerate component function but no coset structure.

	KG-DES (6×4)	BEA-1 secret (10×10)	BEA-1 public (10×10)	DES (6×4)	Random (2,000)
Flagged ($\Gamma \neq \emptyset$)	8/8	4/4	0/4	1/8 (S_4)	0/2000
Non-trivial $\mathbf{U}$	8/8	4/4	0/4	0/8	0/2000
$\dim \mathbf{U}$ recovered	2	5	—	—	—
Agrees with published	Tab. 13	Tab. 36	Tab. 36	—	—

The last step of the detector takes for granted that the supports T_γ together span all of $\mathbf{U}^\perp$. The next proposition says when they do, and what the detector returns when they do not.

Proposition 6 (What the spectral detector guarantees)**.** *Let Γ, $\mathbf{W}$ and $\mathbf{U}$ be the output of the spectral detector for $S : \mathbb{F}_2^m \to \mathbb{F}_2^n$.*

1. *$D_u S(x) \in \mathbf{W}$ for every $u \in \mathbf{U}$ and every x. The detector never reports a truncated differential that does not hold.*

2. *If S has a trapdoor pair $(\mathbf{U}^*, \mathbf{W}^*)$ with $\mathbf{U}^* \neq \{0\}$ and $\mathbf{W}^* \neq \mathbb{F}_2^n$, then $\mathbf{W}^{*\perp} \setminus \{0\} \subseteq \Gamma$. In particular, $\Gamma = \emptyset$ certifies that S has no such pair.*

3. *If in addition $\Gamma \subseteq \mathbf{W}^{*\perp}$, then $\mathbf{W} = \mathbf{W}^*$ and $\mathbf{U} \supseteq \mathbf{U}^*$, and $\mathbf{U} = \mathbf{U}^*$ when S is injective.*

Proof. The first equivalence of Proposition 3 concerns a single mask γ and does not use W, so we apply it mask by mask.

(1) For $\gamma \in \Gamma$ the support of $\widehat{S}(\cdot, \gamma)$ lies in T_γ, and $T_\gamma \subseteq \mathbf{U}^\perp$ by the definition of $\mathbf{U}$. Hence $\gamma \cdot D_u S(x) = 0$ for all $u \in \mathbf{U}$ and all x. This is linear in γ, so it holds for every $\gamma \in \operatorname{span}(\Gamma)$, and $D_u S(x) \in \operatorname{span}(\Gamma)^\perp = \mathbf{W}$.

(2) For $\gamma \in \mathbf{W}^{*\perp} \setminus \{0\}$ the same equivalence gives $\mathrm{supp}_\gamma(S) \subseteq \mathbf{U}^{*\perp}$, a proper subspace because $\mathbf{U}^* \neq \{0\}$. Hence $\dim T_\gamma < m$ and $\gamma \in \Gamma$. As $\mathbf{W}^* \neq \mathbb{F}_2^n$, the set $\mathbf{W}^{*\perp} \setminus \{0\}$ is not empty.

(3) By (2) and the hypothesis, $\Gamma = \mathbf{W}^{*\perp} \setminus \{0\}$, so $\mathrm{span}(\Gamma) = \mathbf{W}^{*\perp}$ and $\mathbf{W} = \mathbf{W}^*$. Every T_γ lies in $\mathbf{U}^{*\perp}$, so the span of their union does too, and $\mathbf{U} \supseteq \mathbf{U}^*$. By (1), S maps every coset of $\mathbf{U}$ into a coset of $\mathbf{W}^*$. If S is injective, then $|\mathbf{U}| \leq |\mathbf{W}^*| = |\mathbf{U}^*|$, and so $\mathbf{U} = \mathbf{U}^*$. □

The hypothesis of part 3 can be checked from the output, since it implies that $\Gamma \cup \{0\}$ is a subspace. It holds on the twelve trapdoored S-boxes of KG-DES and BEA-1, where $\Gamma \cup \{0\} = \mathbf{W}_i^\perp$, with 3 masks for each KG-DES S-box and 31 for each secret S-box of BEA-1. For BEA-1's bijections, part 3 therefore proves that the detector returns the trapdoor, where Table 9 only records that it did. The KG-DES S-boxes are not injective, so part 3 gives only $\mathbf{U} \supseteq \mathbf{U}^*$ there, and equality is a computed fact.

When the hypothesis fails, the detector stays sound but can miss a trapdoor. The counterexample S-box of Table 7 is such a case. Its set Γ consists of the 15 nonzero masks of $\{0,1,2,3\}^\perp$ and one more mask, 17, whose component is invariant under $x \mapsto x \oplus 1$. That single extra mask makes $\Gamma \cup \{0\}$ fail to be a subspace, and the detector returns the trivial $\mathbf{U} = \{0\}$, whereas Algorithm 1 finds $\{0, 8, 16, 24\}$. DES's S_4, which has no trapdoor, is flagged for the same reason. Its Γ has five masks, $\Gamma \cup \{0\}$ is not a subspace, and the detector returns the trivial $\mathbf{U}$. In practice the detector should therefore pass to Algorithm 1 whenever $\Gamma \cup \{0\}$ is not a subspace or its answer fails Check.

Remark 9 (Standing of these results). The results above mix three levels of claim, and we separate them here. Established in the literature are Rijmen and Preneel's construction [RP97] and its break [WBDY98], Paterson's imprimitivity framework [Pat99], and Bannier et al.'s reduction of the cipher-level partition condition to a single-S-box condition, together with their uniformity bounds [BBF16]. Neither Paterson nor Bannier et al. proposes a Rijmen–Preneel-type hidden approximation, and nothing above should be read as attributing one to them. New here are Proposition 3, the detector built from it with its guarantees (Proposition 6), and the measurements of Table 9. Proposition 3 is elementary Fourier analysis and an instance of the correlation-matrix framework of Beyne [Bey21]. Section 4.4 develops this framework for partitions, and Sections 6.10 and 7.8 apply it to the two ciphers. We state it because it is the precise form in which the two families can be compared, not because the technique is new.

We prove, with a checkable condition, that the spectral detector is sound and that it returns the trapdoor itself when the masks it flags are those of one trapdoor (Proposition 6). The condition holds on all twelve trapdoored S-boxes of KG-DES and BEA-1, i.e., eight 6×4 non-bijective boxes with a dimension-2 trapdoor and four 10×10 bijections with a dimension-5 one. The detector reports no trapdoor on any of the twelve controls (DES S_4 is flagged but yields the trivial $\mathbf{U}$). The detector is not complete in general, since on the counterexample S-box it misses the trapdoor that Algorithm 1 finds. Its advantage is concrete, not asymptotic. It has no analogue of the candidate cap of Remark 7. BEA-1 is the case where that distinction bites, since there $|C| = 31$ exceeds the cap of 24 used in practice, and Algorithm 1 returns an answer only because WeakBuild's greedy result happens to pass Check before the fallback is reached. Had it failed, the algorithm would have reported *search incomplete*. The spectral route cannot fail in that way, because it never enumerates subsets, and on BEA-1 its condition holds, so the two methods complement each other. Both cost $O(2^{2m}n)$ for a square S-box, and we assert nothing about S-boxes wider than 10 bits, trapdoors of higher dimension, or partitions preserved only approximately.

5.4 Trapdoor Pairs Under Cipher-Design Equivalences

We now know how to *find* a trapdoor pair (Algorithm 1), and that the method is correct (Section 5.2). A designer, however, need not publish the trapdoored S-box as it is. It may be disguised by one of the standard transformations of S-box design, so we ask which of these transformations *preserve* the trapdoor structure, and in what form. We consider four such operations, ranging from the most rigid (inversion) to the most permissive (general CCZ-equivalence). Duality under inversion (Proposition 7) and plain affine equivalence (Proposition 8) keep the trapdoor intact. Extended affine equivalence (Proposition 9) preserves it only under an explicit extra condition that fails generically, and general CCZ-equivalence (Definition 16 onward) is not guaranteed to preserve it at all. An EA-equivalent of BEA-1's S_0 can even lose every trapdoor pair (Proposition 14). All four results hold for an arbitrary S-box, not just for those of KG-DES. Several of the worked checks below use BEA-1's secret S-boxes instead, because those results apply only to bijections, and KG-DES's 6-bit-to-4-bit S-boxes are not bijective.

Proposition 7 (Duality of $\mathbf{U}$ and $\mathbf{W}$ under inversion)**.** *Let $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a bijection and let U, W be linear subspaces of $\mathbb{F}_2^n$ such that π maps $\mathcal{L}(U)$ to $\mathcal{L}(W)$ (Definition 4). Then π^{-1} maps $\mathcal{L}(W)$ to $\mathcal{L}(U)$, and moreover $|U| = |W|$.*

Proof. By Definition 4, π mapping $\mathcal{L}(U)$ to $\mathcal{L}(W)$ means

$$\{\, \pi(C) : C \text{ a coset of } U \,\} = \{\, D : D \text{ a coset of } W \,\}.$$

Since distinct cosets of U are disjoint and π is injective, the map $C \mapsto \pi(C)$ sends distinct cosets to distinct (indeed disjoint) images. It is therefore injective on the set of U-cosets. Together with the displayed equality, this makes it a bijection from the set of U-cosets onto the set of W-cosets. In particular, counting both sides,

$$\frac{|\mathbb{F}_2^n|}{|U|} = \frac{|\mathbb{F}_2^n|}{|W|}, \qquad \text{so} \qquad |U| = |W|.$$

Now fix an arbitrary coset D of W. By the bijection just established, there is a unique coset C of U with $\pi(C) = D$ (as sets). Since π is injective and $|C| = |U| = |W| = |D|$, the restriction $\pi|_C : C \to D$ is an injective map between finite sets of equal size, and its image is all of D. Hence it is a bijection. Consequently, for every $d \in D$ there is a unique $c \in C$ with $\pi(c) = d$, so that $\pi^{-1}(d) = c$ and

$$\pi^{-1}(D) := \{\, \pi^{-1}(d) : d \in D \,\} = C,$$

a coset of U. As D ranges over every coset of W, the corresponding C (via the inverse of the bijection above) ranges over every coset of U. Hence

$$\{\, \pi^{-1}(D) : D \text{ a coset of } W \,\} = \{\, C : C \text{ a coset of } U \,\} = \mathcal{L}(U),$$

which is the statement that π^{-1} maps $\mathcal{L}(W)$ to $\mathcal{L}(U)$. □

Remark 10 (Duality cross-check for bijective S-boxes). By Proposition 7, if π maps $\mathcal{L}(\mathbf{U})$ to $\mathcal{L}(\mathbf{W})$, then, since π is a bijection, π^{-1} maps $\mathcal{L}(\mathbf{W})$ to $\mathcal{L}(\mathbf{U})$, and $\mathbf{U}$ and $\mathbf{W}$ trade roles. FINALIZE uses this as an independent check on correctness, not to save time. Since CHECK already gives $\mathbf{W}$, running SEARCHTRAPDOOR again on π^{-1} repeats the whole search along a different computational path, and confirms that it arrives at the same answer with the roles swapped. This applies only when π is bijective, because otherwise π^{-1} is not defined. The 6-bit-to-4-bit S-boxes of KG-DES do not qualify, so for them FINALIZE does nothing. We checked this on all four of BEA-1's secret S-boxes (Section 7.2, where π *is* bijective). In every case the $\mathbf{U}$ and $\mathbf{W}$ of π^{-1} are the $\mathbf{W}$ and $\mathbf{U}$ of π, with the roles swapped.

Proposition 8 (Trapdoor subspaces under affine equivalence)**.** *Let $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ map $\mathcal{L}(U)$ to $\mathcal{L}(W)$, and let $A : \mathbb{F}_2^m \to \mathbb{F}_2^m$, $B : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be invertible affine maps, $A(x) = A_0(x) \oplus a$ and $B(y) = B_0(y) \oplus b$ with A_0, B_0 invertible linear and a, b arbitrary constants. Define $\pi' := B \circ \pi \circ A$ (the affine equivalent of π under A, B). Then π' maps $\mathcal{L}(U')$ to $\mathcal{L}(W')$, where*

$$U' \;=\; A_0^{-1}(U), \qquad W' \;=\; B_0(W),$$

and $\dim(U') = \dim(U)$, $\dim(W') = \dim(W)$.

Proof. Only the linear parts A_0 and B_0 decide which subspace the cosets belong to. The constants a and b only relabel which coset of that subspace is reached, since for any subspace X and constant c,

$$A^{-1}(c + X) = A_0^{-1}(c \oplus a) + A_0^{-1}(X)$$

(a coset of $A_0^{-1}(X)$, since A_0^{-1} is linear and invertible), and similarly $B(c + X) = (B_0(c) \oplus b) + B_0(X)$, a coset of $B_0(X)$. By the first identity with $X = U$, A maps every coset of $U' := A_0^{-1}(U)$ onto a coset of U. As A is invertible, this correspondence is a bijection between the two sets of cosets. Fix a coset C' of U'. Then $A(C')$ is a coset of U, so by hypothesis $\pi(A(C'))$ is a single coset of W. By the second identity with $X = W$, $B(\pi(A(C')))$ is then a single coset of $W' := B_0(W)$. Hence $\pi'(C') = B(\pi(A(C')))$ is a single coset of W' for every coset C' of U'. As C' ranges over all cosets of U', the coset $\pi(A(C'))$ ranges over all cosets of W, by the bijection given by A and the hypothesis on π. Since B also gives a bijection, $\pi'(C')$ ranges over all cosets of W', and $\{\pi'(C') : C' \in \mathcal{L}(U')\} = \mathcal{L}(W')$. This is the statement that π' maps $\mathcal{L}(U')$ to $\mathcal{L}(W')$. Finally, A_0 and B_0 are linear isomorphisms, so they preserve dimension, and $\dim(U') = \dim(U)$ and $\dim(W') = \dim(B_0(W)) = \dim(W)$. □

Remark 11 (Scope: this is not how BEA-1 hides its trapdoor). Affine equivalence is a rigid notion. It preserves every cryptographic property of π, such as nonlinearity, differential uniformity and, by Proposition 8, trapdoor subspaces, and it merely relabels coordinates. BEA-1's public and secret S-boxes are related in a quite different way (Section 7.2). They differ at roughly 8–10% of the table entries, and the public version carries *no* detectable trapdoor at all, since Algorithm 1 finds only the trivial subspace in it (Table 36). A random perturbation on that scale is practically never an affine equivalence of the secret S-box. Proposition 8 therefore answers a different question, which design operations preserve a trapdoor, and it says nothing about how BEA-1's backdoor is concealed.

Proposition 9 (Trapdoor subspaces under extended affine (EA) equivalence)**.** *Let $\pi : \mathbb{F}_2^m \to \mathbb{F}_2^n$ map $\mathcal{L}(U)$ to $\mathcal{L}(W)$, let A, B be invertible affine maps as in Proposition 8, and let $\eta : \mathbb{F}_2^m \to \mathbb{F}_2^n$ be an* arbitrary *affine map, $\eta(x) = H_0(x) \oplus h$ with H_0 linear (not necessarily invertible, and not necessarily square) and h an arbitrary constant. Define the EA-equivalent*

$$\pi'(x) \;:=\; B(\pi(A(x))) \;\oplus\; \eta(x).$$

If $A_0^{-1}(U) \subseteq \ker(H_0)$ (the added term vanishes identically on the whole pulled-back subspace, not merely on part of it), then π' sends every coset of $U' = A_0^{-1}(U)$ onto a single coset of $W' = B_0(W)$, with U' and W' as in Proposition 8. If moreover π' is a bijection, π' maps $\mathcal{L}(U')$ to $\mathcal{L}(W')$.

Proof. Fix a coset C' of U' and $u' \in U'$. Since H_0 vanishes on all of $U' \subseteq \ker(H_0)$, we have $\eta(x) \oplus \eta(x \oplus u') = H_0(u') = 0$ for every x and every $u' \in U'$. The added term therefore contributes nothing to any difference taken *within* a U'-coset. Consequently, on any single U'-coset, π' agrees with $B \circ \pi \circ A$ up to the fixed additive constant $\eta(x_0)$, where x_0 is a chosen coset representative. For $x = x_0 \oplus u'$,

$$\pi'(x) = B(\pi(A(x))) \oplus \eta(x_0 \oplus u') = B(\pi(A(x))) \oplus \eta(x_0).$$

Hence $\pi'(C') = B(\pi(A(C'))) \oplus \eta(x_0)$. By the proof of Proposition 8, $B(\pi(A(C')))$ is a single coset of W', and a translate of a coset is again a coset, so $\pi'(C')$ is a single coset of W'. The translation constant $\eta(x_0)$ depends on the coset C', so distinct cosets of U' may in general land on the same coset of W'. If π' is a bijection, they cannot. The images of the pairwise disjoint cosets of U' are then pairwise disjoint, and since there are $2^m/|U'| = 2^n/|W'|$ of them ($m = n$ for a bijection), they exhaust $\mathcal{L}(W')$. Hence π' maps $\mathcal{L}(U')$ to $\mathcal{L}(W')$. □

Remark 12 (Plain affine equivalence is the case $H_0 = 0$). When $\eta \equiv 0$ (equivalently $H_0 = 0$), the hypothesis $A_0^{-1}(U) \subseteq \ker(H_0) = \mathbb{F}_2^m$ holds automatically, and Proposition 9 reduces to Proposition 8. Genuinely extended equivalences ($H_0 \neq 0$) satisfy the hypothesis only when $\ker(H_0)$ happens to contain the whole subspace $A_0^{-1}(U)$, which is five-dimensional for BEA-1. This is generically false for a random H_0, and impossible whenever H_0 is injective, since then $\ker(H_0) = \{0\}$ contains no nonzero subspace. We checked this on one of BEA-1's secret S-boxes with a random invertible 10×10 added term H_0. For $u' \in A_0^{-1}(\mathbf{V}_0)$, the set of output differences turned out to be a *different* 32-element set at each of several inputs x, and not a fixed subspace at all. This is expected, since $\ker(H_0) = \{0\}$ excludes all of $A_0^{-1}(\mathbf{V}_0)$ except 0. With $H_0 = 0$ the same check gave one fixed 32-element set for every x, as predicted.

Remark 13 (Partial intersection: confinement without a full partition map). If H_0 vanishes only on a *proper* subspace $U'' := A_0^{-1}(U) \cap \ker(H_0) \subsetneq A_0^{-1}(U)$, the difference computation in the proof above still shows $\pi'(x) \oplus \pi'(x \oplus u'') \in W'$ for every x and every $u'' \in U''$. Differences confined to U'' still land in W'. However, this is *not* the same as π' mapping $\mathcal{L}(U'')$ to $\mathcal{L}(W')$ in the full sense of Definition 4. Since π is injective on each U-coset (a consequence of CHECK's own verification), it is also injective on the smaller sub-coset $\{A(x) \oplus v : v \in A_0(U'')\}$. The image of this sub-coset therefore has size $|U''| < |U| = |W|$, so it covers only a *proper subset* of its W'-coset. For example, take U'' to be a 2-dimensional subspace of BEA-1's 5-dimensional $\mathbf{V}_0$. The image of a U''-coset under S_0 then has $|U''| = 4$ elements, all in one coset of $\mathbf{W}_0$, but it covers only 4 of that coset's 32 elements. The hypothesis $A_0^{-1}(U) \subseteq \ker(H_0)$ in Proposition 9 is therefore essential. Without it, the conclusion weakens from a partition-preserving map to mere confinement.

Definition 16 (CCZ-equivalence [CCZ98])**.** $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ and $\pi^* : \mathbb{F}_2^n \to \mathbb{F}_2^n$ are *CCZ-equivalent* if there exists an affine permutation $\mathcal{A}$ of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ such that

$$\{(x, \pi(x)) : x \in \mathbb{F}_2^n\} \;=\; \mathcal{A}\big(\{(x, \pi^*(x)) : x \in \mathbb{F}_2^n\}\big). \tag{4}$$

Equivalently, $\mathcal{A}$ carries the graph of π^* onto the graph of π, with no requirement that $\mathcal{A}$ respect the input/output coordinate split at all. EA-equivalence (Proposition 9) is the special case in which $\mathcal{A}$ preserves that split, so that it is "triangular" with respect to the two $\mathbb{F}_2^n$ blocks. A general CCZ-equivalence need not be.

Applying $\mathcal{A}$ to the graph of π and reading off the first coordinate gives a map of the form $x \mapsto T_{11}(x) \oplus T_{12}(\pi(x))$, and this defines a new function graph only if the map is a bijection. We first looked for such maps empirically. For BEA-1's secret S-box S_0 we tried 200 random choices of the linear maps T_{11} and T_{12}, and none of them gave a bijection. The next proposition replaces this search with an exact criterion, and Remark 14 explains the outcome.

Proposition 10 (Bijectivity criterion for block-triangular graph maps)**.** *Let $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ and let $T_{11}, T_{12} : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be linear (not necessarily invertible). Define $f(x) := T_{11}(x) \oplus T_{12}(\pi(x))$. Then f is a bijection of $\mathbb{F}_2^n$ if and only if*

$$T_{11}(a) \;\notin\; T_{12}\big(\mathrm{Image}(D_a\pi)\big) \qquad \text{for every } a \in \mathbb{F}_2^n \setminus \{0\},$$

where $D_a\pi$ is the derivative of Definition 7 and $\mathrm{Image}(D_a\pi) := \{D_a\pi(x) : x \in \mathbb{F}_2^n\}$.

Proof. Since f's domain and codomain have the same finite size, f is a bijection iff it is injective. For $x \neq x'$, write $a := x \oplus x' \neq 0$. By the linearity of T_{11} and T_{12},

$$f(x) \oplus f(x') = T_{11}(x) \oplus T_{11}(x') \oplus T_{12}(\pi(x)) \oplus T_{12}(\pi(x')) = T_{11}(a) \oplus T_{12}\big(\pi(x) \oplus \pi(x')\big).$$

Since $x = x' \oplus a$, $\pi(x) \oplus \pi(x') = \pi(x' \oplus a) \oplus \pi(x') = D_a\pi(x')$ by Definition 7. Hence $f(x) = f(x')$ (a collision) iff $T_{11}(a) = T_{12}(D_a\pi(x'))$ for this a and x'. Thus f has *some* collision iff there exist $a \neq 0$ and $x' \in \mathbb{F}_2^n$ with $T_{11}(a) = T_{12}(D_a\pi(x'))$, in other words iff there exists $a \neq 0$ with $T_{11}(a) \in T_{12}(\{D_a\pi(x') : x' \in \mathbb{F}_2^n\}) = T_{12}(\mathrm{Image}(D_a\pi))$. Negating this, f is injective iff $T_{11}(a) \notin T_{12}(\mathrm{Image}(D_a\pi))$ for every $a \neq 0$. □

Remark 14 (Why the random search was expected to fail). The criterion of Proposition 10 also explains the outcome of that search quantitatively. For $\pi = S_0$, the size $|\mathrm{Image}(D_a\pi)|$ over the 1023 nonzero a has mean 314.76 out of 1024 (about 30.7% of the codomain) and median 323. Treat $T_{12}(\mathrm{Image}(D_a\pi))$ as a random subset of that size, a standard heuristic for a "generic-looking" S-box with no further exploitable structure. For a *random* linear T_{11}, each of the 1023 required exclusions $T_{11}(a) \notin T_{12}(\mathrm{Image}(D_a\pi))$ then fails independently with probability around $314.76/1024 \approx 0.30738$. Requiring all 1023 simultaneously gives a heuristic success probability of order $(1 - 0.30738)^{1023} \approx 6.7 \times 10^{-164}$, so it is no surprise that none of the 200 trials succeeded. To find a CCZ-equivalent function that genuinely mixes input and output, one has to exploit specific algebraic structure of π, as the constructions in the literature do, typically through power maps over $\mathrm{GF}(2^n)$. A generic linear search will not do. BEA-1's secret S-boxes were built without any known field structure of this kind, so they offer no obvious starting point for such a construction.

Remark 15 (A byproduct: an independent detector for the trapdoor itself). The quantity $|\mathrm{Image}(D_a\pi)|$ of Proposition 10 also gives a second, independent way to locate a trapdoor subspace, unrelated to the span-dimension test of Algorithm 1 (Section 5.1). Directions a with an unusually *small* image are candidates for membership in $\mathbf{V}_i$. On S_0 the separation is very clear. Of the 1023 nonzero directions, 31 have $|\mathrm{Image}(D_a\pi)| = 31$, which is the whole of $\mathbf{W}_0 \setminus \{0\}$ (a bijection has no zero difference). The remaining 992 lie between 297 and 348, and nothing falls between 31 and 297. The 31 small-image directions are the elements of $\mathbf{V}_0 \setminus \{0\}$. Minimizing the image size thus confirms the same 5-dimensional trapdoor independently of the span-dimension criterion of Algorithm 1.

The next three propositions record known facts about CCZ-equivalence, from Carlet, Charpin and Zinoviev [CCZ98] and the later literature on vectorial Boolean functions. Unlike the other formal results of this paper, we do not derive them from scratch, because the spectrum-preservation theorem behind them is a substantial result in its own right. What we add is the connection to partition trapdoors, and Proposition 13 also rests on our own Proposition 9.

Proposition 11 (Bijectivity is not preserved)**.** *If π is a bijection, a π^* satisfying Definition 16 may or may not be a bijection.*

Proof. Both directions are exhibited by explicit construction.

May be a bijection: take $\mathcal{A}$ to be the identity map on the graph space. Then $\pi^* = \pi$, which is a bijection by hypothesis.

May not be a bijection: the EA-transform $\pi^*(x) = B(\pi(A(x))) \oplus \eta(x)$ of Proposition 9 is a special (triangular) case of Definition 16, so it suffices to exhibit one instance that fails to be injective. Take A and B to be the identity, fix a nonzero input e and a mask c with $c \cdot e = 1$, put $v := \pi(0) \oplus \pi(e)$, and let $\eta(x) := (c \cdot x)\, v$, a linear map of rank 1. Then $\pi^*(0) = \pi(0)$ and $\pi^*(e) = \pi(e) \oplus v = \pi(0)$, so π^* is not injective, whatever the bijection π. For $\pi = S_0$, one of BEA-1's secret S-boxes, and $e = c = 1$: $v = 186 \oplus 38 = 156$, $\pi^*(0) = \pi^*(1) = 186$, and π^* takes only 784 of the 1024 possible values. □

Proposition 12 (Additive linear structures: frequency preserved, direction not)**.** *Let $a \in \mathbb{F}_2^n \setminus \{0\}$ be an* additive linear structure *of π (Definition 7: $D_a\pi$ is constant). If π and π^* are CCZ-equivalent, then π^* has the same* number *of additive linear structures as π, but not generally at the same directions a.*

Justification. A direction a is a linear structure of π if and only if the row of the difference table at a has a single entry equal to 2^n, because every x gives the same $D_a\pi(x)$. This is the extremal case of the differential spectrum. CCZ-equivalence is known to preserve the differential spectrum as a *multiset* of values [CCZ98], so it preserves the count of entries equal to 2^n, i.e., the number of linear structures. It does not preserve *which* cells of the table those extremal entries occupy, since $\mathcal{A}$ need not respect the input/output split. The directions a that realize linear structures in π^* are therefore in general different from those of π. Unlike the other proofs in this subsection, this one relies on the cited theorem instead of deriving it. The remark that follows checks its consequence on an explicit instance, but it does not replace a proof of the theorem. □

Remark 16 (An explicit instance). None of BEA-1's secret S-boxes has a nonzero linear structure (their largest DDT entry is 46, not 1024), so testing Proposition 12 on them would only confirm the empty case. We therefore take $\pi(x) := g(x \gg 1) \oplus ((x \bmod 2) \cdot K)$ for $x \in \mathbb{F}_2^{10}$, where $g(y) := S_0(y)$ for $y \in [0, 512)$ (one of BEA-1's secret S-boxes, restricted to its first 512 inputs and read as a 9-bit function) and $K = \texttt{0x2a5}$. Since g does not depend on the least significant bit of x, $D_1\pi(x) = K$ identically. Hence $a = 1$ is a linear structure of π by construction, and an enumeration shows that it is the *only* one. A plain affine equivalence $\pi^*(x) = B(\pi(A(x)))$ (Proposition 8, itself a special case of CCZ-equivalence) satisfies $D_a\pi^*(x) = B_0\big(D_{A_0 a}\pi(A(x))\big)$, so π^* has one linear structure, at $a = A_0^{-1}(1)$ with value $B_0(K)$. The count is preserved, and for a random invertible A_0 the direction moves away from 1 with probability 1022/1023. Enumeration on random instances confirms both.

Proposition 13 (Non-trivial subspaces are not necessarily preserved)**.** *If π is bijective and maps $\mathcal{L}(U)$ to $\mathcal{L}(W)$ for some non-trivial U, W, a CCZ-equivalent π^* need not map $\mathcal{L}(U')$ to $\mathcal{L}(W')$ for the* specific *$U' = A_0^{-1}(U)$, $W' = B_0(W)$ induced by the equivalence's own affine data (Proposition 8's construction), i.e. the original trapdoor, transported through the same transformation that carries π to π^*, need not survive.*

Proof. We have shown that EA-equivalence, the triangular special case of Definition 16, does not in general preserve this induced trapdoor. Proposition 9 requires the hypothesis $A_0^{-1}(U) \subseteq \ker(H_0)$, and Remark 12 shows that it fails for a generic added term, in which case no non-trivial trapdoor of the induced form survives. Remark 13 shows that even a partial survival is weaker than genuine partition preservation. Since CCZ-equivalence strictly contains EA-equivalence (Definition 16), this failure is inherited. For the same choice of π, A, B, η, the resulting π^* is also CCZ-equivalent to π, through the corresponding triangular $\mathcal{A}$, and it fails to map $\mathcal{L}(U')$ to $\mathcal{L}(W')$ for the reason given above. □

A stronger statement would say that π^* need not have *any* non-trivial trapdoor pair, not only that the pair transported from π fails to survive. Algorithm 1 cannot decide this on the natural test case. Take an EA-equivalent of one of BEA-1's secret S-boxes with a random invertible H_0, as in Remark 12. Its candidate scan accepts about half of the 1023 nonzero directions (543 on the first instance we tried, 511 on the one below), far beyond any cap, and the result of WeakBuild fails Check. A simple count settles the question without any subset search.

Lemma 2 (A counting condition)**.** *If π has a trapdoor pair (U, W) of dimension d, then at least $2^d - 1$ nonzero directions a have* $\dim \operatorname{span} \operatorname{Image}(D_a\pi) \leq d$.

Proof. Every nonzero $u \in U$ has $\operatorname{Image}(D_u\pi) \subseteq W$, and $\dim W = d$. □

Proposition 14 (A trapdoor need not survive at all)**.** *There is a bijection with a* 5*-dimensional trapdoor pair that has an EA-equivalent, and hence CCZ-equivalent, map with no non-trivial trapdoor pair of any dimension. An example is BEA-1's secret S-box* S_0 *with* $\pi^*(x) = S_0(x) \oplus H_0(x)$ *for an invertible linear* H_0*.*

Proof. Fix one random invertible H_0, the one used in the accompanying script (see the appendix). Over the 1023 nonzero directions, span $\mathrm{Image}(D_a\pi^*)$ has dimension 5 for one direction, 6 for 30, 9 for 480 and 10 for the remaining 512. Only 31 directions therefore have dimension at most 8, fewer than the $2^d - 1$ that Lemma 2 requires for every d from 1 to 8. A pair of dimension 9 would have a hyperplane as its input subspace, all of whose 511 nonzero elements have dimension at most 9. Testing the 1023 hyperplanes shows that none qualifies, so no pair of any dimension from 1 to 9 exists. □

Similarly, no trapdoor pair exists for any of 20 further random invertible H_0 that we tried. The map π^* is not a bijection (it takes 637 of the 1024 values), an instance of Proposition 11, but Definition 4 does not require one, and the argument does not use bijectivity. The 31 directions of small span are the nonzero elements of $\mathbf{V}_0$. The trapdoor still confines their differences, to $\mathbf{W}_0 + \langle H_0(a) \rangle$, but for all of them except one this subspace has dimension 6, one more than a 5-dimensional pair over them would allow.

Remark 17 (Consistency with coarser CCZ-invariance failures)*.* Proposition 13 is consistent with the well-known failure of CCZ-equivalence to preserve even coarser invariants such as the algebraic degree, the property that makes it useful for finding inequivalent APN functions. A fine linear structure such as an imprimitivity trapdoor surviving a general CCZ-equivalence would be the exception, not the rule, and Proposition 14 gives an instance in which it does not survive at all.

5.5 A General Algorithm for Deriving Round-Function Subspaces

A trapdoor pair of each S-box is only half of a partition trapdoor. The linear layer must also carry the output subspaces of one round onto the input subspaces of the next. In this subsection we turn that requirement into a mechanical computation. Given the S-box pairs $\mathbf{U}_i \to \mathbf{W}_i$ and the linear maps before and after the S-box layer, it computes the largest candidate subspace of the round input. It then checks that the round function preserves it, with no ad hoc reasoning about individual bit positions. The notation follows a DES-style round, with an expansion E before eight S-boxes and a bit permutation P after them. Remark 20 covers general linear layers. Section 6.3 applies the computation to KG-DES and Section 7.3 to BEA-1.

5.5.1 Setup

The round function decomposes as

$$R \xrightarrow{E} (g_1, \ldots, g_8) \xrightarrow{S_1,\ldots,S_8} (w_1, \ldots, w_8) \xrightarrow{P} F,$$

where $g_i \in \mathbb{F}_2^6$ is the 6-bit block of $E(R)$ feeding S_i, and $w_i \in \mathbb{F}_2^4$ is its output. Write $\mathrm{pr}_i : \mathbb{F}_2^{48} \to \mathbb{F}_2^6$ for the coordinate projection onto block i, so that $g_i = (\mathrm{pr}_i \circ E)(R)$.

Proposition 15 (Round-Level Candidate Subspace)**.** *Let* $\mathbf{U}_1, \ldots, \mathbf{U}_8 \subseteq \mathbb{F}_2^6$ *be linear subspaces (for KG-DES all equal to* $\mathbf{U} = \{0, 4, 8, 12\}$*, Section 6.2). Define*

$$\mathbf{T}_0 \;=\; \bigcap_{i=1}^{8} (\mathrm{pr}_i \circ E)^{-1}(\mathbf{U}_i).$$

Then $\mathbf{T}_0$ *is a linear subspace of* $\mathbb{F}_2^{32}$*, and it is the* largest *subspace with the property that* $R \in \mathbf{T}_0$ *implies* $g_i \in \mathbf{U}_i$ *for every* i*.*

Proof. Each $\mathrm{pr}_i \circ E$ is linear (a composition of linear maps), so the preimage of a subspace under it is a subspace; an intersection of subspaces is a subspace. Maximality is immediate. Any R with $g_i \notin \mathbf{U}_i$ for some i is excluded from the right-hand side of the defining constraint, and is therefore excluded from $\mathbf{T}_0$, but no R satisfying all eight constraints is excluded. □

5.5.2 Simplification for Selection-Type Linear Layers

Proposition 15 is fully general, but computing the preimage of a subspace under an arbitrary linear map requires Gaussian elimination over $\mathbb{F}_2$ (Remark 20 below). In DES-style ciphers, however, E is not an arbitrary linear map. It is a *selection* map, in which every output bit is a copy of exactly one input bit, with no XOR combination, and then $\mathbf{T}_0$ has a simple closed form.

Algorithm 3 Deriving $\mathbf{T}_0$ for a selection-type expansion E

Require: Expansion table E (as a list of source-bit indices), subspaces $\mathbf{U}_1, \ldots, \mathbf{U}_8 \subseteq \mathbb{F}_2^6$, each spanned by coordinate vectors
Ensure: Set $Z \subseteq \{1, \ldots, 32\}$ of R-bit positions defining $\mathbf{T}_0 = \{\, \delta \in \mathbb{F}_2^{32} : \delta_j = 0 \ \forall j \in Z \,\}$

$Z \leftarrow \emptyset$
for $i = 1$ **to** 8 **do**
  $Z_i \leftarrow \{\, p \in \{1, \ldots, 6\} :$ coordinate p is 0 for every $u \in \mathbf{U}_i \,\}$ ▷ local zero-positions of $\mathbf{U}_i$
  for each $p \in Z_i$ **do**
    $g \leftarrow 6(i-1) + p$ ▷ global position in the 48-bit output of E
    $Z \leftarrow Z \cup \{\, E[g] \,\}$ ▷ the R-bit copied to that position
  end for
end for
return Z

Algorithm 3 runs in $O(1)$ time (a fixed $8 \times 4 = 32$ lookups) and requires no matrix arithmetic, because each constraint "coordinate p of g_i is zero" translates *directly* to "R-bit $E[6(i-1)+p]$ is zero" under a selection map. No linear combination of R-bits is ever involved. The closed form needs each $\mathbf{U}_i$ to be spanned by coordinate vectors, as KG-DES's $\mathbf{U} = \{0, 4, 8, 12\}$ is, so that $\mathbf{U}_i$ is described by zero coordinates alone. Otherwise the constraints on g_i are parities of several coordinates, and the general route of Remark 20 applies.

Proposition 16 (P-Closure Check)**.** *Having fixed* $\mathbf{T}_0$*, let* $\mathbf{W} = \mathbf{W}_1 \oplus \cdots \oplus \mathbf{W}_8 \subseteq \mathbb{F}_2^{32}$ *(the direct sum of the S-box output subspaces, embedded in their respective pre-P coordinates). If*

$$P(\mathbf{W}) \ \subseteq \ \mathbf{T}_0,$$

then $\mathbf{T}_0$ *is a genuine round-function invariant subspace. Since P is a fixed permutation of coordinates, this containment is checked by mapping each free basis coordinate of* $\mathbf{W}$ *through the P-table and confirming it does not land in Z.*

Proof. Everything here concerns differences. Let two inputs R, R' have difference $\Delta R = R \oplus R' \in \mathbf{T}_0$. Since E is linear and the round key cancels, the input difference of S_i is $\Delta g_i = (\mathrm{pr}_i \circ E)(\Delta R)$, which lies in $\mathbf{U}_i$ by construction of $\mathbf{T}_0$ (Proposition 15). By Lemma 3, applied at each of the eight S-boxes, the output difference Δw_i lies in $\mathbf{W}_i$, so $(\Delta w_1, \ldots, \Delta w_8) \in \mathbf{W}$, and the hypothesis $P(\mathbf{W}) \subseteq \mathbf{T}_0$ gives $\Delta F = P(\Delta w_1, \ldots, \Delta w_8) \in \mathbf{T}_0$. Hence $\Delta R \in \mathbf{T}_0 \implies \Delta F \in \mathbf{T}_0$, which is round-function invariance. □

Remark 18 (Only the sufficient direction is claimed). We state Proposition 16 only as a sufficient condition, and only this direction is used in the paper (Lemma 5, Theorem 2, and their BEA-1 analogues in Section 7.3). The converse (invariance $\implies P(\mathbf{W}) \subseteq \mathbf{T}_0$) also holds under two extra conditions. First, each $\mathbf{W}_i$ is the *exact* set $\{D_u S_i(x)\}$ of differences realized, not merely a containing bound (guaranteed by CHECK's own construction). Second, $\mathbf{T}_0$ contains, for each i, differences that activate S_i alone with every input difference in $\mathbf{U}_i$ (for KG-DES, differences confined to nibble i, by Lemma 4). Under both, every $w \in \mathbf{W}_i$ is the output difference of a pair whose input difference lies in $\mathbf{T}_0$ and activates S_i alone, so invariance puts $P(w)$, with w placed in block i, in $\mathbf{T}_0$. These elements span $\mathbf{W}$, and P and $\mathbf{T}_0$ are linear, so $P(\mathbf{W}) \subseteq \mathbf{T}_0$. We do not use the converse, and for the general linear layer of Remark 20, where $\mathbf{T}_0$ need not have this per-block structure, we do not claim it.

Remark 19 (If P-Closure Fails). In general, $\mathbf{T}_0$ from Proposition 15 need not satisfy $P(\mathbf{W}) \subseteq \mathbf{T}_0$. When it does not, the subspaces can be shrunk until it does. Replace each $\mathbf{W}_i$ by its part that P sends into $\mathbf{T}_0$, $\mathbf{W}'_i := \mathbf{W}_i \cap (\text{block } i \text{ of } P^{-1}(\mathbf{T}_0))$. Replace each $\mathbf{U}_i$ by the set of input differences whose derivative stays in it, $\mathbf{U}'_i := \{u \in \mathbf{U}_i : D_u S_i(x) \in \mathbf{W}'_i \text{ for all } x\}$, which is again a subspace because $D_{u\oplus v} S_i(x) = D_u S_i(x) \oplus D_v S_i(x \oplus u)$. Then recompute $\mathbf{T}_1$ from the $\mathbf{U}'_i$ by Proposition 15. Each step can only shrink the subspaces, so the iteration reaches a fixed point within 32 steps, where Proposition 16 applies. The fixed point may be $\{0\}$, and whether a non-trivial invariant survives depends on the S-boxes. For KG-DES (Section 6.3) no iteration is needed, because the check of Proposition 16 already succeeds at $\mathbf{T}_0$.

Remark 20 (Generalization to Arbitrary Linear Layers). Algorithm 3 exploits the fact that DES's E is a selection map. If a cipher's linear layer θ instead computes $\mathbb{F}_2$-linear combinations of input bits (as in, e.g., an MDS matrix or a general binary matrix diffusion layer), Proposition 15 still applies verbatim, but the preimage $(\mathrm{pr}_i \circ \theta)^{-1}(\mathbf{U}_i)$ must be computed by standard linear algebra. Represent $\mathrm{pr}_i \circ \theta$ as a binary matrix M_i, represent $\mathbf{U}_i$ as the common kernel of a set of linear functionals $\{\ell_1, \ldots, \ell_t\}$ (a basis for $\mathbf{U}_i^{\perp}$), and solve

$$(\mathrm{pr}_i \circ \theta)^{-1}(\mathbf{U}_i) \;=\; \ker\big([\ell_1 M_i;\; \ldots;\; \ell_t M_i]\big)$$

by Gaussian elimination over $\mathbb{F}_2$. The intersection $\bigcap_i (\mathrm{pr}_i \circ \theta)^{-1}(\mathbf{U}_i)$ is then the nullspace of the stacked constraint matrix from all 8 boxes, again computable by a single Gaussian elimination. The P-closure check of Proposition 16 and the fixed-point iteration of Remark 19 are unchanged. This makes the method applicable, without modification beyond this linear-algebraic bookkeeping, to any SPN or Feistel cipher with a linear diffusion layer, not only to selection-based expansions such as DES's.

5.6 What the Method Establishes, and What It Does Not

We separate what the method guarantees from what we have only observed.

Proved. Algorithm 1 is *sound*. Nothing is returned without passing CHECK, and CHECK tests Definition 4 coset by coset. A reported pair is therefore a trapdoor pair, and the method has no false positives (Remark 6). It is *complete below the cap*. If the candidate set has at most cap elements and a non-trivial trapdoor pair exists, a pair of maximal dimension is returned (Proposition 5). The target is robust to affine re-encoding. An affine-equivalent S-box carries the transported pair (Proposition 8), and for a bijective S-box the inverse carries the dual pair (Proposition 7), which FINALIZE uses as an independent cross-check. The spectral variant rests on Proposition 3 and needs no cap. At the round level, Proposition 16 gives a sufficient condition for the direct sum of the S-box pairs to be invariant.

Cost. The candidate scan costs $O(2^{2m}n)$ for an $m \times n$ S-box (Remark 8), governed by the input width. An S-box as wide in its output as the Rijmen–Preneel 8×32 tables is therefore cheap to scan. CHECK costs $O(2^m)$ per candidate subspace, the fallback is exponential in $|C|$ and capped, and the round-level closure is a fixed number of table lookups for a selection layer or one Gaussian elimination for a general one. The 10-bit S-boxes of BEA-1 take about a second each.

Outside the method's reach. The method finds exact coset partitions that are visible in single S-boxes, and nothing else. It does not detect probabilistic partitions, the relaxation of Definition 15 that Bannier formalizes from Harpes and Massey's partitioning cryptanalysis [Ban17, Definition 3.11], where no derivative is confined exactly and no spectral support vanishes. No exact test can detect these. The round-level closure verifies one hypothesized configuration, every S-box active with its pair, and does not search over which subset of S-boxes is active, so a trapdoor placed in a partial activation pattern would be missed (Remark 23). Structure outside the S-boxes is invisible to any S-box test. A trapdoor planted in the linear layers or the tweak schedule, as MALICIOUS does, passes through untouched (Section 9). Disguises beyond affine equivalence need not preserve a trapdoor pair (Propositions 9 and 13), so a pair could in principle be transported into a form the search does not recognize; we know of no construction that does so. When the candidate set exceeds the cap the fallback is infeasible, and the algorithm reports an incomplete search rather than a wrong answer (Remark 7). BEA-1, with $|C| = 31$ against a cap of 24, is resolved only because the greedy build passes CHECK. Finally, no analysis of the primitive can establish intent. The method decides the third category of Table 6, not the first.

Observed, not proved. On the twelve trapdoored S-boxes examined in this paper, the eight of KG-DES and the four secret ones of BEA-1, both the algorithm and its spectral variant return the planted pair. For BEA-1's bijections Proposition 6 turns the spectral result into a proof, since its condition holds there. The algorithm finds only the trivial subspace in BEA-1's four public S-boxes, and the spectral variant reports no trapdoor in the eight DES S-boxes or in 2,000 random DES-style S-boxes. These are measurements on specific S-boxes, and we claim nothing for S-boxes wider than 10 bits, trapdoors of higher dimension, or partitions preserved only approximately.

5.7 Comparison with Prior Approaches

Our algorithms address a problem that is closely related to, but different in its goal from, the *subspace trail* algorithms of Leander, Tezcan and Wiemer [LTW18]. Both study chains of growing linear subspaces propagating through an SPN or Feistel round function, but they differ in what is assumed known at the start. To make the comparison concrete, we also bring in our *original* candidate-search procedure, the one that Algorithm 1 corrects and replaces. Its pseudocode is given in [FASA23] and [Fah24, Chapter 3], and we do not repeat it here. Remark 22 below sets out its structure, its cost and the gap that led us to replace it.

Tables 10 and 11 summarize the resulting five-way comparison of two algorithms of Leander et al. [LTW18], our corrected and original candidate-search procedures, and our round-level closure algorithm. The remarks that follow give the argument behind each entry.

Remark 21 (Algorithm 1 of Leander et al. [LTW18]: trace, not discover). Leander et al.'s [LTW18] Algorithm 1 (COMPUTETRAIL) takes as *input* a nonlinear bijective function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ *and* a starting subspace U that is already known. It then traces the forward trail $U = U_0 \subsetneq U_1 \subsetneq \dots$ until the dimension reaches n. For each basis vector of the current

Table 10: Complexity, strong points, and weak points: Leander, Tezcan, and Wiemer's [LTW18] Algorithms 1 (ComputeTrail) and 2 (OneRoundSubspaceTrails) against this paper's Algorithm 1 (SearchTrapdoor), its own superseded predecessor [Fah24] (Chapter 3), and the round-level closure algorithm (Algorithm 3, `derive-T0`). Following Algorithm 1's own convention, m = S-box input width and n = S-box output width (equal for a square S-box); k = number of parallel S-boxes; $|C|$ = candidate-set size.

Algorithm	Complexity	Strong point	Weak point
Leander Alg. 1	$O(n^2)$ evals of F (U grows to dim. n via $t \approx n{+}100$ sampled points per basis vector)	General: any nonlinear bijective F, no S-box structure assumed; cheap once U is known	Needs U as input, cannot discover it; correct only with overwhelming probability ($\approx 1 - 2^{-100}$), not certainty
Leander Alg. 2	$O(2^k n)$, exponential in S-box count k (Corollary 4: exactly 2^k candidates suffice)	Deterministic and provably complete under its precondition; a genuine nonexistence certificate	Precondition (no linear structures) fails exactly for deliberately backdoored S-boxes; degrades to an off-by-one-round bound once it does
Alg. 1 (this work)	$O(2^{2m} n)$ candidate generation; fallback exponential in $\|C\|$, explicitly capped	Discovers $(\mathbf{U}, \mathbf{W})$ from a raw table alone, exhaustively; reports "incomplete" rather than guessing when the cap binds	Costlier in n than tracing a given U, but solves the strictly harder discovery problem; capped fallback exponential in $\|C\|$
Original LS/HW (superseded)	$O(n \cdot 2^{n+2m})$ for its LS table as pseudocoded; a 2^n factor worse than Algorithm 1's candidate scan	Simple, intuitive HW-threshold rule; also correctly recovers BEA-1's trapdoor (Remark 22), not only KG-DES's	No Check analogue: W taken from the first ad hoc coset satisfying two heuristic conditions, never cross-checked against the others, so a returned pair is not self-certified even when (as on both ciphers tested here) it happens to be correct
`derive-T0` (this work)	$O(1)$ for a selection layer; one Gaussian elimination for a general linear layer; no 2^k term	Cheapest of the five once the per-S-box $\mathbf{U}_i$'s are known; no blow-up in S-box count	Assumes the per-S-box $\mathbf{U}_i$'s already discovered; verifies only one hypothesized all-active configuration

subspace it samples $t \approx n + 100$ random points x, evaluates the derivative $D_{u_i}F(x)$ at each, and takes the span. The result is correct with overwhelming probability but not with certainty, and the authors give an error probability of order 2^{-100} for this choice of t. Since $\dim(U)$ grows to at most n, the total cost is $O(n^2)$ evaluations of F. This is efficient, but only because the harder question, which U to start from, is assumed to be answered already.

Algorithm 1 answers that prior question. Given only the lookup table of an S-box, it decides whether a partition-preserving pair $(\mathbf{U}, \mathbf{W})$ exists at all, with no subspace supplied in advance. It does so exhaustively, not by sampling. The candidate step tests every nonzero direction α against every input x, and Check (Algorithm 2) tests every coset, not a sample. The price is a candidate step of cost $O(2^{2m} \cdot n)$ (Remark 8), exponential in the S-box width, against the $O(n^2)$ of tracing a *given* subspace. When the fallback search of Proposition 5 is infeasible, Algorithm 1 reports *search incomplete* (Remark 7), whereas Algorithm 1 of Leander et al. returns an answer with a small error probability by design.

Remark 22 (Our original algorithm). Our original LS/HW procedure [FASA23, Fah24], which Algorithm 1 replaces, belongs next to Leander et al.'s two algorithms, because, unlike either of them, it shares the goal of *discovering* $\mathbf{U}$ from a raw table. Its candidate step is the $O(n \cdot 2^{n+2m})$ construction of the LS table costed in Remark 8, a factor of 2^n

Table 11: Capabilities and limits of the same five algorithms.

Algorithm	What it *can* do	What it *cannot* do
Leander Alg. 1	Trace an already-known starting subspace U forward to the full trail, for *any* nonlinear bijective S-box	Discover U from a raw S-box table; guarantee zero error on any single run
Leander Alg. 2	Exhaustively find the best/longest one-round trail, and thereby prove no better one exists, for S-boxes with no linear structures	Handle S-boxes *with* linear structures exactly (only a near-tight bound); discover an unknown structure from a table
Alg. 1 (this work)	Find a planted trapdoor in a deliberately structured S-box, with a completeness proof (Proposition 5) subject only to an explicit feasibility cap	Scale to very large m without the candidate-generation cost growing exponentially, unavoidable for a search this exhaustive
Original LS/HW (superseded)	Find a genuine trapdoor via an intuitive Hamming-weight threshold on the LS matrix; checked directly, this succeeds on the trapdoors of both KG-DES and BEA-1	Certify its own output (no Check step, so a returned pair is trusted rather than verified); its necessity is not proven in general, only checked on the two ciphers examined here
`derive-T0` (this work)	Propagate an already-known S-box-level trapdoor to the full round/cipher level cheaply, for selection *or* general linear layers	Search over which subset of S-boxes to activate – would miss a partial-activation trail that Leander Alg. 2's exhaustive 2^k sweep would find

above the scan of Algorithm 1. More importantly, its final step ("Searching for Subspace W") has no counterpart of Check. It scans the cosets $i = 0, 1, 2, \ldots$ only until it finds the *first* one whose image contains 0 and XORs to 0, takes W from that coset, and stops. It never tests whether the *other* cosets of the candidate U also map onto cosets of the same W. Every other algorithm in this comparison carries some explicit guarantee on its output. Leander et al.'s Algorithm 2 is exact by the completeness argument of their Corollary 4, Algorithm 1 is exact through the coset-by-coset test of Check, and even their Algorithm 1 is correct up to a stated error probability. The original procedure carries no such guarantee, so a pair it returns is certified only by whatever validation is applied afterwards. Its candidate step does recover the genuine trapdoors of both KG-DES and BEA-1 (Section 5.2 and Remark 34). The missing certification is nevertheless a real gap, and the mandatory Check of Algorithm 1 was introduced to close it.

Remark 23 (Algorithm 2 of Leander et al. [LTW18]: the opposite precondition). Leander et al.'s [LTW18] Algorithm 2 (OneRoundSubspaceTrails) is deterministic and provably complete, but only under a precondition that the authors show to be essential. The S-box layer must have *no non-trivial linear structures*. Under that precondition their Corollary 4 shows that only 2^k starting subspaces need to be checked, where k is the number of parallel S-boxes and each subspace is, per S-box, either $\{0\}$ or the full block. Algorithm 2 therefore enumerates all 2^k activation patterns, each at the cost of one matrix–vector product with the known linear-layer matrix M, for a total of $O(2^k \cdot n)$. This is exponential in the *number* of S-boxes rather than in their width. The backdoored S-boxes of KG-DES and BEA-1 fall outside this regime. Their whole purpose is a planted, non-trivial partition structure (Definition 19, Table 36), which is the kind of structure that Algorithm 2 assumes absent. The authors acknowledge the gap in their Section 5, on S-boxes *with* linear structures. There the clean 2^k bound is lost, the choice of starting subspace "remains huge," and their fallback guarantees only a bound that is "tight... or off by at most one round."

Proposition 15, Algorithm 3 and Remark 20 take the complementary position. Once Algorithm 1 has found the subspaces $\mathbf{U}_i$, one known subspace per S-box instead of 2^k unknowns, the round-level closure $\mathbf{T}_0 = \bigcap_i (\mathrm{pr}_i \circ \theta)^{-1}(\mathbf{U}_i)$, with θ the linear layer before the S-boxes, is computed directly. This takes $O(1)$ for a selection layer such as DES's E

and P (Algorithm 3) and one Gaussian elimination for a general layer such as BEA-1's MixColumns (Remark 20, Proposition 29), with no 2^k enumeration. The trade-off is real. The construction verifies the closure of *one* configuration, fixed in advance, in which every S-box is active. It would not discover a stronger trail that uses only a *subset* of the S-boxes, as an exhaustive 2^k sweep would, and it is efficient precisely because it does not attempt that search.

Remark 24 (Two different research goals). These trade-offs come from a difference in what each line of work tries to establish. Leander et al. [LTW18] build a tool for security proofs. Given a published cipher that is presumed honest, they want to rule out long subspace-trail distinguishers efficiently and provably, or in their words to "provably exclude the existence of such attacks." Their Table 1 accordingly reports results for standard designs (AES, Anubis, Klein, Kuznyechik, Prince, Qarma). We build the converse tool. Given a possibly malicious S-box, we want to discover whatever partition structure has been planted in it, directly from its lookup table and without assuming anything about what that structure looks like. The two toolsets are therefore complementary. An evaluator could first run Algorithm 1 to find a candidate pair $(\mathbf{U}_i, \mathbf{W}_i)$ at the S-box level, a question their algorithms do not address, and then propagate it with the round-level machinery of either approach. At that stage Algorithm 3 and their Algorithm 1 cover overlapping ground. The closed form is preferable when the linear layer is a selection map, and random sampling is adequate when its small error probability is acceptable.

Remark 25 (Earlier detection ideas). Bannier already points to the detection side of his theory [Ban17, §4.4]. He suggests that the subspaces of a trapdoored S-box can be recovered from the largest entries of its correlation and difference-probability matrices. He also gives a criterion that rules out a partition backdoor from the uniformities of the S-boxes alone, which shows, for instance, that AES has none [Ban17, Theorem 4.33], and he restates a criterion of Calderini based on S-boxes that send no large subspace onto a subspace [Ban17, Theorem 4.37]. These criteria certify that a backdoor is absent. Algorithm 1 and its spectral variant work in the other direction. They find the trapdoor when it is present, they read it off the support of the spectrum rather than its largest entries, and they come with proofs of soundness and completeness.

Which technique detects what. Beyond subspace-trail search, the techniques used across this paper are not interchangeable, and the mechanisms of Section 4.5 predict which apply where (Table 12).

The techniques detect different objects, and none of them covers the probabilistic partitions of Definition 15's relaxed form. In the rest of the paper we combine them in the order that Bannier et al.'s reduction justifies. We first screen each S-box for structure, then test whether the structure survives the linear layer, and only then treat it as a trapdoor of the cipher. The second step is what separates a real trapdoor from an S-box anomaly that diffusion destroys, as the comparison of KG-DES with DES's own S_4 shows (Section 6.5).

Table 12: Applicability of the detection techniques used or discussed in this paper. No single technique covers the whole taxonomy of Table 5, which is the main structural obstacle to a general backdoor detector.

Technique	Detects	Fails on	Cost
Walsh spectrum, magnitude at $\lambda = 0$	RP type-II relations	Balanced S-boxes (Prop. 2); relations an expansion builds in the round function (Prop. 1)	$O(m2^{m+n})$ blind
Walsh spectrum, support geometry	Partition trapdoors (Props. 3, 4)	Probabilistic partitions	$O(m2^{m+n})$
Difference-span scan (Alg. 1)	Partition trapdoors	Capped at $\|C\| >$ cap	$O(2^{2m}n)$ + fallback
Linear-structure / DLCT table	Derivative confinement	Structures that die in the linear layer	$O(2^{m+n}m)$
Round-level closure (Alg. 3)	Whether an S-box structure survives diffusion	Partial-activation trails	$O(1)$ or one elimination
Invariant-subspace search [LMR15]	Key-dependent invariant cosets	Key-independent partitions	Varies
Rank test [PW20, Prop. 1]	Reused MALICIOUS tweak pairs	Distinct pairs (Thm. 4)	One rank computation

6 KG-DES: Paterson's Trapdoor in DES

The detector of Section 5 is only worth having if it finds real backdoors, so the next three sections test it on real backdoored ciphers. We run the detector and then the full attack battery on each cipher, and we report where the predictions of the taxonomy of Section 4 hold and where they break. KG-DES is the natural first case. Paterson introduced the imprimitivity framework together with a worked instance, a DES-like cipher with eight new S-boxes, and KG-DES places those S-boxes in standard DES with everything else unchanged. The construction is published, so the trapdoor the algorithm returns can be compared with the one the designer intended. DES's linear layer and its cryptanalysis are also among the best understood in the literature, so every trail the trapdoor opens can be set against a known reference.

In this section we define the cipher, recover the trapdoor from the S-box tables with Algorithm 1, and carry it through DES's expansion and permutation to an invariant subspace of the whole cipher. We then ask what the trapdoor gives an attacker under each standard technique, i.e., connectivity tables, Rijmen–Preneel-type relations, differential, linear, differential-linear and boomerang trails, and the correlation-matrix view. Little of this is in the literature. It records Paterson's construction and its purpose and, from a companion study [FAI+23], the DLCT and FBCT of S_1 at their bounds. Section 6.11 separates what is reproduced from what is new.

6.1 The Cipher

Paterson's imprimitivity is a property of the whole cipher (Section 4.1); Bannier et al.'s reduction brings it down to a single S-box, where it becomes Definition 4. The S-box maps the cosets of a non-trivial input subspace onto the cosets of a non-trivial output subspace, in other words it has a trapdoor pair. A cipher built from such S-boxes can carry that structure, undetected, through the round function to the whole cipher, and this is what makes the construction a *backdoor*. A designer who knows the pair $(\mathbf{U}, \mathbf{W})$ can mount a probability-1 distinguisher or a truncated-differential attack directly (Theorem 2, Corollary 3). An evaluator who sees only the S-box lookup tables finds them balanced and of ordinary nonlinearity, and the trapdoor is stated nowhere. Section 6.5 measures how

far this concealment really extends, and finds the differential spectrum less innocent than the rest. KG-DES, defined next, places the S-boxes of Paterson's own instance in DES. In Section 6.2 we recover $(\mathbf{U}, \mathbf{W})$ from the S-box tables alone with Algorithm 1, as an evaluator would have to.

Definition 17 (KG-DES)**.** KG-DES is standard DES (the same 16-round Feistel structure, 64-bit block, 56-bit key schedule, expansion permutation E ($32 \to 48$ bits) and bit permutation P ($32 \to 32$ bits)) with one change. Each of the eight DES S-boxes is replaced by the corresponding S-box of Paterson's example cipher [Pat99], a deliberately weakened substitute that we denote $S_1, \ldots, S_8$ throughout this paper. Paterson's own cipher uses the same E, P and S-boxes with 32 rounds and an 80-bit key (Remark 26). Paterson numbers them 1 to 8, and our earlier work [Fah24] calls them KG-0,...,KG-7, so that $S_i \leftrightarrow$ KG-$(i-1)$. Each $S_i : \mathbb{F}_2^6 \to \mathbb{F}_2^4$ is engineered so that the trapdoor subspace pair of Definition 19/Lemma 3 holds by design, while remaining unremarkable on the usual balance and nonlinearity criteria. Its differential spectrum is a partial exception, quantified in Remark 31, since the confinement forces a worse differential uniformity than a genuine DES S-box attains. The full 64-entry lookup tables for $S_1, \ldots, S_8$ are given in Paterson's original construction [Pat99] and are not reproduced here. Every quantity we use is derived from them, including the subspaces of Table 13, the trimmed DDT of Table 14 and the DDT counts quoted in Section 6.6. We recomputed each of these directly from the published tables, so all of them can be reproduced from the cited source.

Definition 18 (KG-DES Round Function and Encryption)**.** Write a plaintext block as $P \in \mathbb{F}_2^{64}$ and let $(L_0, R_0) := \mathrm{IP}(P) \in \mathbb{F}_2^{32} \times \mathbb{F}_2^{32}$, where IP is DES's standard initial permutation. The round function is

$$F(R, K) \;:=\; P\big(\pi(E(R) \oplus K)\big), \qquad R \in \mathbb{F}_2^{32},\ K \in \mathbb{F}_2^{48},$$

where $E : \mathbb{F}_2^{32} \to \mathbb{F}_2^{48}$ is DES's expansion permutation, $\pi := (S_1, \ldots, S_8) : \mathbb{F}_2^{48} \to \mathbb{F}_2^{32}$ applies the eight backdoored S-boxes of Definition 17 to the eight consecutive 6-bit blocks of its argument, and $P : \mathbb{F}_2^{32} \to \mathbb{F}_2^{32}$ is DES's bit permutation (traced explicitly in Table 16). For round keys $K_1, \ldots, K_{16} \in \mathbb{F}_2^{48}$, derived from a 56-bit master key by DES's own, unmodified key schedule, sixteen rounds are applied:

$$L_i := R_{i-1}, \qquad R_i := L_{i-1} \oplus F(R_{i-1}, K_i), \qquad i = 1, \ldots, 16,$$

and the ciphertext is $C := \mathrm{IP}^{-1}(R_{16}, L_{16})$ (note the final swap). KG-DES replaces only π, and IP, E, P, IP^{-1} and the key schedule are DES's own. IP and IP^{-1} are fixed, keyless, invertible bit permutations that only relabel which wire carries which bit. Every result in this paper is a property of the round function F applied to (L_i, R_i), and a relabelling of coordinates changes neither the dimension of a subspace nor any differential probability. We therefore omit IP and IP^{-1} from the per-round analysis.

Remark 26 (Paterson's own example has 32 rounds). Paterson's example cipher [Pat99] uses the same E, P and S-boxes, but with 32 rounds and an 80-bit key whose schedule he leaves unspecified. We place the S-boxes in standard sixteen-round DES, as in our earlier work [Fah24], because DES supplies a fixed key schedule and a reference cryptanalysis for every comparison. Most of our results do not depend on this choice. The invariant subspace of Theorem 2 and the distinguishers built on it (Corollary 3, Propositions 26 and 27, and the known-plaintext partition distinguisher of Section 6.10) hold for any number of rounds and any key schedule, so they apply unchanged to Paterson's cipher. The S-box and round-function results do not change either, and the same exhaustive search gives the optimal 32-round characteristic exactly. It has probability $2^{-43.87}$ with 31 active S-boxes, and $B_{31} = 42.45$. Two results are specific to sixteen rounds. At 32 rounds the last-round key recovery of Section 6.7 would need at least about 2^{47} chosen pairs per S-box,

by the estimate $2^{B_{31}+5}$. Its master-key step also relies on the DES key schedule, and Paterson's example does not specify a schedule. The linear characteristic of Proposition 25 reaches only correlation 2^{-32} over 32 rounds, which would need about the whole codebook.

6.2 Recovering the Trapdoor from the S-Box Tables

We ran Algorithm 1 on each of the eight lookup tables. It returns a trapdoor pair of dimension 2 for every S-box, with the same input subspace for all eight and the output subspaces of Table 13. We did not guess either subspace or read it off by inspection. The definition, lemma and table below record the output of the algorithm, and Example 2 re-derives $\mathbf{W}_1$ from the DDT by hand as a check. In every S-box the candidate set has seven elements and the greedy build fails Check, so the pair comes from the fallback (Remark 7). Our earlier, uncorrected criterion recovers the same pairs. The intermediate version, which trusted the greedy build without Check, returns a spurious 8-element set on every S-box (Remark 34).

The input half of the pair is the same 2-dimensional subspace for all eight S-boxes, and every later result is phrased in terms of it.

Definition 19. Let $S_i : \mathbb{F}_2^6 \to \mathbb{F}_2^4$ denote the i-th S-box of KG-DES, $i = 1, \ldots, 8$. Define

$$\mathbf{U} = \{0, 4, 8, 12\} \subset \mathbb{F}_2^6, \qquad \mathbf{U}^* = \mathbf{U} \setminus \{0\} = \{4, 8, 12\}.$$

$\mathbf{U}$ is a 2-dimensional linear subspace of $\mathbb{F}_2^6$, generated by $\{4, 8\}$, since $4 \oplus 8 = 12$. Name the input bits of an S-box $(b_1, \ldots, b_6)$, with b_1 the most significant, so that b_1, b_6 select the row and $b_2 b_3 b_4 b_5$ is the nibble of R that feeds it. Then $\mathbf{U} = \mathrm{span}\{b_3, b_4\}$, the two middle bits, which are the only two input positions that DES's expansion does not share with a neighbouring S-box (Lemma 4).

Lemma 3 (Probability-1 Truncated Differential)**.** *For each $i \in \{1, \ldots, 8\}$ there exists a 2-dimensional linear subspace $\mathbf{W}_i \subset \mathbb{F}_2^4$ such that*

$$\forall\, x \in \mathbb{F}_2^6, \quad \forall\, u \in \mathbf{U} : \qquad S_i(x) \oplus S_i(x \oplus u) \;\in\; \mathbf{W}_i. \tag{5}$$

In other words, the truncated differential $\mathbf{U} \to \mathbf{W}_i$ holds with probability exactly 1, as opposed to the ordinary point differentials of Table 14, each of which holds with probability at most 24/64.

Proof. The claim is finite and is verified by exhaustive computation. Since $|\mathbb{F}_2^6| = 64$ and $|\mathbf{U}^*| = 3$, verifying Equation (5) for one S-box requires checking $64 \times 3 = 192$ instances; across all eight S-boxes this is 1536 direct table lookups. Doing so for $S_1, \ldots, S_8$ yields, in every case, exactly the sets $\mathbf{W}_i$ listed in Table 13, each of size 4 with $0 \in \mathbf{W}_i$ and closed under $\oplus$, confirming that each $\mathbf{W}_i$ is a 2-dimensional linear subspace of $\mathbb{F}_2^4$. □

Table 13: Output subspaces $\mathbf{W}_i$ realizing the probability-1 truncated differential $\mathbf{U} \to \mathbf{W}_i$, for each of the 8 KG-DES S-boxes.

S-box(es)	$\mathbf{W}_i$
$S_1,\ S_2,\ S_3,\ S_6,\ S_7$	$\{0, 1, 2, 3\}$
S_4	$\{0, 2, 8, 10\}$
S_5	$\{0, 1, 4, 5\}$
S_8	$\{0, 2, 4, 6\}$

Table 14: Trimmed DDT of S_1 (KG-0), restricted to input differences $u \in \mathbf{U}^* = \{4, 8, 12\}$ and output differences $\Delta_{\text{out}} \in \{1, 2, 3\}$; entry (u, Δ_{out}) is $|\{x \in \mathbb{F}_2^6 : S_1(x) \oplus S_1(x \oplus u) = \Delta_{\text{out}}\}|$. Every other entry of the full 64×16 DDT of S_1 is 0 for $u \in \mathbf{U}^*$, which is Lemma 3's claim restricted to S_1.

$u \backslash \Delta_{\text{out}}$	1	2	3
4	24	24	16
8	24	16	24
12	16	24	24

Example 2 (Computing $\mathbf{W}_1$ by hand)**.** Table 14 (the trimmed DDT of S_1, restricted to inputs in $\mathbf{U}^*$) shows that every entry is nonzero and that no column outside $\{1, 2, 3\}$ is ever populated. Each row also sums to 64, so no probability mass escapes this set. Hence the image of $S_1(x) \oplus S_1(x \oplus u)$ over all x and all $u \in \mathbf{U}^*$ is exactly $\{1, 2, 3\}$, and together with the trivial case $u = 0$ (which contributes the difference 0), $\mathbf{W}_1 = \{0, 1, 2, 3\}$. Closure under $\oplus$ is immediate, since $1 \oplus 2 = 3$, $1 \oplus 3 = 2$ and $2 \oplus 3 = 1$ all remain in the set. Hence $\mathbf{W}_1$ is a linear subspace, spanned by $\{1, 2\}$.

Table 14 gives the *ordinary* (point) differential probabilities *within* $\mathbf{U} \to \mathbf{W}_1$, as used in Example 2. For instance, $u = 4 \to 1$ holds with probability 24/64, $u = 4 \to 2$ with probability 24/64 and $u = 4 \to 3$ with probability 16/64, which sum, as required, to $64/64 = 1$. Lemma 3 is the coarser, set-valued statement that subsumes all three rows of Table 14 simultaneously, with certainty rather than probability 24/64.

6.3 From the S-Boxes to the Cipher

A trapdoor pair in every S-box does not yet make a trapdoor of the cipher. DES's expansion duplicates some input bits into two S-boxes, and its permutation scatters the output of each S-box over the next round. In this subsection we show that both carry the trapdoor intact, first by inspecting the two tables and then by the mechanical closure of Section 5.5.

6.3.1 Through the Expansion

Lemma 3 is stated at the level of a single S-box, whose 6-bit input is a 6-bit block of $E(\Delta R)$. For the property to survive intact through the full round function, E must not mix a "safe" nibble difference (one confined to $\mathbf{U}$) with the row-selection bits shared between adjacent S-boxes. Otherwise, activating one S-box could perturb its neighbours unpredictably.

Lemma 4 (Non-Interference Under E)**.** *Let $\Delta R \in \mathbb{F}_2^{32}$ have every 4-bit nibble confined to $\{0, 2, 4, 6\}$ (equivalently: bit positions 1 and 4 of every nibble, counting from the most significant bit, are zero). Then for each $i \in \{1, \ldots, 8\}$, the 6-bit block of $E(\Delta R)$ feeding S_i is nonzero only if the i-th nibble of ΔR is nonzero, and in that case it equals an element of $\mathbf{U}^*$.*

Proof. By construction of the DES expansion table E, the 6-bit block feeding S_i consists of the last bit of nibble $i - 1$, the four bits of nibble i, and the first bit of nibble $i + 1$ (indices taken modulo 8). Suppose every nibble of ΔR has first and last bits (r_1, r_4) equal to zero. Then the boundary bits contributed by nibbles $i \mp 1$ are zero whatever their other values, and the 6-bit block reduces to $(0, r_1, r_2, r_3, r_4, 0) = (0, 0, r_2, r_3, 0, 0)$, where $r_1 r_2 r_3 r_4$ are the bits of nibble i and $r_1 = r_4 = 0$. This is an element of $\mathbf{U} = \{0, 4, 8, 12\}$, the span of positions 3 and 4 of the 6-bit block. These are the only two positions that E does not duplicate into a neighbouring block. It is zero precisely when nibble i is zero. An

enumeration of all 48 output bits of E against all 8 nibble positions confirms this, and Section 5.5 recasts it as the output of a general algorithm. □

6.3.2 Through the Bit Permutation

Table 13 records each $\mathbf{W}_i$ as a set, but not which of the four output-nibble bit positions are *free* (vary across $\mathbf{W}_i$) and which are identically zero. Writing nibble bits in DES numbering ($p = 1$ most significant, $p = 4$ least significant, so position p carries weight 2^{4-p}), Table 15 makes this explicit.

Table 15: Free bit position(s), within its own pre-P nibble, of each $\mathbf{W}_i$. These are *not* the same for every box, and only S_8's free positions happen to be $\{2, 3\}$.

S-box(es)	$\mathbf{W}_i$	Free pre-P position(s)
$S_1,\ S_2,\ S_3,\ S_6,\ S_7$	$\{0, 1, 2, 3\}$	$\{3, 4\}$
S_4	$\{0, 2, 8, 10\}$	$\{1, 3\}$
S_5	$\{0, 1, 4, 5\}$	$\{2, 4\}$
S_8	$\{0, 2, 4, 6\}$	$\{2, 3\}$

Lemma 5 (Closure Under P)**.** *Let ΔR satisfy the hypothesis of Lemma 4, and let ΔF be the corresponding round-function output difference. Then every nibble of ΔF also has its first and last bits equal to zero, i.e. ΔF again has every nibble confined to $\{0, 2, 4, 6\}$.*

Proof. By Lemma 3, every active S_i contributes an output difference lying in $\mathbf{W}_i$, a 2-dimensional subspace of $\mathbb{F}_2^4$. As Table 15 shows, the two free coordinates of $\mathbf{W}_i$ do *not* occupy the same pre-P nibble position for every box. Only the free positions of S_8 are already $\{2, 3\}$. The rest are $\{3, 4\}$ (S_1, S_2, S_3, S_6, S_7), $\{1, 3\}$ (S_4) or $\{2, 4\}$ (S_5). What the eight boxes have in common is not their pre-P position but their *destination* once P is applied. Embed the local free position p of box i into the pre-P 32-bit coordinate as $g = 4(i - 1) + p$, and read off the post-P position as the index k with $P[k] = g$. We trace all 16 free bit positions (two per S-box) through the P-table (Table 16). Every one of them lands on bit position 2 or 3 of *some* nibble of the 32-bit output, and never on position 1 or 4, although they start from different positions within their own nibbles. Inactive S-boxes, and the non-free bit positions of active S-boxes, are identically zero by Lemma 3. Hence no bit position 1 or 4 of any output nibble can be set, for any x and any branch realized at the active S-boxes. □

Table 16: Verification of Lemma 5 by direct enumeration: every one of the 16 free pre-P bit positions, once traced through the actual DES P-table, lands on nibble-local position 2 or 3 of the post-P output, even though (Table 15) the free positions themselves are not uniform across boxes. The same enumeration discharges the containment $P(\mathbf{W}) \subseteq \mathbf{T}_0$ required by Proposition 16.

Box i	local pos. p	pre-P global g	post-P global k	nibble-local position of k
1	3	3	23	3
1	4	4	31	3
2	3	7	2	2
2	4	8	18	2
3	3	11	30	2
3	4	12	6	2
4	1	13	26	2
4	3	15	10	2
5	2	18	14	2
5	4	20	3	3
6	3	23	11	3
6	4	24	19	3
7	3	27	22	2
7	4	28	7	3
8	2	30	27	3
8	3	31	15	3

6.3.3 The Invariant Subspace and a Two-Plaintext Distinguisher

Paterson proves that the pairs of cosets of this subspace form a complete block system for the group generated by the round functions, for every round key [Pat99, Lemmas 2 and 3]. We restate his result in the form of differences, which is the form the trails below use, and give the short proof for completeness.

Theorem 2 (Invariant subspace of KG-DES, after Paterson [Pat99])**.** *Let*

$$\mathbf{S} \;=\; \big\{\, \delta \in \mathbb{F}_2^{32} \;:\; \textit{every nibble of } \delta \textit{ lies in } \{0,2,4,6\} \,\big\}.$$

$\mathbf{S}$ *is a* 16*-dimensional linear subspace of* $\mathbb{F}_2^{32}$*. For any round key* $K \in \mathbb{F}_2^{48}$*: if* $\Delta R \in \mathbf{S}$ *then the resulting round-function output difference* $\Delta F \in \mathbf{S}$*, with probability* 1*. Consequently, for KG-DES with any key schedule and any number of rounds* r*: if* $(\Delta L_0, \Delta R_0) \in \mathbf{S} \times \mathbf{S}$*, then* $(\Delta L_r, \Delta R_r) \in \mathbf{S} \times \mathbf{S}$ *for every* $r \geq 0$*, with probability* 1*.*

Proof. $\mathbf{S}$ is defined by the vanishing of 16 fixed linear coordinates (two per nibble), so it is a linear subspace of dimension $32 - 16 = 16$. The one-round claim combines Lemmas 4 and 5. If $\Delta R \in \mathbf{S}$, then $\Delta F \in \mathbf{S}$, independently of the round key K, since K cancels in the difference $F(R, K) \oplus F(R', K)$ whatever its value. The Feistel update sets $\Delta L_{i+1} = \Delta R_i$, which lies in $\mathbf{S}$ by the inductive hypothesis, and $\Delta R_{i+1} = \Delta L_i \oplus \Delta F(\Delta R_i)$. Since $\mathbf{S}$ is a linear subspace it is closed under $\oplus$, so $\Delta L_i, \Delta R_i \in \mathbf{S}$ together imply $\Delta L_{i+1}, \Delta R_{i+1} \in \mathbf{S}$. The full-cipher statement follows by induction on r, and the argument at each step does not depend on the round key. □

Corollary 3 (Probability-1 distinguisher for KG-DES)**.** *For any master key* $K \in \mathbb{F}_2^{56}$ *and any plaintext pair* (P_i, P_j) *with* $\mathrm{IP}(P_i \oplus P_j) \in \mathbf{S} \times \mathbf{S}$*, let* $D(c, c') := [\,\mathrm{IP}(c \oplus c') \in \mathbf{S} \times \mathbf{S}\,]$ *be the predicate an adversary tests on the resulting ciphertext pair. By Theorem 2, applied to the sixteen rounds,* IP *of the ciphertext difference,* $(\Delta R_{16}, \Delta L_{16})$*, lies in* $\mathbf{S} \times \mathbf{S}$ *for every key. The final swap exchanges two components that both lie in* $\mathbf{S}$*. Hence*

$$\Pr\big[D(\mathcal{E}_{\mathrm{KG}}(K, P_i), \mathcal{E}_{\mathrm{KG}}(K, P_j)) = 1\big] \;=\; 1.$$

For an ideal random permutation Π *on* $\mathbb{F}_2^{64}$ *in place of* $\mathcal{E}_{\mathrm{KG}}(K,\cdot)$*, the same predicate holds only for the fraction of ciphertext differences lying in* $\mathbf{S}\times\mathbf{S}$*:*

$$\Pr\big[D(\Pi(P_i),\Pi(P_j))=1\big] \;=\; \frac{|\mathbf{S}\times\mathbf{S}|}{2^{64}} \;=\; \left(\frac{|\mathbf{S}|}{2^{32}}\right)^2 = 2^{-32}$$

(up to a $O(2^{-64})$ *correction for the excluded event* $\Pi(P_i)=\Pi(P_j)$*). The adversary that submits this single pair and outputs* D *therefore achieves*

$$\mathrm{Adv}^{\mathrm{prp}}_{\mathcal{A}} \;=\; 1-2^{-32}$$

from two chosen plaintexts and one comparison.

Remark 27 (The distinguisher in concrete terms). Paterson's basic attack [Pat99, Section 5] uses the same block structure with 2^{32} chosen plaintexts to recover the permutation that the cipher induces on the blocks. As a distinguisher, two plaintexts are enough. In operational terms, Corollary 3 is a two-query bit-pattern test. The ciphertext difference carries 32 constrained bit positions, positions 1 and 4 of each of the sixteen nibbles, eight per half. Every one of them is 0 with certainty, whatever the key and the round count, while the other 32 bits are unconstrained. An adversary needs only to request the pair and read those 32 bits. The final permutation IP^{-1} is fixed and public, so it moves the constrained bits to other ciphertext positions but changes neither their number nor the advantage (Definition 18).

Remark 28 (Reachability). The round function does not only preserve the invariant subspace $\mathbf{S}$. Interestingly, it also explores all of it. We followed 3,000 trajectories of 16 rounds, with random round keys and random starting differences in $\mathbf{S}\times\mathbf{S}$. The observed differences ΔR_r span all of $\mathbf{S}$, and for each of the 28 pairs of nibble positions (adjacent or not) their projections take all 16 values of $\{0,2,4,6\}^2$. The trajectories are therefore not confined to a smaller subspace of $\mathbf{S}$. A single trajectory of 16 rounds does not reach all 16 values of a window, so the statement concerns the union over trajectories. Unlike Theorem 2, this is an empirical observation (Table 35) and not a proof, which is why we state it as a remark.

Example 3 (Reachability on a two-nibble window)**.** Restricted to any two of the eight nibble positions of ΔR, the set of values observed across many rounds is

$$\{00,02,04,06,20,22,24,26,40,42,44,46,60,62,64,66\}\ \text{(hex)},$$

a subspace of dimension 4, the direct sum $\{0,2,4,6\}\times\{0,2,4,6\}$, spanned by $\{02,04,20,40\}$. This makes Remark 28 concrete. Only 16 of the 256 possible byte values ever occur at that window, and, by the reachability property, all 16 of them eventually do.

6.3.4 The Same Subspace by Machine

Proposition 15 and Algorithm 3 reproduce $\mathbf{S}$ from $\mathbf{U}$ and the E-table alone, with no reasoning about individual bit positions.

Example 4 (Algorithm 3 applied to S_1 alone)**.** For $\mathbf{U}=\{0,4,8,12\}$, the local zero-positions are $Z_1=\{1,2,5,6\}$ (positions $3,4$, corresponding to nibble bits r_2,r_3, remain free). For $i=1$, the global positions in the 48-bit output of E are $g=6(1-1)+p=p$ for $p\in\{1,2,5,6\}$, i.e. positions $1,2,5,6$ of E's output. Consulting the DES expansion table at these positions gives source bits $E[1]=32$, $E[2]=1$, $E[5]=4$, $E[6]=5$. Hence S_1 alone contributes the raw constraints

$$Z \;\supseteq\; \{32,1,4,5\}$$

to the running union. We repeat this for $i = 2, \ldots, 8$ and take the union of all $8 \times 4 = 32$ raw constraints. Because each boundary bit is claimed by two neighbouring boxes, the union collapses to the 16 distinct positions

$$Z = \{1, 4, 5, 8, 9, 12, 13, 16, 17, 20, 21, 24, 25, 28, 29, 32\}.$$

Grouped by nibble, these are positions 1 and 4 of every nibble. The algorithm thus reproduces $\mathbf{S}$ of Theorem 2 mechanically, and Lemma 4 becomes an instance of it.

For KG-DES we carried out the P-closure check of Proposition 16 on all 16 free output-bit coordinates, two per S-box (listed in Table 15 and traced through P in Table 16). It holds in every case, so $\mathbf{T}_0 = \mathbf{S}$ is an invariant subspace, and Lemma 5 becomes an instance of Proposition 16.

6.3.5 Why DES's Linear Layer Admits the Trapdoor

KG-DES keeps DES's E and P; only the S-boxes change. The trapdoor subspaces turn out to be the ones DES's linear layer admits, and the layer's branch number, 2 (Table 47), is what later lets single trapdoor transitions chain from round to round.

Proposition 17 (DES's linear layer determines KG-DES's subspaces)**.** *Let* $\mathbf{S}$ *be spanned by the* 16 *bits of* R *that* E *does not duplicate, bits* 2 *and* 3 *of every nibble. Then* $E(\mathbf{S})$ *occupies exactly the two middle input bits of every S-box, which is* $\mathbf{U} = \{0, 4, 8, 12\}$*. For every* i*,* P *sends exactly two of the four output bits of* S_i *into* $\mathbf{S}$*, and the span of those two bits is* $\mathbf{W}_i$*, for all eight S-boxes, including the irregular* $\mathbf{W}_4 = \{0, 2, 8, 10\}$*,* $\mathbf{W}_5 = \{0, 1, 4, 5\}$ *and* $\mathbf{W}_8 = \{0, 2, 4, 6\}$*. Consequently* $\mathbf{W}_i$ *is the largest subspace of* $\mathbb{F}_2^4$ *spanned by output bits of* S_i *with* $P(\mathbf{W}_i) \subseteq \mathbf{S}$*.*

Proof. The middle bits of each nibble appear once in E, at input positions 3 and 4 of one S-box, and the end bits appear twice, as the E table shows. For P we look up the 32 output bits directly. For S_1, S_2, S_3, S_6 and S_7 they are bits 3 and 4 (values 2 and 1), for S_4 bits 1 and 3 (values 8 and 2), for S_5 bits 2 and 4 (values 4 and 1), and for S_8 bits 2 and 3 (values 4 and 2). The spans are the stated $\mathbf{W}_i$. □

Interestingly, routing two output bits of every S-box to middle bits and two to end bits is one of DES's own published design criteria for P [Cop94], chosen for diffusion in the 1970s and published in 1994, before Paterson's construction. A random bit permutation of 32 bits has the property with probability $6^8/\binom{32}{16} = 2^{-8.48}$. The evidence therefore points one way. DES's linear layer was not built for the trapdoor. Paterson built the trapdoor around it, as he explains himself [Pat99, Section 5], and the subspaces he used, down to the irregular $\mathbf{W}_4$, $\mathbf{W}_5$ and $\mathbf{W}_8$, are the ones E and P dictate. All of the design freedom is in the S-boxes. It follows that the linear layer of standard DES is ready for the trapdoor as it stands. Installing in DES any S-boxes with $D_u S_i(x) \in \mathbf{W}_i$ for $u \in \mathbf{U}$ produces the invariant subspace of Theorem 2.

6.4 Connectivity Tables of the Trapdoor S-Boxes

A trapdoor that confines every derivative to a proper subspace should leave a visible trace in the Differential-Linear Connectivity Table (Definition 11). This table counts how often a linear mask annihilates a derivative. In this subsection we show that for KG-DES the trace is in fact extremal, and we derive this from Lemma 3 through the Walsh spectrum. We then explain the remaining extremal entries by the one quadratic trapdoor component of each S-box, and show that the Feistel Boomerang Connectivity Table (Definition 12) is extremal as well. KG-DES's S-boxes are 6×4 and non-square. That rules out the Boomerang Connectivity Table, since it needs a bijection. It does not rule out the DLCT

or the FBCT, since both are defined for any function and were introduced for Feistel round functions of exactly this kind [BHL+20]. Section 7.5 later runs the same analysis against BEA-1's square, bijective S-boxes, where all three tables apply.

Proposition 18 (DLCT attains its exact upper bound on KG-DES's trapdoor directions)**.** *For every $i \in \{1,\dots,8\}$, every $u \in \mathbf{U}^* = \{4, 8, 12\}$, and every γ in the annihilator of $\mathbf{W}_i$ (a 2-dimensional subspace of $\mathbb{F}_2^4$, so 3 nonzero γ's qualify for each i),* $\mathrm{DLCT}_{S_i}(u,\gamma) = 2^5 = 32$ *exactly: the theoretical maximum for this 6×4 S-box.*

Proof. Fix i and a nonzero $\gamma \in \mathbf{W}_i^{\perp}$, write $m = 6$, and let $f := \gamma \cdot S_i$. The argument has three steps.

The DLCT entry is half an autocorrelation, and a quadratic form in the correlations. The count in Definition 11 and the number of x with $\gamma \cdot D_\Delta S_i(x) = 1$ add up to 2^m, so $\mathrm{DLCT}_{S_i}(\Delta,\gamma)$ is half their difference. Expanding the square of the Walsh transform,

$$\sum_{\lambda}(-1)^{\lambda\cdot\Delta}\,\widehat{S_i}(\lambda,\gamma)^2 \;=\; \sum_{x,y}(-1)^{f(x)\oplus f(y)}\sum_{\lambda}(-1)^{\lambda\cdot(x\oplus y\oplus\Delta)},$$

and the inner sum is 2^m if $y = x \oplus \Delta$ and 0 otherwise. Together,

$$\mathrm{DLCT}_{S_i}(\Delta,\gamma) \;=\; \tfrac{1}{2}\sum_{x\in\mathbb{F}_2^m}(-1)^{f(x)\oplus f(x\oplus\Delta)} \;=\; 2^{-m-1}\sum_{\lambda\in\mathbb{F}_2^m}(-1)^{\lambda\cdot\Delta}\,\widehat{S_i}(\lambda,\gamma)^2. \tag{6}$$

In the correlations $C^{S_i}_{\gamma,\lambda} = 2^{-m}\widehat{S_i}(\lambda,\gamma)$ of Definition 10 this reads

$$\frac{\mathrm{DLCT}_{S_i}(\Delta,\gamma)}{2^{m-1}} \;=\; \sum_{\lambda}(-1)^{\lambda\cdot\Delta}\big(C^{S_i}_{\gamma,\lambda}\big)^2 :$$

the differential-linear correlation is a signed sum of squared linear correlations.

Parseval caps it. For $\gamma \neq 0$, $\sum_\lambda \big(C^{S_i}_{\gamma,\lambda}\big)^2 = 1$. Hence $\mathrm{DLCT}_{S_i}(\Delta,\gamma) \le 2^{m-1}$, with equality if and only if $C^{S_i}_{\gamma,\lambda} = 0$ for every λ with $\lambda\cdot\Delta = 1$, i.e., if and only if $\mathrm{supp}_\gamma(S_i) \subseteq \Delta^{\perp} := \{\lambda : \lambda\cdot\Delta = 0\}$.

The trapdoor makes the other correlations vanish. By Lemma 3, $D_u S_i(x) \in \mathbf{W}_i$ for every x and every $u \in \mathbf{U}$, and γ annihilates $\mathbf{W}_i$, so $f(x\oplus u) = f(x)$, and f is constant on every coset of $\mathbf{U}$. Grouping the Walsh sum by cosets $t \oplus \mathbf{U}$, with t running over a set of coset representatives,

$$\widehat{S_i}(\lambda,\gamma) \;=\; \sum_{t}(-1)^{f(t)\oplus\lambda\cdot t}\sum_{u\in\mathbf{U}}(-1)^{\lambda\cdot u},$$

and the inner sum vanishes unless $\lambda \in \mathbf{U}^{\perp}$. Hence $\mathrm{supp}_\gamma(S_i) \subseteq \mathbf{U}^{\perp} \subseteq u^{\perp}$ for every $u \in \mathbf{U}$, and the bound is attained, with $\mathrm{DLCT}_{S_i}(u,\gamma) = 2^{-7}\cdot 2^{12} = 2^5$. □

The proof gives an exact criterion, not only the bound. For any π, $\mathrm{DLCT}_\pi(\Delta,\gamma) = 2^{m-1}$ if and only if $\mathrm{supp}_\gamma(\pi) \subseteq \Delta^{\perp}$. Proposition 3 applies the same identity to a whole subspace of directions at once. Algebraic degree plays no part. In every S_i, two of the three trapdoor components $\gamma\cdot S_i$, $\gamma \in \mathbf{W}_i^{\perp}\setminus\{0\}$, have degree 3 and one has degree 2 (for S_1, degree 3 at $\gamma = 4, 8$ and 2 at $\gamma = 12$), and all three attain the bound. The quadratic one accounts for the table's remaining maximizers (Proposition 19).

We checked Proposition 18 exhaustively, on all 63×15 nonzero pairs (Δ,γ) per S-box. The DLCT uniformity is 32 for every one of $S_1,\dots,S_8$, which independently reproduces the value published earlier for S_1 by a companion study [FAI+23]. In every S-box the bound is attained at 13 pairs, the $9 = 3\times 3$ of the proposition and 4 more, all in the column of the quadratic trapdoor component. The next proposition derives those four from the degree.

Proposition 19 (The quadratic trapdoor component and the remaining DLCT maximizers)**.** *In every S_i exactly one of the three components $\gamma \cdot S_i$, $\gamma \in \mathbf{W}_i^{\perp} \setminus \{0\}$, has algebraic degree 2. Let γ_i^{q} be its mask (12 for S_1, S_2, S_3, S_6, S_7; 5 for S_4; 10 for S_5; 9 for S_8) and $f := \gamma_i^{\mathrm{q}} \cdot S_i$. The directions Δ in which $D_\Delta f$ is constant form a 4-dimensional subspace $\mathbf{N}_i \supseteq \mathbf{U}$; those in which $D_\Delta f \equiv 0$ form a 3-dimensional subspace $\mathbf{N}_i^0 = \mathbf{U} \cup (d_i \oplus \mathbf{U})$ for some $d_i \notin \mathbf{U}$; and*

$$\mathrm{DLCT}_{S_i}(\Delta, \gamma_i^{\mathrm{q}}) \;=\; \begin{cases} 2^5 & \Delta \in \mathbf{N}_i^0 \setminus \{0\}, \\ -2^5 & \Delta \in \mathbf{N}_i \setminus \mathbf{N}_i^0, \\ 0 & \Delta \notin \mathbf{N}_i. \end{cases}$$

In particular $(\Delta, \gamma_i^{\mathrm{q}})$ attains the bound at exactly four directions outside $\mathbf{U}$, those of $d_i \oplus \mathbf{U}$.

Proof. The degrees are read off the algebraic normal forms; the discussion after the proof explains them. Write $\Phi(a, x) := f(x \oplus a) \oplus f(x) \oplus f(a) \oplus f(0)$. Because f has degree 2, Φ is a symmetric bilinear form vanishing on the diagonal, and

$$D_a f(x) \;=\; \Phi(a, x) \oplus f(a) \oplus f(0)$$

is affine in x. If a lies in the radical $\{a : \Phi(a, \cdot) \equiv 0\}$, then $D_a f$ is the constant $f(a) \oplus f(0)$ and, by Equation (6), the DLCT entry is $\pm 2^5$ according to that constant. Otherwise $D_a f$ is a nonconstant affine function, hence balanced, and the entry is 0. Thus $\mathbf{N}_i$ is the radical, a subspace. Φ vanishes on it, so $a \mapsto f(a) \oplus f(0)$ is linear on $\mathbf{N}_i$, and $\mathbf{N}_i^0$ is the kernel of this map, so it is either all of $\mathbf{N}_i$ or a hyperplane in it. This gives the three cases of the display.

The dimensions follow from balance. The proof of Proposition 18 shows that f is constant on the cosets of $\mathbf{U}$, so $\mathbf{U} \subseteq \mathbf{N}_i^0$ and $f = g \circ q$ for the quotient map $q : \mathbb{F}_2^6 \to \mathbb{F}_2^6/\mathbf{U} \cong \mathbb{F}_2^4$ and a quadratic g whose form has the same rank as Φ. That rank is even and nonzero, so it is 2 or 4. Rank 4 would make g a bent function of 4 variables, of weight 6 or 10, and f of weight 24 or 40. However, each row of S_i is a permutation of $\mathbb{F}_2^4$, so every nonzero component of S_i has weight 32. Hence the rank is 2 and $\dim \mathbf{N}_i = 6 - 2 = 4$. If $\mathbf{N}_i^0$ were all of $\mathbf{N}_i$, then f would be constant on the cosets of $\mathbf{N}_i$ and would factor through $\mathbb{F}_2^6/\mathbf{N}_i \cong \mathbb{F}_2^2$ by a quadratic with nondegenerate form. This is a bent function of 2 variables, of weight 1 or 3, which would make f of weight 16 or 48, again not 32. Therefore $\mathbf{N}_i^0$ is a hyperplane of $\mathbf{N}_i$, of dimension 3, and it contains $\mathbf{U}$. □

Propositions 18 and 19 together account for all 13 maximizers, and also for the 8 entries at the minimum -2^5, which are the directions in $\mathbf{N}_i \setminus \mathbf{N}_i^0$. For S_1, $d_1 = 34$ and $\mathbf{N}_1 = \mathbf{U} \oplus \langle 18, 34 \rangle$. The second direction is weaker than the first. For $\Delta \in d_i \oplus \mathbf{U}$ the derivative $D_\Delta S_i$ is confined only to the 3-dimensional subspace of $\mathbb{F}_2^4$ that γ_i^{q} annihilates, not to $\mathbf{W}_i$, so that one mask is the only one reaching the bound there. $D_{34} S_1$ takes all 8 values of that subspace; in S_4 and S_6, two of the four directions of $d_i \oplus \mathbf{U}$ reach only 7 of them. The direction also moves a boundary bit and so does not survive the expansion (Remark 29).

The degrees come from the row structure of a DES-type S-box. In the bit names of Definition 19, a row, fixed by (b_1, b_6), is a permutation of $\mathbb{F}_2^4$ and is the union of the four cosets of $\mathbf{U}$ indexed by (b_2, b_5). Each of these cosets goes into one coset of $\mathbf{W}_i$, and the four images are disjoint, so they are the four cosets of $\mathbf{W}_i$. Hence $(b_2, b_5) \mapsto S_i(x) \bmod \mathbf{W}_i$ is a bijection $\mathbb{F}_2^2 \to \mathbb{F}_2^4/\mathbf{W}_i$. Every bijection between 2-dimensional spaces over $\mathbb{F}_2$ is affine, so each row r has a linear part $A_{i,r} \in \mathrm{GL}_2(\mathbb{F}_2)$. A trapdoor component $\gamma \cdot S_i$ is therefore affine in (b_2, b_5) on each row and does not depend on b_3, b_4. Its only possible cubic monomials are $b_1 b_6 b_2$ and $b_1 b_6 b_5$, and their coefficients are the two entries of $\gamma \circ A_i$. Here $A_i := \bigoplus_r A_{i,r}$ is the sum over $\mathbb{F}_2$ of the four row maps (a single 2×2 matrix, not a direct sum), and γ is read as a linear form on $\mathbb{F}_2^4/\mathbf{W}_i$. The component has degree at most 2 exactly when γ annihilates the image of A_i, so S_i has $2^{2-\operatorname{rank} A_i} - 1$ trapdoor components of degree at

most 2 and degree 3 for the rest. All eight KG-DES S-boxes have rank $A_i = 1$. For four row maps drawn independently and uniformly from $\mathrm{GL}_2(\mathbb{F}_2)$, rank 1 has probability 5/9, rank 0 has probability 2/27, and rank 2, which would leave no quadratic component and no second direction, has probability 10/27. Paterson's own design criteria rule out rank 2.

Proposition 20 (Paterson's criteria force a quadratic trapdoor component)**.** *Let S be a 6×4 S-box with the structure above: each row a permutation, and each coset of $\mathbf{U}$ sent onto a coset of $\mathbf{W}$. If S satisfies Paterson's criteria S2 and S6 [Pat99, §4], then* $\operatorname{rank} A \le 1$*, so at least one trapdoor component of S has degree exactly* 2*.*

Proof. Write row r of the label map as $y \mapsto A_r y \oplus c_r$, where $y = (b_2, b_5)$, and fix a trapdoor component $f = \gamma \cdot S$. Since f does not depend on b_3, b_4, its Walsh transform vanishes unless the mask λ is zero on those bits. For such a λ, with row part λ_r and label part λ_y,

$$\widehat{S}(\lambda,\gamma) \;=\; 16 \sum_{r\,:\,\gamma\circ A_r=\lambda_y} (-1)^{\gamma(c_r)\oplus\lambda_r\cdot r}.$$

S2 bounds $|\widehat{S}|$ by 32. If exactly three rows had the same form $\gamma \circ A_r = \beta$, we could take $\lambda_y = \beta$ and choose λ_r so that the three signs agree, because any three distinct rows give two independent differences. That would give 48, so no form occurs in exactly three rows. S6 says that inputs differing in $11xy00$ never collide. Such inputs lie in rows r and r' that differ in b_1 and in labels y and $y \oplus (1,0)$. If $A_r y \oplus c_r = A_{r'}(y \oplus (1,0)) \oplus c_{r'}$ for some y, the two cosets would map onto the same coset of $\mathbf{W}$, and two of their inputs would collide. Hence the affine map $y \mapsto (A_r \oplus A_{r'})y \oplus A_{r'}(1,0) \oplus c_r \oplus c_{r'}$ never takes the value 0, and $B_0 := A_{00} \oplus A_{10}$ and $B_1 := A_{01} \oplus A_{11}$ are singular.

Suppose $\operatorname{rank} A = 2$, where $A = B_0 \oplus B_1$. Then B_0 and B_1 both have rank 1, with left kernels $\{0, \gamma_0\}$ and $\{0, \gamma_1\}$, and $\gamma_0 \neq \gamma_1$, since otherwise $\gamma_0 A = 0$. Put $\beta := \gamma_0 A_{00} = \gamma_0 A_{10}$. The forms $\gamma_0 A_{01}$ and $\gamma_0 A_{11}$ differ, because $\gamma_0 B_1 \neq 0$, and neither equals β, since β would then occur in exactly three rows. They are therefore the two other nonzero forms. In the same way, with $\beta_1 := \gamma_1 A_{01} = \gamma_1 A_{11}$, the forms $\gamma_1 A_{00}$ and $\gamma_1 A_{10}$ are the two nonzero forms other than β_1. If $\beta \neq \beta_1$, one of these two is β, and then $(\gamma_0 \oplus \gamma_1)A_r = 0$ for $r = 00$ or 10, which is impossible for an invertible A_r. Thus $\beta = \beta_1$. The two nonzero forms other than β sum to β, so $(\gamma_0 \oplus \gamma_1)A = \gamma_0 B_1 \oplus \gamma_1 B_0 = \beta \oplus \beta_1 = 0$, contradicting $\operatorname{rank} A = 2$. Finally, a trapdoor component of degree at most 1 would be affine and would violate S2, so the components of degree at most 2 have degree exactly 2. □

Rank 0, with three quadratic components, is not excluded. We drew 1.5×10^7 S-boxes uniformly with this structure and kept the 564 that satisfy all of S2 and S4–S8. Of these, 563 have rank 1 and one has rank 0. Paterson's criteria therefore explain why every KG-DES S-box has a quadratic trapdoor component, and they make exactly one such component the typical case.

The companion study [FAI+23] also reports the Feistel Boomerang Connectivity Table of S_1 at its bound. That too follows from the trapdoor, one derivative further on.

Proposition 21 (FBCT attains its bound on KG-DES's trapdoor directions)**.** *For every i and all distinct $u, u' \in \mathbf{U}^*$, $\mathrm{FBCT}_{S_i}(u, u') = 2^6 = 64$, the largest value the table (Definition 12) can take.*

Proof. The four inputs x, $x \oplus u$, $x \oplus u'$, $x \oplus u \oplus u'$ form the coset $x \oplus \mathbf{U}$, and S_i maps that coset into a single coset of $\mathbf{W}_i$ (Lemma 3). On every coset the map is injective, as a direct check of all 16 cosets of each S-box shows, so it is a bijection between two 4-element affine spaces. Every such bijection is affine, since there are 24 permutations of four points and 24 affine bijections of $\mathbb{F}_2^2$. An affine map sends the four points of a 2-dimensional coset to four points summing to 0, so the defining equation holds for every x. □

Exhaustive computation over all ordered pairs of distinct nonzero 6-bit differences finds FBCT uniformity 64 for every S_i, attained at exactly the 6 ordered pairs of Proposition 21 and nowhere else. The DES S-boxes reach between 20 and 48. Section 7.6.2 shows that the same property makes the whole cipher affine on every coset of $\mathbf{S} \times \mathbf{S}$, with a boomerang of probability 1 as a consequence.

6.5 Rijmen–Preneel-Type Relations

KG-DES is already a backdoored cipher, so one can ask whether it also carries the structure of Section 3, by design or as a side effect of the partition. Its S-boxes cannot, because every row is a permutation of $\mathbb{F}_2^4$, so every component is balanced (Proposition 2). Two questions remain. What does a screen of the S-boxes find, and does the round function carry such a relation?

Remark 29 (What the screen does find: a larger linear-structure subspace). Running the corresponding screen for linear structures (all (α, γ) with $\gamma \cdot D_\alpha S_i$ constant in x, so $|\mathrm{DLCT}_{S_i}(\alpha, \gamma)| = 32$) returns considerably more than Proposition 18's 9 pairs, and the surplus has an exact description. For every one of the eight S-boxes the set of directions admitting *some* such γ is

$$\mathbf{U}_i^{\mathrm{ls}} \;=\; \{0\} \cup \{\alpha : \exists\, \gamma \neq 0,\ \gamma \cdot D_\alpha S_i \equiv \mathrm{const}\},$$

which in every case is a 4-dimensional linear subspace of $\mathbb{F}_2^6$ containing $\mathbf{U}$, with 21 pairs in total split as 13 with $\gamma \cdot D_\alpha S_i \equiv 0$ and 8 with $\gamma \cdot D_\alpha S_i \equiv 1$. Its three nonzero cosets modulo $\mathbf{U}$ carry a fixed profile, identical for all eight boxes:

Coset of $\mathbf{U}$ in $\mathbf{U}_i^{\mathrm{ls}}$	$\dim \mathrm{span}\{D_\alpha S_i(x)\}$	γ's	Pairs
$\mathbf{U}^*$ itself (e.g. $\{4, 8, 12\}$)	2	3	9
one coset (e.g. $34 \oplus \mathbf{U}$)	3	1	4
two cosets (e.g. $18, 48 \oplus \mathbf{U}$)	4	1 ($\equiv 1$)	8

The 13 pairs with constant 0 are exactly the DLCT $= +32$ maximizers counted in Table 35; the remaining 8 reach -32 instead, being genuine linear structures in the sense of Definition 7 rather than confinements. Proposition 19 derives the whole profile. $\mathbf{U}_i^{\mathrm{ls}}$ is the subspace $\mathbf{N}_i$ of that proposition, the coset with 4 pairs is $d_i \oplus \mathbf{U}$, and the two cosets with 8 make up $\mathbf{N}_i \setminus \mathbf{N}_i^0$. The proposition gives $\mathbf{N}_i \subseteq \mathbf{U}_i^{\mathrm{ls}}$; the exhaustive screen gives equality, as no other component has a linear structure outside $\mathbf{U}$.

The position of these directions in the input explains why those outside $\mathbf{U}$ do no harm. In the bit names of Definition 19, $b_2b_3b_4b_5$ is the nibble and b_1, b_2, b_5, b_6 are the four bits the expansion shares with the neighbouring S-boxes. $\mathbf{U} = \mathrm{span}\{b_3, b_4\}$ is the pair of *middle* nibble bits, whereas the extra cosets are generated by directions such as $18 = b_2 \oplus b_5$ and $34 = b_1 \oplus b_5$, every one of which moves a boundary bit. By Lemma 4, a difference touching b_1, b_2, b_5 or b_6 perturbs an adjacent S-box, so none of these cosets satisfies the non-interference hypothesis and none propagates past the expansion. $\mathbf{U}$ is the part of $\mathbf{U}_i^{\mathrm{ls}}$ that avoids the boundary, which is why it, and only it, survives to the round-function level.

Remark 30 (The screen's discriminating power, measured). A detector is only as informative as its null distribution. We therefore ran the same screen against two controls, the eight genuine DES S-boxes and 3,000 random DES-style S-boxes, each with four rows that are independent random permutations of $\{0, \ldots, 15\}$. Table 17 reports the outcome.

The type-II row is zero everywhere, as Proposition 2 predicts. The linear-structure row separates the designs completely. Not one of the 3,000 random S-boxes has a single linear structure, while KG-DES has 21 in every S-box. This is what one expects. Since $\gamma \cdot D_\alpha S(x)$ takes the same value at x and $x \oplus \alpha$, a fixed pair (α, γ) faces 32 independent

Table 17: Screening KG-DES against two controls. Type-II bias is identically zero everywhere, as Proposition 2 requires. The linear-structure count separates KG-DES from both controls outright. Random baselines use 3,000 trials for the first two rows and 400 for the two uniformity rows, the latter being the more expensive to compute.

Statistic	KG-DES	DES	Random
$\max_{\beta\neq 0} \lvert\mathrm{LAT}(0,\beta)\rvert$ (RP type-II bias)	0 (all 8)	0 (all 8)	0 (all trials)
Linear structures per S-box	21 (all 8)	0, except 9 in S_4	0 in 3000/3000
Differential uniformity	24 (all 8)	16 (all 8)	24 in 1/400
Linear uniformity	16 (all 8)	14–20	16 in 57/400

constraints, and it is a linear structure of a random S-box only with probability about $2 \cdot 2^{-32} = 2^{-31}$.

The DES column is the more instructive control, because it is not clean. Its S_4 carries 9 structures on the direction set $\{1, 46, 47\}$, itself a 2-dimensional subspace. A screen that flagged only KG-DES would be suspect. One that also flags a known anomaly in real DES behaves like a measurement and not like an oracle. What separates the two cases is not the bare count but three qualitative features. KG-DES's structures appear in *all eight* boxes rather than one. Their confinements share a single subspace, since every $u \in \mathbf{U}^*$ has $D_u S_i$ in the same $\mathbf{W}_i$, and this is what makes a partition. DES S_4 also confines two directions to 2-dimensional subspaces, but to different ones, $D_{46}S_4$ to $\{0, 5, 10, 15\}$ and $D_{47}S_4$ to $\{0, 6, 9, 15\}$, while $D_1 S_4$ lies in the affine plane $\{5, 6, 9, 10\}$, so the three do not combine into a partition. And, decisively, KG-DES's core $\mathbf{U}$ is aligned with the expansion so as to survive it, whereas $\{1, 46, 47\}$, every element of which moves a boundary bit, is not. The last criterion is the one that generalizes. An anomaly of an S-box that the linear layer destroys is a curiosity, and one that the linear layer preserves is a trapdoor. Proposition 15 and Algorithm 3 test this, and this is why our detection pipeline runs the S-box screen first and the round-level closure second.

Remark 31 (The trapdoor is not free: differential uniformity betrays it). Table 17 records one cost of the construction, and Definition 17 should not be read as denying it. The cost is forced by the confinement. For $u \in \mathbf{U}^*$ the 64 input pairs distribute over the $|\mathbf{W}_i| = 4$ values of $\mathbf{W}_i$ alone, so by pigeonhole

$$\max_b \mathrm{DDT}_{S_i}(u, b) \;\geq\; \frac{2^6}{|\mathbf{W}_i|} \;=\; 16$$

for any dimension-2 confinement in a 6×4 S-box. Paterson's S-boxes go beyond this bound, to 24 in all eight, and he states this himself as one of their design properties (his property S7). The value compares with 16 for all eight genuine DES S-boxes, and is reached by only 1 in 400 random DES-style S-boxes. A plain differential-uniformity check therefore does separate KG-DES from DES, without any knowledge of the trapdoor. Paterson anticipates this. He notes that any trapdoor of this linear kind leaves a regularity in the XOR tables of small S-boxes, and that his S-boxes are incomplete, so that a close examination would reveal the trapdoor [Pat99, Section 6]. Our screen confirms his assessment. This is the differential half of Corollary 1. The spectral half, $\max_\lambda |\widehat{S}_i(\lambda, \gamma)| \geq 16$, holds with room to spare as well, since the linear uniformity 16 of Table 17 counts in LAT units and $\widehat{S} = 2\,\mathrm{LAT}$, so the maximum is 32. Neither bound is tight. Balance and linear uniformity, by contrast, are unremarkable for a single S-box (Table 17), although Section 6.8 shows that where the maximal linear entries sit, rather than how large they are, is what exposes the cipher to linear cryptanalysis. Consequently, the S-boxes pass as ordinary on those two statistics, but not on the differential spectrum.

6.5.1 A Type-II Relation of the Round Function

Balance rules out a Rijmen–Preneel relation in the S-boxes, but not in the round function, a possibility that Rijmen and Preneel themselves point out for ciphers with an expansion [RP97, §§3.1 and 4]. By Proposition 1, an expansion in front of the S-boxes gives $\theta_{\text{in}}^{\top}$ a kernel, of dimension 16 for DES's E (Table 47). The round function then acquires type-II entries from S-box approximations whose nonzero input masks cancel on the shared bits, even when every S-box is balanced. Their sign is $(-1)^{\mu \cdot K}$, a parity of key bits. For $A = \texttt{81000000}$ exactly one μ contributes, the masks $\texttt{03}$ on S_4 and $\texttt{30}$ on S_5 that read the two bits R_{16}, R_{17} both S-boxes receive, and

$$C_{A,0}^{F_K} \;=\; \pm C_{1,\texttt{03}}^{S_4}\, C_{8,\texttt{30}}^{S_5} \;=\; \pm 2^{-2},$$

with the sign given by four key bits. With the DES S-boxes the same trail gives $\pm 2^{-6}$. Both values hold for all 2^{12} values of the key bits entering S_4 and S_5. This entry is the round-function approximation behind the linear characteristic of Proposition 25.

The comparison with Rijmen and Preneel is therefore sharper than a screen of the S-boxes can state. KG-DES has no Rijmen–Preneel relation in its S-boxes, but its round function has one, of correlation 2^{-2} and key-dependent sign, assembled by DES's expansion from S-box approximations that the trapdoor places on the shared bits. Iterated as in Section 3.2, it gives a two-round characteristic whose key mask is a sum of round-key parities rather than 0.

6.6 Differential Characteristics

The fixed-round characteristics of Section 6.6.1 below are each built by repeatedly selecting, at every active S-box, the highest-probability entry of the trimmed DDT (Table 14). This greedy strategy is optimal at each individual step, but it does not in general return the state $(\Delta L, \Delta R)$ to its starting value. Continued past round 5, the difference spreads into additional nibbles and the cumulative probability collapses, as Section 6.6.7 quantifies.

An *iterative* characteristic, by contrast, returns exactly to its own input difference after a fixed number of rounds, and can therefore be self-concatenated indefinitely, in the manner of the classical Biham–Shamir characteristic for standard DES [BS93]. Finding one requires deliberately selecting the *non-default* branch of a tied DDT entry at specific rounds. We searched the input differences whose S-box inputs lie in $\mathbf{U} = \{0, 4, 8, 12\}$, restricted to one, two and three simultaneously active nibbles. For each candidate we explored all branches tied for the maximum DDT count at every round, and checked whether the resulting trajectory returns to its starting state. In this subsection we report the results of that search. By Theorem 2, every trajectory stays inside $\mathbf{S} \times \mathbf{S}$, so the search runs over the effective $16 + 16$-bit state space of the cipher rather than the full $32 + 32$-bit difference space. Even there the search is not exhaustive, and Sections 6.6.3 and 6.6.4 replace it with exhaustive searches over the full space.

6.6.1 The Fixed-Round Family

Before turning to iterative characteristics, we record the best *fixed*-round characteristics, contrasted with the iterative family by Section 6.6.7 below. We use the trimmed DDT of Table 14 and always select each round's highest-probability entry, 24/64, which every $u \in \mathbf{U}^*$ achieves against its two best output differences. In this way one trail with input differences confined to $\mathbf{U}$ is built for each round count $r \in \{2, 3, 4, 5\}$. Each activates exactly $r - 1$ S-boxes, one per round after the first, at probability 24/64 apiece. Table 18 gives the resulting probabilities, each $(24/64)^{r-1}$.

The specific 5-round trail referred to in Section 6.6.7 activates 5 S-boxes rather than the 4 of Table 18's uniform one-per-round pattern (an extra active S-box appears at

Table 18: Best fixed-round characteristics for input differences confined to **U**, built greedily by taking each round's highest-probability DDT entry at every active S-box. The exhaustive search of Section 6.6.4 shows that they are optimal among all characteristics (Table 23).

Characteristic	Active S-boxes	Probability
2-round	1	$24/64 = 2^{-1.415}$
3-round	2	$(24/64)^2 = 2^{-2.830}$
4-round	3	$(24/64)^3 = 2^{-4.245}$
5-round	4	$(24/64)^4 = 2^{-5.660}$

the transition into round 5), so it costs one further factor of 24/64 and achieves $2^{-7.075}$ instead of the table's $2^{-5.660}$. It is weaker, not stronger, since every additional active S-box multiplies in another 24/64. All per-round rates quoted in this paper use the same 24/64 per-S-box factor.

6.6.2 Best Iterative Characteristic

The strongest characteristic the hand search found is a period-6 cycle with input difference $\Delta R_0 =$ `0x02040000`, $\Delta L_0 = 0$. Its rate per round is the best of any period up to 6 (Section 6.6.3), although a longer period does better (Section 6.6.4). Table 19 gives the full round-by-round trail.

Table 19: 6-Round Iterative Characteristic (period 6)

Round	ΔL	ΔR	Active S-boxes (probability)
0	00000000	02040000	—
1	02040000	00004040	$S_2 : 4 \to 1$ (24/64), $S_4 : 8 \to 8$ (24/64)
2	00004040	00000000	$S_5 : 8 \to 4$ (24/64), $S_7 : 8 \to 1$ (24/64)
3	00000000	00004040	(free round, $p = 1$)
4	00004040	02040000	$S_5 : 8 \to 4$ (24/64), $S_7 : 8 \to 1$ (24/64)
5	02040000	00000000	$S_2 : 4 \to 1$ (24/64), $S_4 : 8 \to 8$ (24/64)
6	00000000	02040000	(free round, $p = 1$)

Round 6 reproduces the round 0 state exactly, so the characteristic concatenates with itself. Its probability is

$$Pr = \left(\frac{24}{64}\right)^2 \times \left(\frac{24}{64}\right)^2 \times \left(\frac{24}{64}\right)^2 \times \left(\frac{24}{64}\right)^2 = \left(\frac{24}{64}\right)^8 = 2^{-11.32} \tag{7}$$

per 6 rounds, i.e. an average rate of $2^{-1.887}$ per round.

Example 5 (Self-concatenation)**.** Applying the characteristic of Table 19 k times yields a $6k$-round characteristic of probability $\left(2^{-11.32}\right)^k$. For $k = 2$, a 12-round characteristic, $Pr = 2^{-22.64}$. For $k = 3$, an 18-round characteristic that exceeds one full application of KG-DES's 16 rounds, $Pr = 2^{-33.96}$. For comparison, the best iterative characteristic of standard (non-backdoored) DES achieves only $2^{-3.93}$ per round [BS93]. The trapdoor S-boxes give a characteristic just over two bits stronger per round ($3.93 - 1.887 \approx 2.04$), which remains useful well beyond the full round count of the cipher. The best rate any iterative characteristic of KG-DES attains is lower still, 1.415 bits per round, and the best 16-round characteristic has probability $2^{-21.23}$ (Proposition 22).

6.6.3 MILP-Certified Global Optimality

One may object that Table 19 was found by a manual, greedy search with explicit branching over tied optima (Section 6.6.7) and not by an exhaustive search, so it could be a strong but suboptimal local optimum. To settle this, we encoded the round function of KG-DES independently as an exact, bit-level mixed-integer linear program (MILP). The complete difference distribution table of each S-box is captured by one binary indicator variable per nonzero transition (a, b). The standard four-inequality XOR constraints link the Feistel halves, and a periodicity constraint $(\Delta L_r, \Delta R_r) = (\Delta L_0, \Delta R_0)$ closes the trail into a cycle of the requested period. The objective minimizes $\sum -\log_2(\mathrm{DDT}(a, b)/64)$ over the active transitions, i.e., it maximizes the probability. We solved the model to proven global optimality (the appendix lists the software).

We ran two searches. The first has no restriction at all, and $(\Delta L_r, \Delta R_r)$ ranges over the full $32 + 32$-bit difference space. The second confines every round's state to $\mathbf{S} \times \mathbf{S}$ (Theorem 2), the space the hand search above explored.

Table 20: Global-optimal periodic characteristics over the full unrestricted $32 + 32$-bit difference space: the MILP optimum, and the same quantity from the independent branch-and-bound search of Section 6.6.4, which also covers periods 5 and 6.

Period	MILP optimum (bits)	Exact search (bits)
1	8.3561	8.3561
2	8.3561	8.3561
3	9.6601	9.6601
4	11.9053	11.9053
5	—	12.4902
6	—	11.3203

Table 21: Global-optimal periodic characteristics restricted to $\mathbf{S} \times \mathbf{S}$, from the MILP and from the independent branch-and-bound search. "Infeasible": no closed trail of that period exists in $\mathbf{S}$.

Period	MILP optimum (bits)	Exact search (bits)
1	—	infeasible
2	infeasible	infeasible
3	9.6601	9.6601
4	11.9053	11.9053
5	—	12.4902
6	11.3203	11.3203

Three conclusions follow.

First, at periods 3 and 4 the two MILP searches return the same optimum, and the exact search of Section 6.6.4 finds the same at periods 5 and 6. Since every $\mathbf{S}$-restricted trail is also an unrestricted trail, the restricted feasible set is a subset of the unrestricted one, and the restricted optimum can therefore never be better than the unrestricted one. Equality means that the unrestricted optimum is *attained inside* $\mathbf{S}$. At these periods the best differential structure anywhere in the cipher's full $32 + 32$-bit difference space lies in the trapdoor-invariant subspace, so confining the search to $\mathbf{S}$ discarded nothing. This is not an obvious property of the construction, and it justifies, after the fact and for periods 3 to 6, the restriction that the hand search imposed from the start.

Second, at period 6, the case directly comparable to Table 19, the solver's branch-and-bound terminated with the bound closed rather than at the time limit, certifying $2^{-11.3203}$

as the global optimum *within* $\mathbf{S} \times \mathbf{S}$. This matches the hand-found value of $2^{-11.32}$ to the precision originally reported, and its per-round rate, $11.3203/6 = 1.887$ bits, reproduces the $2^{-1.887}$ of Equation (7). The period-6 characteristic of Table 19 is therefore not a strong greedy guess but a certified optimum. The *unrestricted* period-6 MILP was not completed, because the model size grows with the round count, and the period-4 instance took 155 of its 180 seconds. The branch-and-bound search of Section 6.6.4 settles it. The unrestricted period-6 optimum is also $2^{-11.3203}$, so no period-6 trail outside $\mathbf{S}$ does better. That search also reproduces every MILP entry of Tables 20 and 21, with an implementation that shares nothing with the MILP model. Period-6 optimality does not make Table 19 the best iterative characteristic overall, since a period-12 characteristic has a lower cost per round (Proposition 22).

Third, periods 1 and 2 are provably infeasible within $\mathbf{S}$. This does not conflict with their unrestricted optima in Table 20. A period-1 trail needs $\Delta L_0 = \Delta R_0$ and a vanishing round-function output difference, and a period-2 trail needs the output difference to vanish in both rounds. Inside $\mathbf{S}$ neither can happen, because for $u \in \mathbf{U}^*$ no KG-DES S-box has output difference 0 ($\mathrm{DDT}_{S_i}(u, 0) = 0$ for all eight S-boxes), so any active S-box produces a nonzero output difference. The unrestricted optima lie outside $\mathbf{S}$, and the restriction removes them. Only an exhaustive search can give a negative result of this kind. A greedy extension cannot.

However, the solver's optimal period-6 trail is not the same explicit trail as Table 19. It activates different S-boxes at the same total weight, so the global optimum is attained more than once. We checked this trail separately from the solver, by simulating the round function (expansion, DDT-based substitution and permutation).

Table 22: MILP-certified alternate optimal 6-round characteristic, $Pr = (24/64)^8 = 2^{-11.32}$, matching Table 19's probability exactly with a different explicit trail.

Round	ΔL	ΔR	Active S-boxes (probability)
0	00000000	00000404	—
1	00000404	00200020	$S_6 : 8 \to 2$ (24/64), $S_8 : 8 \to 4$ (24/64)
2	00200020	00000000	$S_3 : 4 \to 2$ (24/64), $S_7 : 4 \to 2$ (24/64)
3	00000000	00200020	(free round, $p = 1$)
4	00200020	00000404	$S_3 : 4 \to 2$ (24/64), $S_7 : 4 \to 2$ (24/64)
5	00000404	00000000	$S_6 : 8 \to 2$ (24/64), $S_8 : 8 \to 4$ (24/64)
6	00000000	00000404	(free round, $p = 1$)

6.6.4 Optimal Characteristics for Every Round Count

The MILP of Section 6.6.3 certifies the best *periodic* characteristic of a given period. An attack on a fixed number of rounds needs a different quantity, the best r-round characteristic B_r with no periodicity required, and the periodic optima themselves deserve a check that shares no code with the MILP. Both come from Matsui's branch-and-bound search [Mat95]. Write a_i for the difference entering the round function in round i and b_i for the difference leaving it. Then $a_{i+1} = a_{i-1} \oplus b_i$ with $a_0 = \Delta L_0$, so a_1 and a_2 are free and every later a_i is forced. The probability of round i is the product of $\mathrm{DDT}_{S_j}(e_j, o_j)/64$ over the 6-bit blocks e_j of $E(a_i)$ and the output nibbles o_j with $b_i = P(o)$. Rounds are pruned against the already known optima for fewer rounds. For a periodic search the last two output differences are forced by the cycle condition, and because every rotation of a cycle is again a cycle of the same weight, round 1 may be taken to be a cheapest round. As a check, on the DES S-boxes the search returns Biham and Shamir's iterative characteristic `19600000` with probability 1/234 as the best of period 2, and their 13-round characteristic of probability $2^{-47.22}$ as B_{13} [BS93].

Table 23: Optimal weight $B_r = -\log_2 p$ of an r-round differential characteristic of KG-DES and of DES, by exhaustive branch-and-bound search. Every KG-DES entry is a multiple of $\log_2(64/24) = 1.415$, since each optimal characteristic uses only the 24/64 entries of the trapdoor directions.

r	1	2	3	4	5	6	7	8
KG-DES	0	1.42	2.83	4.25	5.66	8.49	8.49	9.91
DES	0	2	4	9.61	13.22	19.96	23.61	30.48
r	9	10	11	12	13	14	15	16
KG-DES	11.32	12.74	14.15	16.98	16.98	18.40	19.81	21.23
DES	31.48	38.35	39.35	46.22	47.22	54.10	55.10	61.97

Proposition 22 (Optimal differential characteristics of KG-DES)**.** *The following hold for the S-boxes of KG-DES.*

1. *The best characteristic over the full* 16 *rounds has probability* $2^{-21.23}$*, i.e.,* $(24/64)^{15}$*. It is attained by the characteristic of Table 24, all of whose differences lie in* $\mathbf{S}$*.*

2. *Every characteristic of KG-DES over* $6k$ *rounds, for any* $k \geq 1$*, has probability at most* $2^{-1.415}$ *per round on average; hence so does every iterative characteristic, whatever its period. A period-*12 *characteristic attains this rate. From* $\Delta L = 0$ *and* $\Delta R =$ 00000004 *it returns to the same difference after* 12 *rounds with* 12 *active S-boxes, with probability* $2^{-16.98}$*. (Characteristics over other round counts can average less than* 1.415 *bits per round, because their first round can be free, as in* $B_{16}/16 = 1.33$*.)*

3. *The best periodic characteristics of periods* 1 *to* 6 *have weights* 8.356*,* 8.356*,* 9.660*,* 11.905*,* 12.490 *and* 11.320*, with or without the restriction to* $\mathbf{S} \times \mathbf{S}$*, except that periods* 1 *and* 2 *have no characteristic inside* $\mathbf{S} \times \mathbf{S}$*.*

Proof. Parts 1 and 3 are the results of the exhaustive search. For part 2, split any characteristic over $6k$ rounds into k blocks of six rounds. Each block is a 6-round characteristic, so the total weight is at least $k\,B_6 = 8.49\,k$, which is 1.415 bits per round. An iterative characteristic of period p, concatenated 6 times, is a characteristic over $6p$ rounds, so the bound applies to it. The period-12 characteristic is rounds 3 to 14 of Table 24, which begin and end in the same state. □

Three consequences follow. First, the fixed-round characteristics of Table 18 are optimal, with $B_r = (r-1) \times 1.415$ for $r = 2, \ldots, 5$. Second, the period-6 characteristic of Table 19 is optimal among periods up to 6, now without the restriction to $\mathbf{S}$ that the MILP needed at period 6, but it is not the best iterative characteristic. The period-12 cycle gains almost half a bit per round, and over 16 rounds the gap is seven bits ($2^{-21.23}$ against about $2^{-28.3}$ for the best concatenation of Table 19). Third, the distance to DES is far larger than a per-round comparison suggests, at $2^{-21.23}$ against $2^{-61.97}$ over the full cipher. As a check on the probability model, 2^{20} random pairs through 6 rounds of KG-DES with independent round keys followed the input and output differences of the optimal 6-round characteristic 2868 times, a rate of $2^{-8.51}$ against the predicted $2^{-8.49}$.

Paterson [Pat99, Section 4.2] bounds the characteristics of his 32-round cipher that have an active S-box in every round by 2^{-41} over 29 rounds, and reports characteristics close to that bound. Interestingly, characteristics with inactive rounds do better. The same exhaustive search gives $B_{29} = 39.62$ bits, and over his full 32 rounds $B_{32} = 43.87$ bits, with 31 active S-boxes.

Table 24: An optimal 16-round differential characteristic of KG-DES: 15 active S-boxes, each with DDT = 24/64, probability $2^{-21.23}$. Rounds 3–14 form the period-12 cycle of Proposition 22.

Round r	ΔL_r	ΔR_r	Active S-boxes in round r
0	00020020	00000004	—
1	00000004	00000000	$S_8 : 8 \to 6$
2	00000000	00000004	(free round)
3	00000004	00000020	$S_8 : 8 \to 4$
4	00000020	00000404	$S_7 : 4 \to 2$
5	00000404	00200000	$S_6 : 8 \to 2$, $S_8 : 8 \to 4$
6	00200000	00000400	$S_3 : 4 \to 2$
7	00000400	00000000	$S_6 : 8 \to 2$
8	00000000	00000400	(free round)
9	00000400	00200000	$S_6 : 8 \to 2$
10	00200000	00000404	$S_3 : 4 \to 2$
11	00000404	00000020	$S_6 : 8 \to 2$, $S_8 : 8 \to 4$
12	00000020	00000004	$S_7 : 4 \to 2$
13	00000004	00000000	$S_8 : 8 \to 4$
14	00000000	00000004	(free round)
15	00000004	00000020	$S_8 : 8 \to 4$
16	00000020	00000404	$S_7 : 4 \to 2$

6.6.5 Comparison of Iterative Characteristics

Table 25 summarizes every iterative characteristic found by the hand search (restricted to ≤ 3 simultaneously active nibbles at round 0), ranked by overall probability, together with the period-12 characteristic of Proposition 22. The latter has the best rate of all.

Table 25: Iterative characteristics of KG-DES ($\Delta L_0 = 0$). Ranks 1–4 are those of the hand search, ranked by overall probability; the last row is the period-12 characteristic of Table 24 (rounds 3–14), whose rate $2^{-1.415}$ per round is the best any iterative characteristic can reach (Proposition 22).

Rank	Period	Probability	Rate/round	ΔR_0
1 (best)	6	$2^{-11.32}$	$2^{-1.887}$	0x02040000
2	6	$2^{-14.15}$	$2^{-2.358}$	0x06040000
3	8	$2^{-16.98}$	$2^{-2.123}$	0x02020000
4	6	$2^{-16.98}$	$2^{-2.830}$	0x22020000
best rate	12	$2^{-16.98}$	$2^{-1.415}$	0x00000004

Ranks 3 and 4 have the same overall probability ($2^{-16.98}$) but different periods. Rank 3 spreads the same total cost over 8 rounds instead of 6, so its rate per round is better ($2^{-2.123}$ against $2^{-2.830}$), although it ranks lower by total probability. Depending on whether an attack targets a fixed round count or seeks the best asymptotic rate, rank 3 may be preferable to rank 4.

6.6.6 Additional Characteristics

For completeness, Tables 26–28 give the full round-by-round trails for ranks 2–4.

Remark 32 (Verification). We checked every transition in Tables 19, 22 and 26–28 by simulating the round function (expansion, DDT-based S-box substitution and permutation), separately from the search. In each case the listed probability is the exact product of the

Table 26: 6-Round Iterative Characteristic (rank 2), $Pr = (24/64)^{10} = 2^{-14.15}$

Round	ΔL	ΔR	Active S-boxes (probability)
0	00000000	06040000	—
1	06040000	00404040	$S_2 : c \to 1$ (24/64), $S_4 : 8 \to a$ (24/64)
2	00404040	00000000	$S_3 : 8 \to 1$, $S_5 : 8 \to 4$, $S_7 : 8 \to 1$ (all 24/64)
3	00000000	00404040	(free round, $p = 1$)
4	00404040	06040000	$S_3 : 8 \to 1$, $S_5 : 8 \to 4$, $S_7 : 8 \to 1$ (all 24/64)
5	06040000	00000000	$S_2 : c \to 1$ (24/64), $S_4 : 8 \to a$ (24/64)
6	00000000	06040000	(free round, $p = 1$)

Table 27: 8-Round Iterative Characteristic (rank 3), $Pr = (24/64)^{12} = 2^{-16.98}$

Round	ΔL	ΔR	Active S-boxes (probability)
0	00000000	02020000	—
1	02020000	00404000	$S_2 : 4 \to 1$ (24/64), $S_4 : 4 \to 2$ (24/64)
2	00404000	06060000	$S_3 : 8 \to 1$ (24/64), $S_5 : 8 \to 4$ (24/64)
3	06060000	00000000	$S_2 : c \to 1$ (24/64), $S_4 : c \to 2$ (24/64)
4	00000000	06060000	(free round, $p = 1$)
5	06060000	00404000	$S_2 : c \to 1$ (24/64), $S_4 : c \to 2$ (24/64)
6	00404000	02020000	$S_3 : 8 \to 1$ (24/64), $S_5 : 8 \to 4$ (24/64)
7	02020000	00000000	$S_2 : 4 \to 1$ (24/64), $S_4 : 4 \to 2$ (24/64)
8	00000000	02020000	(free round, $p = 1$)

DDT entries at the active S-boxes, as the search reports. For Table 22 the simulation is independent of the solver that produced the trail, so the MILP objective and the simulation arrive at the same 11.3203-bit figure by two separate routes. We simulated all five tables, and Table 24, once more with a second implementation of the round function, and every transition and every probability agrees.

Table 28: 6-Round Iterative Characteristic (rank 4), $Pr = (24/64)^{12} = 2^{-16.98}$

Round	ΔL	ΔR	Active S-boxes (probability)
0	00000000	22020000	—
1	22020000	00004042	$S_1 : 4 \to 1$, $S_2 : 4 \to 1$, $S_4 : 4 \to 8$ (all 24/64)
2	00004042	00000000	$S_5 : 8 \to 1$, $S_7 : 8 \to 1$, $S_8 : 4 \to 2$ (all 24/64)
3	00000000	00004042	(free round, $p = 1$)
4	00004042	22020000	$S_5 : 8 \to 1$, $S_7 : 8 \to 1$, $S_8 : 4 \to 2$ (all 24/64)
5	22020000	00000000	$S_1 : 4 \to 1$, $S_2 : 4 \to 1$, $S_4 : 4 \to 8$ (all 24/64)
6	00000000	22020000	(free round, $p = 1$)

6.6.7 Why the Fixed-Round Family Is Not Iterative

The fixed-round characteristics of Section 6.6.1 do *not* belong to the iterative family above, despite sharing the same trapdoor subspace $\mathbf{U}$ and $\mathbf{W}$. Continue a greedy trail past round 5 under the rule used to build it, taking the first maximal DDT entry at every active S-box. From each of the 24 single-nibble starting differences $\Delta L_0 \in \mathbf{S}$ with $\Delta R_0 = 0$, the first five rounds cost $2^{-7.075}$ (the five-S-box variant described after Table 18), but what follows depends on the start. From 20 of the starts the number of active S-boxes grows, usually to three in a single round by round 6 and to five or more later, and the weight after 20 rounds lies between 39.6 and 82.1 bits. From the other four the extension falls into a period-12 cycle at the optimal rate of Proposition 22 and reaches only 26.9 bits. A greedy extension can therefore land on the optimal iterative characteristic, but only because of its starting point. It cannot tell that it has done so, and it cannot rule out anything better. The iterative characteristics above were found by branching over *both* tied optima at each active S-box and searching for trajectories that close back onto their own input difference. The two families are therefore best presented as separate results. Section 6.6.1 establishes short, fixed-round characteristics optimized for a specific round count, while Section 6.6.2 establishes a self-concatenating characteristic suitable for attacking arbitrarily many rounds. Section 6.6.4 supersedes both with the optimal characteristic for every round count.

6.7 From the Characteristic to a Key Recovery

The characteristics of Section 6.6 are distinguishers. Here we turn the optimal one into a last-round key recovery. It shows that the trapdoor decides which key bits the attack yields. The differential recovers the coset-label bits of the last-round subkey and leaves the in-coset bits untouched.

The attack is the standard one-round differential technique [BS93], applied to the round function of Definition 18. We fix the optimal 16-round characteristic of Table 24. Its last active round is S_7 with the transition $4 \to 2$, and its first 15 rounds form a characteristic of probability $2^{-19.81}$ (the weight B_{15} of Table 23) from the plaintext difference to $(\Delta L_{15}, \Delta R_{15})$. We request N chosen-plaintext pairs with that plaintext difference. For a right pair the state after 15 rounds follows the characteristic. The input difference to the round function in round 16 is then ΔR_{15}, and $\Delta L_{16} = \Delta R_{15}$ is known, so we discard every pair whose ciphertext halves do not differ by ΔR_{15} in the left half. For each surviving pair the round-16 output difference is

$$\Delta F_{16} \;=\; \Delta R_{16} \oplus \Delta L_{15},$$

where ΔR_{16} is read from the ciphertext and ΔL_{15} from the characteristic. Furthermore, we discard the pairs for which $P^{-1}(\Delta F_{16})$ activates an S-box other than S_7. What remains, for each pair, is the input pair of S_7 (obtained from the ciphertext through E) and the

required output difference $P^{-1}(\Delta F_{16})$ restricted to S_7. We now guess the six subkey bits k that S_7 receives in round 16 and count the pairs consistent with that output difference. Every right pair counts the correct subkey, and a wrong subkey is counted only by chance.

Proposition 23 (The recovery reaches the coset bits, not the whole subkey)**.** *Let the last active round of the characteristic be S-box S_i with input difference $u \in \mathbf{U} \setminus \{0\}$. The counter of the 1R attack takes the same value on the four subkeys k, $k \oplus 4$, $k \oplus 8$, $k \oplus 12$, and on no other coset of $\mathbf{U}$ does it match. The attack therefore recovers the four coset-label bits of the round-16 subkey at S_i, and the two bits selecting the representative within the $\mathbf{U}$-coset are invisible to it.*

Proof. For a pair with S_i-input difference u, the two S-box inputs under a guess k are $x \oplus k$ and $x \oplus u \oplus k$, and their output difference is $D_u S_i(x \oplus k)$. Since $u \in \mathbf{U}$, the derivative $D_u S_i$ is constant on each coset of $\mathbf{U}$ (Lemma 3 and the affine coset maps of Section 6.3.5), so $D_u S_i(y)$ depends only on $y \oplus \mathbf{U}$. Replacing k by $k \oplus u'$ for $u' \in \mathbf{U}$ shifts every input by $u' \in \mathbf{U}$, leaving its $\mathbf{U}$-coset and hence $D_u S_i$ unchanged, so the counter is identical for all four such k. For a guess k'' outside the coset, $\delta := (k'' \oplus k) \bmod \mathbf{U}$ is nonzero, and the two guesses read $D_u S_i$ on $\mathbf{U}$-cosets that differ by δ. Here one property of the tables is needed. For every i and every $u \in \mathbf{U}^*$ the period group of $D_u S_i$ modulo $\mathbf{U}$ is trivial, in other words the only shift under which $D_u S_i$ is invariant is a member of $\mathbf{U}$ itself. A direct check of the eight S-box tables confirms this. Consequently $D_u S_i$ takes different values on the two cosets for at least one input, so k'' misses every right pair whose S-box input falls in that coset and its counter is strictly below the true one. The recovery therefore fixes the four coset-label bits and no more. □

The two hidden bits sit where $\mathbf{U}$ sits in the S-box input, the middle bits of each 6-bit block (Lemma 4). The four recovered bits are the outer bits, the S-box's coset label. We ran the attack on the full 16-round cipher with independent random round keys and 2^{25} chosen-plaintext pairs. About 40 right pairs survive the two filters, and in each of eight trials the largest counter lies in the true $\mathbf{U}$-coset, as Proposition 23 predicts. The data cost follows the estimate $2^{B_{15}+5}$, which yields about thirty right pairs, and reduced-round runs track B_{r-1} in the same way (Table 29).

Table 29: Last-round differential key recovery on KG-DES. The attack recovers the four coset-label bits of the last-round subkey at the characteristic's final S-box; the run reports the fraction of trials in which the true $\mathbf{U}$-coset is the maximal counter.

Rounds	Last S-box	B_{r-1}	Chosen pairs	Coset recovered
7	S_6	8.49	2^{13}	40/40
10	S_3	11.32	2^{16}	20/20
13	S_8	16.98	2^{22}	12/12
16	S_7	19.81	2^{25}	8/8

A single characteristic reaches only one S-box. To recover the whole last-round subkey we need, for each of the eight S-boxes, a characteristic that ends on it alone. A branch-and-bound search over trapdoor trails finds one for every S-box, all of weight 15 over 16 rounds, so each S-box costs about 2^{26} chosen pairs. Together the eight recover all 32 outer bits of the round-16 subkey and none of the 16 in-coset bits. This is the split that the key schedule itself shows. Of the 48 bits of K_{16}, 32 sit in outer positions and 16 in the middle, $\mathbf{U}$-carrying positions (Section 6.3). The remaining bits are not a differential problem, because the $\mathbf{U}$-bits are hidden by construction. Splitting the key in this way, first the bits that decide which coset is reached and then the bits inside the coset, is Paterson's own divide-and-conquer attack [Pat99, §§3.2 and 5]. He reaches the first part by exhaustive search over a reduced key space, which needs a deliberately weakened key

schedule. The differential reaches it without one. The trapdoor therefore does more than open a distinguisher. It also draws the line between the key material that differential cryptanalysis exposes and the material it cannot reach.

The two halves combine into a full key recovery. Under the DES key schedule, the 32 recovered outer bits of the last-round subkey are 32 bits of the master key. The other 24 are the 16 **U**-bits of the subkey and the 8 master bits that the schedule leaves out of that round, and we find them by exhaustive search against a single known block. With the eight weight-15 characteristics the recovery runs at the full sixteen rounds. We recover the eight cosets from about 2^{26} chosen plaintext pairs each, the residual 2^{24} search returns the key, and the reconstructed 56-bit master key matches the one we set. On six rounds, where the characteristics are lighter, the same program needs only about 2^{14} pairs per S-box. (The 8 parity bits of the 64-bit key lie outside the 56-bit key that PC1 keeps and never affect encryption, so there is nothing to recover there.)

The key schedule lets us shrink the residual. Each master bit appears in several round keys, and every one of the 56 occupies an outer position in some round key. The outer positions of K_{16} and K_{15} together already pin 49 of them. Recovering those of K_{15} needs a second stage that peels round 16 first, using $L_{15} = R_{16} \oplus F(L_{16}, K_{16})$. This seems to require all of K_{16}, hidden **U**-bits included. It does not, because the same trapdoor that hides those bits also makes them irrelevant to the peeling.

Proposition 24 (Peeling the last round without its hidden bits)**.** *Let K'_{16} agree with K_{16} on its* 32 *outer bits, with arbitrary* **U***-bits. Take a pair of texts whose round-*16 *S-box input differences all lie in* **U***. Computing L_{15} with K'_{16} in place of K_{16} gives the correct difference ΔL_{15}, and it moves the input of every round-*15 *S-box, in both texts, by the same element of* **U***.*

Proof. For text j the two computations of L_{15} differ by $\varepsilon_j := F(L_{16}^j, K_{16}) \oplus F(L_{16}^j, K'_{16})$. In S-box t the two keys differ by some $v_t \in \mathbf{U}$, so its contribution to ε_j is $D_{v_t} S_t(x_t^j)$, where x_t^j is the true S-box input. Every coset map of KG-DES is affine (proof of Proposition 21), so $D_{v_t} S_t$ is constant on each coset of **U**. By hypothesis x_t^1 and x_t^2 differ by an element of **U**, so they lie in the same coset, and $\varepsilon_1 = \varepsilon_2$. Hence ΔL_{15} is unchanged. Each contribution lies in $\mathbf{W}_t$, and P followed by E carries $\bigoplus_t \mathbf{W}_t$ into **U** at every S-box of the next round (Theorem 2 and Lemma 4). Hence ε_1 shifts every round-15 S-box input by an element of **U**, the same for both texts. □

A right pair of a characteristic that stays in the trapdoor block satisfies the hypothesis at round 16, and it can be recognized from the ciphertexts. By Proposition 23 the counter of the round-15 attack determines only the coset of the subkey, and a common shift by an element of **U** moves each candidate within its coset. The second stage therefore recovers the outer bits of K_{15} from the recovered outer bits of K_{16} alone. We ran both stages at the full sixteen rounds. With a characteristic of weight 14 ending on each S-box in round 15, about $2^{24.8}$ further chosen pairs per S-box recover all eight cosets of K_{15}. Together with K_{16} they pin 49 master bits, and a search over the remaining 2^7 returns the full 56-bit key. The whole run takes under eight minutes. On 6, 8 and 10 rounds the two stages pin 45 bits, and all fifteen runs with random keys recovered the key. The same argument peels round after round, so a third stage would pin the last seven bits, but a search over 2^7 keys is negligible. The in-coset bits stay out of reach of the differential in every round. The key schedule, which places them at outer positions in other rounds, is what brings them back.

6.8 Linear Characteristics

Linear cryptanalysis [Mat94b] looks for an approximation $\lambda_{\text{in}} \cdot P \oplus \lambda_{\text{out}} \cdot C = \kappa \cdot K$ whose correlation c is far from 0, and needs on the order of c^{-2} known plaintexts to use it. A linear characteristic (Definition 13) is only an estimate of that correlation. For a fixed key

the approximation sums over its whole linear hull, so the estimate has to be confirmed by measurement. KG-DES has no linear characteristic of correlation 1, since no S-box has a linear approximation of correlation ± 1, inside or outside the trapdoor subspaces.

In round i write X_i for the mask on the input R_{i-1} of the round function and Y_i for the mask on its output. Summing the round approximations, the intermediate halves cancel exactly when $X_i = Y_{i-1} \oplus Y_{i+1}$ for $1 < i < n$, and each round costs the product of its S-box correlations, where $X_i = E^\top \alpha_i$ and $Y_i = P\beta_i$ for S-box masks $\alpha_i \to \beta_i$. Matsui's branch-and-bound search [Mat95] returns the optimal weight B_r^{L} of an r-round characteristic. On the DES S-boxes our search reproduces Matsui's published values, among them the bias 1.49×2^{-24} at 16 rounds. With the KG-DES S-boxes the same search gives the first row of Table 30.

Table 30: Optimal weight $B_r^{\mathrm{L}} = -\log_2 |c|$ of an r-round linear characteristic, found by Matsui's branch-and-bound search. The DES row is Matsui's [Mat95] and serves as a check on the search.

r	1	2	3	4	5	6	7	8
KG-DES	1	1	2	3	4	6	6	8
DES	0.68	0.68	1.36	3.03	4.71	7.03	8.03	9.71
r	9	10	11	12	13	14	15	16
KG-DES	8	10	10	12	12	14	14	16
DES	12.07	13.39	14.07	15.75	17.42	19.75	20.75	22.42

Proposition 25 (An iterative linear characteristic of KG-DES)**.** *Let $A = \texttt{81000000} = P\beta$, where β is 1 on the output of S_4, 8 on the output of S_5 and 0 elsewhere. For every round key K, the function $R \mapsto A \cdot F(R, K)$ has correlation exactly $\pm 2^{-2}$ with the constant 0. Consequently the characteristic that leaves the odd rounds inactive and uses, in each even round, S_4 with $\alpha = \texttt{03} \to \beta = 1$ and S_5 with $\alpha = \texttt{30} \to \beta = 8$ has $X_i = 0$ and $Y_i = A$ in every active round. Over $2k$ rounds it is a characteristic for $A \cdot R_0 \oplus A \cdot R_{2k}$ of correlation $\pm 2^{-2k}$ for every key. For $2k = 16$ this is $\pm 2^{-16}$, which equals B_{16}^{L}, so the characteristic is optimal.*

Proof. The mask $\alpha = \texttt{03}$ on S_4 reads its last two input bits and $\alpha = \texttt{30}$ on S_5 its first two. The expansion E feeds the same two bits of R to both positions, so the two masks cancel in $E^\top$ and $X_i = 0$. Both output masks annihilate the trapdoor subspaces, $1 \in \mathbf{W}_4^\perp$ and $8 \in \mathbf{W}_5^\perp$, so by Proposition 3 each component $\beta_j \cdot S_j$ ignores the two middle input bits. Hence $A \cdot F(R, K)$ depends on K only through the 12 key bits entering S_4 and S_5, and on R only through the six outer input bits of the two S-boxes. Computing over all 2^{12} values of those key bits gives correlation $+2^{-2}$ for half of them and -2^{-2} for the other half. Since $X_i = 0$, each active round contributes $A \cdot R_i = A \cdot R_{i-2} \oplus A \cdot F(R_{i-1}, K_i)$, and the stated correlation over $2k$ rounds is the product of the k round correlations. Optimality is the search result $B_{16}^{\mathrm{L}} = 16$. □

The trapdoor is what makes this possible, and it does so in a way that a check of the S-boxes one at a time cannot see. Every KG-DES S-box has maximal linear correlation 2^{-1}, below the 40/64 that S_5 of DES reaches. What differs is where the maximal entries sit. All eight of them, in each of the eight S-boxes, lie in the trapdoor block, with input mask in $\mathbf{U}^\perp$ and output mask in $\mathbf{W}_i^\perp$. By Proposition 3 such an approximation reads only the outer input bits, and those are exactly the bits E shares with the neighbouring S-box. Two neighbours can therefore meet on their shared bits and cancel, leaving a round-function approximation with zero input mask and correlation $(2^{-1})^2$. This is a type-II relation of the round function in Matsui's sense, and its sign depends on the key

(Proposition 1). With the DES S-boxes the same masks give $\pm 2^{-6}$ per round. Figure 4 draws the characteristic.

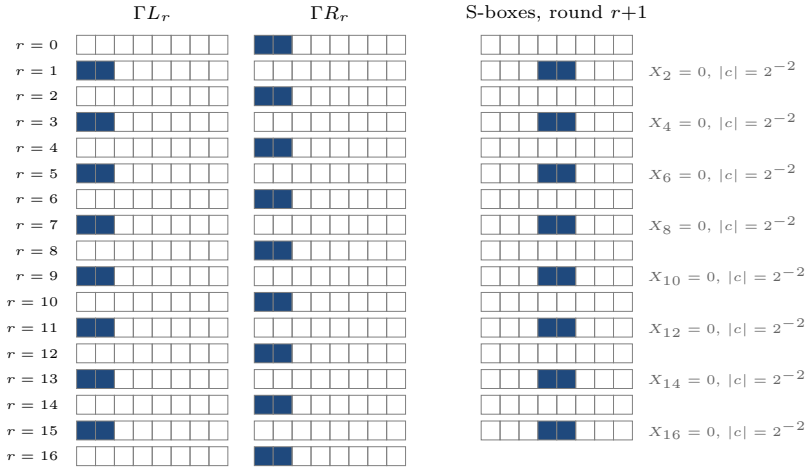

Figure 4: Optimal 16-round linear characteristic of KG-DES, one row per round: the nibble activity of the masks $(\Gamma L_r, \Gamma R_r)$ on the state after r rounds, and the S-boxes active in round $r + 1$. Blue: nonzero mask. The state mask alternates between $(0, A)$ and $(A, 0)$ with $A = \texttt{81000000}$. Each even round activates S_4 and S_5, whose input masks cancel under $E^\top$, so the round function is approximated with input mask $X = 0$ at correlation $\pm 2^{-2}$. The characteristic has correlation $\pm 2^{-16}$ for every key.

Table 31 measures the approximation on the cipher. For a fixed key its correlation is a sum over the linear hull, and other characteristics with the same outer masks add to this one or cancel part of it. Averaged over keys the characteristic predicts the correlation well. Over 64 keys at 10 rounds and 32 keys at 12 rounds, the mean of c^2 is $2^{-19.81}$ and $2^{-23.69}$, against 2^{-20} and 2^{-24} for the characteristic alone, and the median of $|c| \cdot 2^r$ is 1.01 in both cases. Individual keys scatter by a factor of about two either way. The value $|c| \cdot 2^r$ ranges from 0.46 to 2.24 at 10 rounds and from 0.44 to 1.79 at 12. Of the two keys measured over the full 16 rounds, one gives $2^{-16.35}$ and the other $2^{-18.35}$, within one standard deviation of zero at 2^{36} plaintexts, so for that key the hull cancels most of the characteristic. On the DES round function the same masks give $2^{-6.00}$ at two rounds and $2^{-12.06}$ at four, six bits per active round instead of two. The practical reading is that full 16-round KG-DES is distinguished from a random permutation, for a typical key, with on the order of 2^{32} known plaintexts, where the best DES characteristic ($2^{-22.42}$) would need about 2^{45}. The distinguisher uses nothing but the published S-box tables. It needs neither chosen plaintexts, as the invariant-subspace distinguisher of Corollary 3 does, nor knowledge of the partition. With knowledge of the partition, about 2^{17} known plaintexts suffice (Section 6.10).

Table 31: Measured correlation of $A \cdot R_0 \oplus A \cdot R_r$ on r-round KG-DES with independent random round keys, against the characteristic of Proposition 25. The standard deviation of each estimate is about $2^{-\ell/2}$ for 2^ℓ plaintexts; the spread between keys comes from the linear hull. The DES rows use the same masks with the DES S-boxes.

Cipher	Rounds	Plaintexts	Keys	Measured c per key	Characteristic
KG-DES	2	2^{30}	2	$-2^{-2.00}$, $-2^{-2.00}$	2^{-2}
KG-DES	4	2^{30}	2	$-2^{-3.97}$, $-2^{-4.12}$	2^{-4}
KG-DES	6	2^{30}	2	$-2^{-5.78}$, $+2^{-5.89}$	2^{-6}
KG-DES	8	2^{30}	2	$-2^{-8.02}$, $+2^{-8.29}$	2^{-8}
KG-DES	10	2^{30}	2	$+2^{-10.01}$, $-2^{-10.39}$	2^{-10}
KG-DES	12	2^{32}	3	$-2^{-12.22}$, $+2^{-11.78}$, $-2^{-12.63}$	2^{-12}
KG-DES	14	2^{34}	2	$+2^{-14.71}$, $-2^{-13.39}$	2^{-14}
KG-DES	16	2^{36}	2	$+2^{-16.35}$, $+2^{-18.35}$	2^{-16}
DES S-boxes (control)	2	2^{30}	1	$+2^{-6.00}$	–
DES S-boxes (control)	4	2^{30}	1	$-2^{-12.06}$	–

6.9 Differential-Linear Trails and Boomerangs

The two hybrid techniques combine a differential with a second property, a linear approximation or a second differential. In KG-DES both combinations reach their maximum, and for the same reason. The partition holds with certainty.

6.9.1 A Correlation-1 Differential-Linear Trail

Differential-linear cryptanalysis [LH94] joins a differential over the first part of a cipher to a linear approximation over the rest. For an N-bit cipher E, an input difference Δ and an output mask λ, its figure of merit is the differential-linear correlation $c_E(\Delta, \lambda)$ of Definition 11, and a distinguisher built on it needs on the order of c_E^{-2} pairs. Bar-On et al. [BODKW19] split the cipher as $E = E_1 \circ E_m \circ E_0$, with E_m a single S-box layer, and estimate $c_E \approx p\, r\, q^2$. Here p is the probability of the differential through E_0, r the correlation read from the DLCT of E_m (Definition 11, normalized by 2^{m-1}), and q the correlation of the linear approximation through E_1. In KG-DES all three factors can be made exactly 1.

Proposition 26 (Correlation-1 differential-linear trails of KG-DES)**.** *Let* $\mathbf{S}^\perp \subset \mathbb{F}_2^{32}$ *be the annihilator of* $\mathbf{S}$*, spanned by bit positions* 1 *and* 4 *of each nibble. For every key schedule, every number of rounds, every nonzero* $\Delta \in \mathbf{S} \times \mathbf{S}$ *and every nonzero* $\lambda \in \mathbf{S}^\perp \times \mathbf{S}^\perp$*,* $c_E(\Delta, \lambda) = 1$*. No differential-linear distinguisher on KG-DES has larger absolute correlation, for any number of rounds.*

Proof. By Theorem 2, $E(x) \oplus E(x \oplus \Delta) \in \mathbf{S} \times \mathbf{S}$ for every x and every key, so $\lambda \cdot (E(x) \oplus E(x \oplus \Delta)) = 0$ and every term of the sum is $+1$. Optimality is the bound $|c_E| \le 1$. □

The proof does not need the DLCT, but the three-part split shows where the correlation sits in the standard framework. Table 32 gives one explicit trail. The mask λ on R_{16} reaches the round-16 S-box outputs as $P^\top \lambda$, whose only nonzero component is $\gamma = 1$ on S_4. This γ annihilates $\mathbf{W}_4 = \{0, 2, 8, 10\}$, so Proposition 18 gives the maximal DLCT entry whatever input difference in $\mathbf{U}$ reaches S_4.

Figure 5 draws the trail round by round, in the convention of Hadipour, Derbez and Eichlseder [HDE24]. A red upper-left triangle marks a cell that the upper, differential part may activate, and a blue lower-right triangle marks a cell carrying the lower, linear mask. One difference from their figures matters. Their upper parts are truncated differentials in which an active cell may hold any nonzero value, and the probability is paid in the number of active S-boxes. Here every red cell is confined to a trapdoor subspace, which is why a fully active state still costs nothing. From $r = 5$ on every cell is red, and the probability is

Table 32: An explicit correlation-1 differential-linear trail of KG-DES, in the $E_1 \circ E_m \circ E_0$ decomposition of [BODKW19]. Any nonzero Δ and λ from the subspaces of Proposition 26 give the same correlation, with $2^{32} - 1$ choices of each.

Δ	$(\Delta L_0, \Delta R_0) = (\texttt{00000000}, \texttt{02040000})$
λ	$\texttt{80000000}$ on R_{16}, 0 on L_{16}
E_0, p	rounds 1–15 and the expansion of round 16; difference lands in $\bigoplus \mathbf{U}$ with $p = 1$
E_m, r	round-16 S-boxes, $\gamma = 1$ on S_4; DLCT $= 32 = 2^5$ for every $u \in \mathbf{U}^*$, $r = 1$
E_1, q	P and the Feistel XOR with L_{15}; λ annihilates $\mathbf{S} \times \mathbf{S}$, $q = 1$
c_E	1

still 1. The differences a single encryption actually produces fill in these patterns, with occasional cancellations to zero, and all of them stay inside the subspaces.

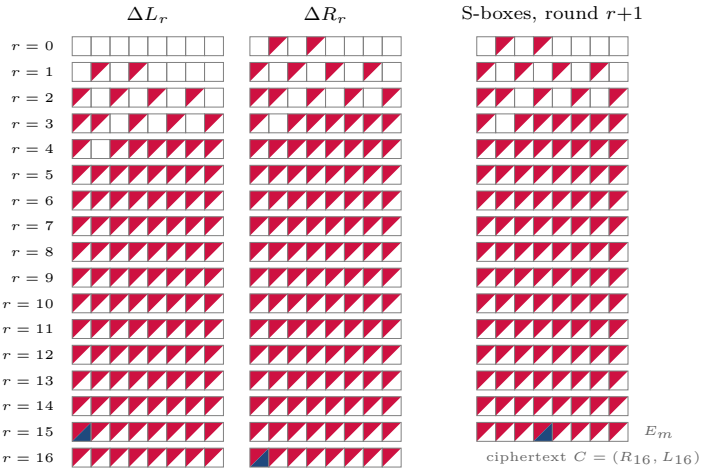

Figure 5: Correlation-1 differential-linear trail through the 16-round KG-DES, one row per round: the nibble activity of $(\Delta L_r, \Delta R_r)$ after r rounds, and the S-boxes active in round $r + 1$. Red: nibble difference possibly nonzero, and then always in $\{0, 2, 4, 6\}$; red S-boxes receive an input difference in $\mathbf{U}^*$. Blue: the mask $\lambda = \texttt{80000000}$ on R_{16}, which reaches L_{15} through the Feistel XOR and the round-16 S-box S_4 as $\gamma = 1$ through $P^\top$; that S-box, marked E_m, carries both colours. Round keys, IP and the final swap are not drawn. From $r = 5$ on, every cell is red. RU = 15, RM = 1, RL = 0; $p = r = q = 1$.

We verified the trail in two independent ways. First, we checked every layer exhaustively, including the expansion and P on bases of the relevant subspaces, each S-box on all of its inputs and all differences in $\mathbf{U}$, and the DLCT entries of E_m in full. Together these cover every key, so they amount to a proof by computation of the trail in Table 32. Second, we ran the full cipher with random keys, next to a control of identical structure without the trapdoor, i.e., the genuine DES S-boxes in the KG-DES round function (Table 33). The control returns correlations indistinguishable from 0, as a random cipher would, so the measurement discriminates. KG-DES returns 1 on every pair.

Remark 33 (What the trail adds, and what it does not). Proposition 26 is the invariant-subspace distinguisher of Corollary 3 written in differential-linear form, and it gives an attacker no capability beyond that distinguisher. It is a chosen-plaintext attack. With a single mask bit, each further pair halves the probability that a random permutation passes. Testing every mask in $\mathbf{S}^\perp \times \mathbf{S}^\perp$ at once means testing membership of the output difference in $\mathbf{S} \times \mathbf{S}$, which a random permutation passes with probability 2^{-32}. Whether any pair

Table 33: Full-cipher measurement of the trail of Table 32, over 20 random keys per row, with independent random round keys and without IP. The 32-round rows run the same trail through twice the specified number of rounds and show that the correlation does not decay with the round count. The control correlations are within about two standard deviations ($\approx 1/\sqrt{\text{pairs}}$) of 0.

Cipher	Pairs	$c_E(\Delta, \lambda)$	Difference in $\mathbf{S} \times \mathbf{S}$
KG-DES, 16 rounds	10,000	+1.0000	10,000
DES S-boxes, 16 rounds (control)	10,000	+0.017	0
KG-DES, 32 rounds	10,000	+1.0000	10,000
DES S-boxes, 32 rounds (control)	10,000	+0.020	0

(Δ, λ) outside these subspaces also reaches correlation 1 we have not established, since an exhaustive search over 64-bit differences is out of reach. The open differential-linear question concerns standard DES, where the best trail must be found by search and its correlation is far below 1.

6.9.2 A Probability-1 Boomerang

A boomerang quartet starts from a plaintext pair $P_2 = P_1 \oplus \alpha$, shifts both ciphertexts by δ, $C_3 = C_1 \oplus \delta$ and $C_4 = C_2 \oplus \delta$, and decrypts. It returns when $P_3 \oplus P_4 = \alpha$, which for a random permutation of N-bit blocks happens with probability about 2^{-N}. The standard estimate of the return probability [CHP+18] splits the cipher into a top part with a differential characteristic of probability p, a switch, and a bottom part with a characteristic of probability q:

$$\Pr[\text{return}] \;\approx\; p^2\, q^2 \prod \frac{\mathrm{BCT}(\Delta, \nabla)}{2^n},$$

the product running over the S-boxes of the switch where both trails are active. An S-box active in only one trail costs nothing, since $\mathrm{BCT}(\Delta, 0) = \mathrm{BCT}(0, \nabla) = 2^n$ (the *ladder* switch). The estimate follows a single characteristic in each half, and for both KG-DES and BEA-1 it turns out to be far too pessimistic. We therefore compute or measure the return probabilities below directly.

The affine coset maps behind Proposition 21 extend from one S-box to the whole cipher, and they make the boomerang exact.

Proposition 27 (KG-DES is affine on the cosets of $\mathbf{S} \times \mathbf{S}$)**.** *For every key K, every $x \in \mathbb{F}_2^{64}$ and every $s \in \mathbf{S} \times \mathbf{S}$,*

$$E_K(x \oplus s) \;=\; E_K(x) \oplus M_{K,[x]}(s),$$

where $M_{K,[x]}$ is a linear bijection of $\mathbf{S} \times \mathbf{S}$ that depends only on K and on the coset $[x] = x \oplus (\mathbf{S} \times \mathbf{S})$.

Proof. By the proof of Proposition 21, each S_i restricted to a coset $y \oplus \mathbf{U}$ is an affine bijection onto a coset of $\mathbf{W}_i$. In other words, $S_i(y \oplus u) = S_i(y) \oplus \lambda_{i,y}(u)$, with $\lambda_{i,y} : \mathbf{U} \to \mathbf{W}_i$ linear and depending only on the coset of y. For $s \in \mathbf{S}$, every 6-bit block of $E(s)$ lies in $\mathbf{U}$ (Lemma 4), so each S-box input moves within its coset of $\mathbf{U}$, and

$$F(R \oplus s) \;=\; F(R) \oplus \Lambda_{[R]}(s), \qquad \Lambda_{[R]}(s) = P\Big(\bigoplus\nolimits_i \lambda_i\big(E(s)_i\big) \Big),$$

with $\Lambda_{[R]}$ linear from $\mathbf{S}$ to $\mathbf{S}$ (Lemma 5) and depending only on the coset of R and the round key. One round therefore maps $(L \oplus s_1, R \oplus s_2)$ to its image of (L, R) shifted by

$(s_2,\ s_1 \oplus \Lambda_{[R]}(s_2))$, an invertible linear function of (s_1, s_2). The coset of the state after each round depends only on the coset of the input (Theorem 2), so the linear parts compose round by round into $M_{K,[x]}$. □

Corollary 4 (Probability-1 boomerang and second-order differentials)**.** *For every key and all $\alpha, \delta \in \mathbf{S} \times \mathbf{S}$, the boomerang quartet on full 16-round KG-DES returns with probability 1, and every second-order derivative $D_s D_t E_K$ with $s, t \in \mathbf{S} \times \mathbf{S}$ vanishes identically.*

Proof. P_1 and P_2 lie in one coset, so $C_2 = C_1 \oplus M(\alpha)$ with $M = M_{K,[P_1]}$, and C_1, C_2, C_3, C_4 lie in one coset of the image. Decryption on that coset is the inverse affine map, so $P_3 = P_1 \oplus M^{-1}(\delta)$ and $P_4 = P_2 \oplus M^{-1}(\delta)$, and $P_3 \oplus P_4 = \alpha$. For the second statement, $E_K(x) \oplus E_K(x \oplus s) \oplus E_K(x \oplus t) \oplus E_K(x \oplus s \oplus t) = M(s) \oplus M(t) \oplus M(s \oplus t) = 0$. □

On the full cipher with independent random round keys, 4,000 of 4,000 quartets with random $\alpha, \delta \in \mathbf{S} \times \mathbf{S}$ returned. With the DES S-boxes in the same round function, none did. One quartet thus distinguishes KG-DES, in the adaptive chosen-plaintext-and-ciphertext setting, where a random permutation returns with probability 2^{-64}. The characteristic estimate misses this completely. Even with a free switch, the best split of Table 23 ($B_{r_0} + B_{r_1} = 18.40$ over $r_0 + r_1 = 15$ rounds) predicts about $2^{-36.8}$, because it follows one sequence of differences through each half, whereas an affine map returns every quartet of the coset at once.

6.10 The Partition in the Correlation Matrix

Both halves of the Fourier condition of Proposition 4 hold for all eight KG-DES S-boxes with their pairs $(\mathbf{U}, \mathbf{W}_i)$, and both fail for every DES S-box tested against the same subspaces. For these non-bijective S-boxes the two halves are independent conditions, so both had to be checked. Figure 6 shows the zero pattern for S_1 once the masks of $\mathbf{U}^\perp$ and $\mathbf{W}_1^\perp$ are listed first.

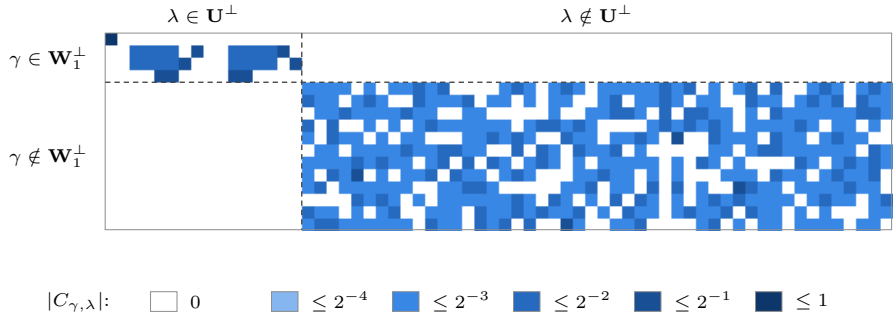

Figure 6: The correlation matrix of KG-DES S_1, all 16×64 entries (rows: output masks γ; columns: input masks λ), with the masks of $\mathbf{W}_1^\perp$ and $\mathbf{U}^\perp$ listed first; white is exactly 0. Both off-diagonal blocks vanish, the pattern of Proposition 4, and the top-left 4×16 block holds the trapdoor components. $C_{0,0} = 1$, and row and column 0 are otherwise 0 because the S-box is balanced.

At the level of the cipher, Theorem 2 says that E_K maps every coset of $\mathbf{S} \times \mathbf{S}$ into a coset of $\mathbf{S} \times \mathbf{S}$ for every key, and the final swap preserves $\mathbf{S} \times \mathbf{S}$. Theorem 1 therefore applies with $\mathbf{T} = \mathbf{T}' = \mathbf{S} \times \mathbf{S}$, $N = 64$ and $d = 32$: for every key, the block of C^{E_K} on $\mathbf{S}^\perp \times \mathbf{S}^\perp$ is the correlation matrix of a permutation of the 2^{32} cosets, an exact 32-bit cipher hidden inside KG-DES. Every subspace of the corresponding mask trail is spanned by single bits. $\mathbf{S}^\perp$ is spanned by nibble positions 1 and 4, $\mathbf{U}^\perp$ by the four outer input bits and $\mathbf{W}_i^\perp$ by two output bits. The trail can therefore be drawn bit by bit through the real

E and P (Figure 7). In mask terms, $P^\top$ carries $\mathbf{S}^\perp$ onto the output bits of $\bigoplus_i \mathbf{W}_i^\perp$, the S-boxes connect those only to input masks in $\bigoplus \mathbf{U}^\perp$ (Proposition 4), and $E^\top$ carries these back into $\mathbf{S}^\perp$. The Feistel state therefore keeps both halves of its mask in $\mathbf{S}^\perp$.

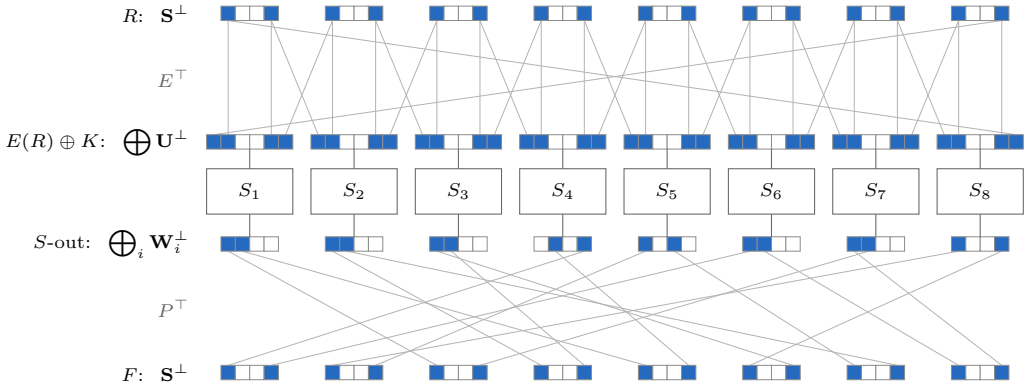

Figure 7: The mask subspace trail of KG-DES through the round function, the geometric content of Theorem 1. Filled bits may carry mask, white bits carry none. The wires are the actual E and P for the filled bits, and every one lands on a filled bit.

By Corollary 2, every nonzero input mask of the block has an output mask with $|C^{E_K}_{v,u}| \geq (2^{32}-1)^{-1/2} > 2^{-16}$. The characteristic of Proposition 25 has its masks in $\mathbf{S}^\perp \times \mathbf{S}^\perp$. Over the full 16 rounds its correlation, 2^{-16}, is the root mean square of a column of the block, so at that length it is typical of the block, not exceptional. The two 16-round measurements of Table 31, $2^{-16.35}$ and $2^{-18.35}$, are of that order. Over fewer rounds the characteristic stands far above this level.

The known-plaintext distinguisher of Section 4.4 needs about $2^{16.5}$ texts for one expected pair in a common coset of $\mathbf{S} \times \mathbf{S}$. At $2^{17.5}$ texts about four pairs are expected, and at least one occurs with probability $1 - e^{-4} \approx 0.98$. A random permutation puts the ciphertexts of such a pair in one coset with probability only 2^{-32}. Table 34 confirms the count on the full cipher. The distinguisher lowers the known-plaintext data of the linear characteristic, about 2^{32}, to about 2^{17}, at the price of knowing the partition.

Table 34: The known-plaintext partition distinguisher on the full KG-DES, with independent random round keys. A pair is two known plaintexts in one coset of $\mathbf{S} \times \mathbf{S}$; the expected number is $N(N-1)/2$ divided by the number of cosets, 128 per key. The last column counts pairs whose ciphertexts lie in one coset of $\mathbf{S} \times \mathbf{S}$. The control uses the same plaintexts, subspaces and keys.

Cipher	Texts per key	Keys	Pairs	Ciphertexts in one coset
KG-DES, 16 rounds	2^{20}	4	501	501
DES S-boxes (control)	2^{20}	4	501	0

6.11 What the Analysis of KG-DES Shows

We have reported three kinds of result, and it helps to keep them apart. The literature supplies the construction and its purpose. From Paterson it also supplies the invariant subspace of the whole cipher, as a block system that holds for every key, and the basic chosen-plaintext attack on it. In addition, it supplies the differential uniformity 24 of the S-boxes and the observation that the trapdoor is detectable [Pat99]. A companion

study [FAI+23] supplies the maximal DLCT and FBCT entries of S_1. Run blind on the tables alone, the algorithm reproduces the trapdoor pair that Paterson designed in (Section 6.2). Theorem 2 restates his block system in the form of differences, and Corollary 3 draws the immediate two-plaintext distinguisher from it. Everything else is new. On the differential side this includes the quadratic trapdoor component, the second, weaker near-trapdoor direction (Proposition 19) and the proof that Paterson's criteria force them (Proposition 20). It also includes the optimal differential characteristic for every round count, including Paterson's 32 rounds, the optimal iterative rate (Proposition 22, Remark 26), and the two-stage key recovery with its residual search over 2^7 possibilities (Section 6.7). On the linear side it includes the linear characteristic of correlation 2^{-16} and the type-II relation of the round function behind it (Proposition 25, Section 6.5.1). Finally there is the probability-1 boomerang (Corollary 4) and the perfect 32-bit approximation with its known-plaintext distinguisher (Section 6.10).

The partition turns out to be more than a chosen-plaintext trapdoor for its holder. DES's linear layer has branch number 2 and the cosets have only four elements, so the coset maps are affine. As a result the trapdoor also opens point differentials, linear characteristics and boomerangs that anyone holding the published tables can use. In the terms of Table 6, the backdoor causes a structural weakness. Two limitations apply. First, the differential of Section 6.7 recovers only the coset-label bits of each round subkey it reaches, and the in-coset bits stay out of its reach. The key schedule brings most of them back through other rounds, and a search over 2^7 possibilities completes the master key. The other attacks in this section are distinguishers and not key recoveries. Second, the linear distinguisher's data varies with the key through the linear hull (Table 31), so its 2^{32} is a typical value and not a bound. Table 35 records how we checked each result of this section and of Section 5.

Table 35: Independent computational verification of the results of Sections 5 and 6; every claim was checked by direct computation, separately from its proof. The corresponding summaries for BEA-1 and for the trails of both ciphers are Tables 51 and 52.

Result	Verification method	Outcome
$\mathbf{W}_i$ sets (Lemma 3)	Exhaustive enumeration, 1536 table lookups	Exact match, 0 exceptions
Non-interference under E (Lemma 4)	Enumeration of all 48 output bits of E against all 8 nibble positions	Confirmed
Closure under P (Lemma 5)	Enumeration of all 16 free output bits through the P-table	Confirmed, 0 violations
Full-cipher invariant (Theorem 2)	20,000 random trials, full 16 rounds, independently random round keys per trial	0 failures
Reachability (Remark 28)	3,000 trials, 16 rounds, union of two-nibble-window differences	Exact match to predicted 16-element subspace
Algorithm 3 output	Recomputed independently of the by-hand derivation of Section 6.3.1	Identical 16-bit result
DLCT uniformity (Proposition 18)	Exhaustive search, all 63×15 nonzero (Δ, γ) pairs, all 8 S-boxes	Exact at 32/32 for every S_i, at 13 pairs: the 9 predicted and the 4 of Proposition 19
FBCT uniformity (Proposition 21)	Exhaustive, all ordered pairs of distinct nonzero differences, all 8 S-boxes; injectivity on all 16 cosets of $\mathbf{U}$	64/64 exactly at the 6 trapdoor pairs, nowhere else
Quadratic trapdoor component (Proposition 19)	Algebraic normal forms of the three trapdoor components; linear structures, kernel and DLCT column of the quadratic one; the row maps and $\operatorname{rank} A_i$	One quadratic component per box; $\dim \mathbf{N}_i = 4$, $\dim \mathbf{N}_i^0 = 3$, $\operatorname{rank} A_i = 1$, all 8
Rijmen–Preneel screen (Proposition 2)	Zero-input-mask row of the LAT, linear structures, uniformities; DES and 3,000 random controls (Table 17)	Type-II bias 0 everywhere; 21 linear structures per box, 0 in controls except DES S_4
Spectral detector (Proposition 3)	Blind run on all 12 trapdoored and 12 control S-boxes, 2,000 random (Table 9)	Correct $(\mathbf{U}, \mathbf{W})$ on all 12; no false trapdoor

7 BEA-1: A Partition Trapdoor in an SPN

KG-DES leaves two questions open. Its trapdoor is small, a 2-dimensional subspace in 6×4 S-boxes whose coset maps are automatically affine, and it sits in DES's selection-based linear layer of branch number 2. Both facts made the analysis of Section 6 easier and the trapdoor more useful, so we cannot tell how much of what we found depends on them. BEA-1, the second published instance of the partition family, removes both. Its 10-bit S-boxes are bijections carrying 5-dimensional trapdoor subspaces, and its MixColumns is a general linear map of maximal branch number. It is therefore the demanding test. Does the method of Section 5 carry over to a cipher of a different shape, and does a partition trapdoor still open ordinary trails when the linear layer diffuses well? We follow the order of the previous section, with the construction and its known analysis first, then the recovery of the trapdoor with Algorithm 1 and its propagation through the round function, and then the trails of each type.

7.1 The Cipher and Its Known Analysis

Bannier and Filiol [BF17] designed BEA-1, and Bannier's PhD thesis [Ban17] develops it in full. Their purpose was to extend the partition backdoors of Paterson and of Harpes to substitution-permutation networks. BEA-1 is an SPN whose design the authors describe as resembling AES [BF17]. It has an 80-bit block, a 120-bit master key and 11 rounds, each an AddRoundKey/Substitution/Diffusion composition (Section 7.3 gives the diffusion layer in full). The four 10-bit S-boxes of the substitution layer carry the same kind

of partition-preserving structure as those of KG-DES, in a cipher of a different shape. What makes BEA-1 a backdoor, and not an accidental structural weakness, is the way its specification handles disclosure. It publishes each S-box *twice*. The first is the real, secret table, which carries the trapdoor. The second is a *modified*, publicly deployable table, altered at roughly 8–10% of its entries so that the public version shows no detectable structure. Below, the *public*, *modified* and *deployed* S-boxes all mean these tables. Bannier accordingly classes BEA-1 as a *probabilistic* partition-based backdoor cipher [Ban17, Ch. 5]. The exact partition belongs to the secret S-boxes, and our analysis of it below uses those. An evaluator with only the public specification sees an ordinary-looking SPN. The designer, who holds the secret tables, keeps a probability-1 distinguisher of the kind that Section 6 established for KG-DES.

Definition 20 (BEA-1 Round Function and Encryption)**.** Write a plaintext block as $P = (p_0, \dots, p_7) \in (\mathbb{F}_2^{10})^8$ (eight 10-bit bundles). Define

$$\mathrm{SubBundles}(x_0, \dots, x_7) := \big(S_0(x_0), S_1(x_1), S_2(x_2), S_3(x_3), S_0(x_4), S_1(x_5), S_2(x_6), S_3(x_7)\big),$$

applying $S_{i \bmod 4}$ to bundle i; $\mathrm{ShiftRows}(x_0, \dots, x_7) := (x_{\sigma(0)}, \dots, x_{\sigma(7)})$ for the fixed bundle permutation $\sigma = (0, 5, 2, 7, 4, 1, 6, 3)$ (Lemma 6); and MixColumns applies the 40×40 $\mathbb{F}_2$-linear map M independently to bundles $\{0, 1, 2, 3\}$ and $\{4, 5, 6, 7\}$, each group read as one element of $\mathbb{F}_2^{40}$. Writing $X_0 := P$ and $k_0, \dots, k_{11} \in (\mathbb{F}_2^{10})^8$ for the round keys (derived from the 120-bit master key by BEA-1's own key schedule [Ban17]), encryption proceeds

$$X_{i+1} := \mathrm{MixColumns}\big(\mathrm{ShiftRows}(\mathrm{SubBundles}(X_i \oplus k_i))\big), \qquad i = 0, \dots, 9,$$

followed by one final, MixColumns-free round and a closing key whitening,

$$X'_{10} := \mathrm{ShiftRows}\big(\mathrm{SubBundles}(X_{10} \oplus k_{10})\big), \qquad C := X'_{10} \oplus k_{11}.$$

This gives 11 substitution layers in total, the "11 rounds" stated above, driven by 12 round keys $k_0, \dots, k_{11}$. Every result in Section 7.3 concerns this iteration at the level of XOR differences. For differences, AddRoundKey (each step $\oplus\, k_i$) cancels whatever the key, as in the proof of Theorem 3. The particular nonlinear construction of the key schedule therefore plays no role in any of our claims.

The designers' own analysis is the reference point for what follows. Bannier proves that, with the secret S-boxes, the round function preserves the partition $\mathcal{L}(V)$ and the last round maps $\mathcal{L}(V)$ to $\mathcal{L}(W)$, whatever the round keys [Ban17, Theorem 5.2]. For the deployed cipher he shows that the public and secret ciphers encrypt a given plaintext identically, in all eleven rounds, with probability about 2^{-11} [Ban17, Theorem 5.4]. He turns this into a statistical key-recovery attack ([Ban17, Algorithm 8]), and we confirmed that it recovers a full BEA-1 key from 2^{16} chosen plaintexts. Everything else in this section is ours, i.e., the blind recovery of the subspaces, the mechanical derivation of the invariant subspace, and every trail and distinguisher.

7.2 Recovering the Trapdoor from the S-Box Tables

We ran Algorithm 1 on both versions of each of the four 10-bit S-boxes, the secret one and the public one (Table 36).

Every one of the four secret S-boxes yields a 32-element subspace $\mathbf{V}_i$ with a consistent companion $\mathbf{W}_i$ of the same size. Check (Algorithm 2) verifies the pair across *every* one of its 32 cosets (all 1024 shifts), and not only on the coset containing 0. This is the same phenomenon as in Table 13, found again, independently, in a cipher that we did not construct. In contrast, the same algorithm run on the four modified (deployed) S-boxes

Table 36: Result of running Algorithm 1 against BEA-1's secret S-boxes (each recovering a dimension-5, all-cosets-verified trapdoor subspace, independently reconfirmed by FINALIZE's duality cross-check on π^{-1}) versus its publicly deployed, modified counterparts (each recovering only the trivial subspace).

S-box	$\dim(\mathbf{V}_i)$	$\lvert\mathbf{V}_i\rvert$	Verified against CHECK	FINALIZE duality cross-check
Secret S_0	5	32	all 32 cosets	S_i^{-1} has the pair $(\mathbf{W}_i, \mathbf{V}_i)$
Secret S_1	5	32	all 32 cosets	S_i^{-1} has the pair $(\mathbf{W}_i, \mathbf{V}_i)$
Secret S_2	5	32	all 32 cosets	S_i^{-1} has the pair $(\mathbf{W}_i, \mathbf{V}_i)$
Secret S_3	5	32	all 32 cosets	S_i^{-1} has the pair $(\mathbf{W}_i, \mathbf{V}_i)$
Modified S_0–S_3 (deployed)	0	1	trivial only	n/a (no non-trivial pair)

finds nothing beyond the trivial subspace $\{0\}$. This matches the specification's own claim that the deployed cipher resists standard differential and linear cryptanalysis, while the secret tables carry the actual trapdoor.

Remark 34 (The original criterion also succeeds here). Perhaps surprisingly, our *original, uncorrected* candidate criterion (the row-sharing test described in Section 5.1) also succeeds on BEA-1's secret S-boxes. Its count is 32 for each of the four S-boxes, not the 1024 one might expect for a 10-bit S-box. In each case exactly the 31 trapdoor directions satisfy the criterion, and the coset search of the original algorithm then recovers the correct $\mathbf{W}$ as well. BEA-1 is therefore not a counterexample to the earlier criterion, either at the candidate step or in the coset search.

The version that does fail is an intermediate one. It already used the corrected candidate step, the difference-span test at the top of Algorithm 1, but it still trusted the closure of the greedy build as sufficient, without CHECK. That version passed the BEA-1 test above, but it failed on the KG-DES S-boxes, whose correct pairs were already known. On each of the eight S-boxes of Table 13 it took in the second maximizing coset $d_i \oplus \mathbf{U}$ of Proposition 19 and returned the 8-element set $\mathbf{N}_i^0$. This set is closed under $\oplus$ but fails the definition. On six of the S-boxes the coset containing 0 already collapses to four images. On S_2 and S_5, where that coset keeps eight distinct images, four of the other seven cosets collapse. This failure is why we made CHECK a mandatory final test instead of trusting closure, and why Algorithm 1 checks its own output before returning it, whichever candidate step feeds it.

7.3 From the S-Boxes to the Cipher

Remark 20 extends the closure of Section 5.5 to any linear diffusion layer. In this subsection we carry it out for BEA-1. We derive the invariant subspace of the round function from the S-box subspaces by the same mechanical procedure as for KG-DES, instead of taking it from Bannier's theorem or sampling it.

7.3.1 A Structural Difference from KG-DES

BEA-1's round function decomposes as

$$R \xrightarrow{\text{SubBundles}} (w_0, \ldots, w_7) \xrightarrow{\text{ShiftRows}} (w'_0, \ldots, w'_7) \xrightarrow{\text{MixColumns}} F,$$

where $R \in \mathbb{F}_2^{80}$ is the round input (eight 10-bit bundles $r_0, \ldots, r_7$), bundle i feeds S-box $S_{i \bmod 4}$ directly, and MixColumns acts independently on the two groups $\{0, 1, 2, 3\}$ and $\{4, 5, 6, 7\}$ via the same 40×40 $\mathbb{F}_2$-matrix M. Three differences from the KG-DES setting of Section 5.5 follow immediately:

- There is no expansion step before the S-boxes, so pr_i (the projection onto S-box i's input) is simply bundle i of R, the identity selection. Proposition 15 therefore gives a *trivial* preimage at this layer, and Algorithm 3 has nothing to do.
- The linear layer sits entirely *after* the S-boxes (as with DES's P), but unlike P it is not a selection map, since MixColumns computes $\mathbb{F}_2$-linear combinations of its 40 input bits. This is the case anticipated by Remark 20, not the closed-form case of Algorithm 3.
- BEA-1 has no Feistel split. The entire 80-bit state is transformed every round, so the inductive argument of Theorem 2 needs no $\Delta L/\Delta R$ bookkeeping, with one round, one state and one invariant.

Proposition 28 (Round-Level Candidate Subspace for BEA-1)**.** *Let $\mathbf{V}_0, \ldots, \mathbf{V}_3 \subset \mathbb{F}_2^{10}$ be the S-box-level subspaces recovered in Section 7.2 (Table 36), and define, in the notation of Proposition 15, $\mathbf{T}_0 = \bigcap_{i=0}^{7} \mathrm{pr}_i^{-1}(\mathbf{V}_{i \bmod 4})$, where pr_i is the projection of R onto bundle i. Then*

$$\mathbf{T}_0 \;=\; \bigoplus_{i=0}^{7} \mathbf{V}_{i \bmod 4}$$

(each $\mathbf{V}_{i \bmod 4}$ embedded in its own 10-bit window of $\mathbb{F}_2^{80}$), a 40-dimensional linear subspace.

Proof. Since pr_i reads bundle i's 10 bits and ignores the other 70, $\mathrm{pr}_i^{-1}(\mathbf{V}_{i \bmod 4})$ is exactly $\mathbf{V}_{i \bmod 4}$ embedded at bundle i, direct-summed with all of $\mathbb{F}_2^{10}$ at every other bundle. Intersecting these eight constraints over $i = 0, \ldots, 7$ leaves each bundle constrained by exactly one $\mathbf{V}_{i \bmod 4}$, giving the stated direct sum. Its dimension is $5 \times 8 = 40$, since each $\mathbf{V}_j$ has dimension 5 (Table 36). □

This $\mathbf{T}_0$ is precisely the subspace $\mathbf{V}$ of BEA-1's own specification [Ban17, §5.2.1]. Proposition 28 re-derives it mechanically from the S-box data alone, just as Example 4 re-derived KG-DES's $\mathbf{S}$ from $\mathbf{U}$ and the E-table.

7.3.2 Closure under ShiftRows: A Selection-Map Step

Lemma 6 (ShiftRows Closure)**.** *BEA-1's bundle permutation is $\sigma = (0, 5, 2, 7, 4, 1, 6, 3)$ (bundle j of the output equals bundle $\sigma(j)$ of the input). Then $\sigma(i) \equiv i \pmod 4$ for every $i \in \{0, \ldots, 7\}$, and hence* ShiftRows *maps $\mathcal{L}(\mathbf{W})$ to itself, where $\mathbf{W} = \bigoplus_{i=0}^{7} \mathbf{W}_{i \bmod 4}$.*

Proof. We use direct table lookup, as for KG-DES's P (Table 16). The permutation $\sigma = (0, 5, 2, 7, 4, 1, 6, 3)$ gives $\sigma(0) = 0, \sigma(1) = 5, \sigma(2) = 2, \sigma(3) = 7, \sigma(4) = 4, \sigma(5) = 1, \sigma(6) = 6, \sigma(7) = 3$. Reducing each value mod 4 reproduces $(0, 1, 2, 3, 0, 1, 2, 3)$, so $\sigma(i) \bmod 4 = i \bmod 4$ in every case. ShiftRows is a coordinate permutation, i.e., a selection map, as E and P are for KG-DES. A value in $\mathbf{W}_{\sigma(i) \bmod 4}$ at bundle $\sigma(i)$ before the permutation is therefore a value in $\mathbf{W}_{i \bmod 4}$ at bundle i after it, so the subspace class is the same and only relabelled. □

7.3.3 Closure under MixColumns: The General Case

Proposition 29 (MixColumns Closure)**.** *Let M be BEA-1's 40×40 $\mathbb{F}_2$-MixColumns matrix and let $\mathbf{W}_{04} = \mathbf{W}_0 \oplus \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \mathbf{W}_3 \subset \mathbb{F}_2^{40}$ (one 4-bundle group, each $\mathbf{W}_j$ embedded at its own 10-bit window). Then*

$$M(\mathbf{W}_{04}) \;=\; \mathbf{V}_0 \oplus \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \mathbf{V}_3 \;=:\; \mathbf{V}_{04}.$$

Proof. As Remark 20 anticipates, M is a general linear map, not a selection map, so containment cannot be read off a table as in Lemma 6. We verify it by linear algebra instead. $\mathbf{W}_{04}$ has a basis of 20 vectors (the 5 local basis vectors of each $\mathbf{W}_j$, $j = 0, \ldots, 3$, embedded at bundle position j). Since M and membership-in-$\mathbf{V}_{04}$ are both $\mathbb{F}_2$-linear, $M(\mathbf{W}_{04}) \subseteq \mathbf{V}_{04}$ if and only if $M(b) \in \mathbf{V}_{04}$ for every one of these 20 basis vectors b, and there is no need to check all 2^{20} elements. A computation confirms $M(b) \in \mathbf{V}_{04}$ for all 20. Since M is invertible (MixColumns must be, for the cipher to be decryptable) and $\dim \mathbf{W}_{04} = \dim \mathbf{V}_{04} = 20$, the image $M(\mathbf{W}_{04})$ has dimension exactly 20, so containment forces equality. □

Table 37: A representative sample (one basis vector per bundle position; the full computation checks all 20) of the MixColumns-closure check of Proposition 29, verifying $M(\mathbf{W}_{04}) \subseteq \mathbf{V}_{04}$ exactly, in the manner of Table 16 but via linear algebra instead of table lookup, as required by a non-selection linear layer.

Local basis of $\mathbf{W}_j$	Bundle position j	M(embedded vector)	$\in \mathbf{V}_{04}$?
`0x16a`	0	(`0x3fa`, `0x285`, `0x37e`, `0x0e5`)	yes
`0x04b`	1	(`0x2df`, `0x334`, `0x3bd`, `0x0dc`)	yes
`0x1a9`	2	(`0x354`, `0x0d4`, `0x12c`, `0x039`)	yes
`0x0f0`	3	(`0x18b`, `0x251`, `0x0c3`, `0x3f6`)	yes
(16 further basis vectors omitted; all 20 checked, all yes)			

7.3.4 The Full-Round and Full-Cipher Invariant

Theorem 3 (Invariant Subspace of BEA-1)**.** *Let* $\mathbf{T}_0 = \bigoplus_{i=0}^{7} \mathbf{V}_{i \bmod 4} \subset \mathbb{F}_2^{80}$ *as in Proposition 28. For any round key* $k \in \mathbb{F}_2^{80}$*: if* $\Delta R \in \mathbf{T}_0$ *then the resulting full-round output difference* $\Delta F \in \mathbf{T}_0$*, with probability* 1*. Consequently, for BEA-1 with the secret S-boxes, any key schedule, and any number of rounds* r*: if* $\Delta X_0 \in \mathbf{T}_0$ *then* $\Delta X_r \in \mathbf{T}_0$ *for every* $r \geq 0$*, with probability* 1*, where* X_r *is the state of Definition 20; and the final (MixColumns-free) partial round maps* $\mathbf{T}_0$ *into* $\mathbf{W} = \bigoplus_{i=0}^{7} \mathbf{W}_{i \bmod 4}$*.*

Proof. AddRoundKey leaves any XOR difference unchanged regardless of the key value, since $(x \oplus k) \oplus (x' \oplus k) = x \oplus x'$ identically. Given $\Delta R \in \mathbf{T}_0$, Table 36 gives SubBundles sending each bundle's difference from $\mathbf{V}_{i \bmod 4}$ to $\mathbf{W}_{i \bmod 4}$ exactly (verified exhaustively on all 32 cosets of each S-box), so the state after SubBundles lies in $\mathbf{W} = \bigoplus_i \mathbf{W}_{i \bmod 4}$. Lemma 6 gives ShiftRows($\mathbf{W}$) = $\mathbf{W}$. Applying Proposition 29 independently to each of the two 4-bundle groups gives MixColumns($\mathbf{W}$) = $\mathbf{T}_0$. Composing: $\Delta R \in \mathbf{T}_0 \Rightarrow \Delta F \in \mathbf{T}_0$, for every round key, proving the one-round claim. The full-cipher statement follows by induction on r as in Theorem 2, but without any Feistel bookkeeping. BEA-1 has no $\Delta L / \Delta R$ split, so each round's output difference is simply the next round's input difference. The final round omits MixColumns, so its output stops one layer earlier, in $\mathbf{W}$ rather than back in $\mathbf{T}_0$. □

Example 6 (Theorem 3 on a concrete trace)**.** Theorem 3 places Δ in $\mathbf{T}_0$ before every SubBundles and in $\mathbf{W}$ after it, for every key and all 11 rounds. We checked this layer by layer on 10 000 random triples (K, P_1, P_2) with $P_1 \oplus P_2$ drawn uniformly from $\mathbf{T}_0$, and found no violation. One such run, with a single active bundle so that the propagation is easy to follow, is shown in full. Take the master key and first plaintext

$$K = (\texttt{0f3}, \texttt{28e}, \texttt{0d2}, \texttt{1c9}, \texttt{35d}, \texttt{3ed}, \texttt{387}, \texttt{1eb}, \texttt{005}, \texttt{0a5}, \texttt{0e2}, \texttt{24c}),$$
$$P_1 = (200, 920, 23, 1004, 643, 430, 813, 515),$$

and let P_2 differ from P_1 only in bundle 0, by $360 = \texttt{0x168}$, a nonzero element of $\mathbf{V}_0$. Table 38 lists the difference between the two encryptions at two checkpoints of every round. The first is just after SubBundles, where the theorem predicts $\Delta \in \mathbf{W} = \bigoplus_i \mathbf{W}_{i \bmod 4}$, and the second just after MixColumns, where it predicts $\Delta \in \mathbf{T}_0 = \bigoplus_i \mathbf{V}_{i \bmod 4}$, the input class of the next round. AddRoundKey leaves the difference unchanged and ShiftRows only moves bundles within their S-box class (Lemma 6), so neither needs a column of its own.

Table 38: Full 11-round trace of Example 6. Round 10 is the final round, without MixColumns, and row C is the ciphertext difference after the final ShiftRows and key addition. Every entry of the left column lies in $\mathbf{W}$ and every entry of the right column in $\mathbf{T}_0$; with the plaintext difference $(360, 0, \ldots, 0) \in \mathbf{T}_0$ these are 23 exact membership checks. The support of Δ does not stay small. MixColumns spreads bundle 0 over bundles 0–3 in round 0, and in round 1 the ShiftRows swaps $(1, 5)$ and $(3, 7)$ carry it into the second group, so all eight bundles are active after round 1. Later rounds show zero bundles appearing and disappearing again. Neither the support nor any individual bit is preserved; only the class, $\mathbf{T}_0$ or $\mathbf{W}$, is.

Round r	Δ after SubBundles ($\in \mathbf{W}$)	Δ after MixColumns ($\in \mathbf{T}_0$)
0	$(763, 0, 0, 0, 0, 0, 0, 0)$	$(383, 155, 594, 355, 0, 0, 0, 0)$
1	$(325, 541, 566, 773, 0, 0, 0, 0)$	$(23, 920, 894, 684, 395, 227, 496, 220)$
2	$(992, 454, 879, 612, 992, 809, 675, 856)$	$(454, 79, 759, 893, 793, 120, 964, 229)$
3	$(113, 213, 961, 818, 396, 383, 852, 660)$	$(951, 338, 102, 488, 185, 593, 102, 355)$
4	$(362, 213, 675, 595, 283, 912, 240, 106)$	$(293, 390, 759, 269, 465, 55, 894, 178)$
5	$(809, 185, 460, 173, 362, 570, 817, 818)$	$(951, 844, 307, 585, 293, 357, 307, 624)$
6	$(94, 383, 460, 773, 509, 751, 616, 1013)$	$(360, 714, 0, 0, 555, 172, 865, 624)$
7	$(47, 308, 0, 0, 419, 809, 263, 962)$	$(951, 510, 406, 920, 782, 212, 186, 0)$
8	$(466, 809, 345, 713, 325, 75, 263, 0)$	$(360, 212, 186, 269, 852, 120, 0, 178)$
9	$(809, 454, 710, 927, 325, 912, 0, 106)$	$(572, 212, 657, 763, 360, 357, 792, 810)$
10	$(562, 837, 932, 675, 509, 712, 566, 660)$	no MixColumns
C	$(562, 712, 932, 660, 509, 837, 566, 675)$	

Remark 35 (Consistency with Bannier's own statement). Theorem 3 is the same claim as Bannier's [Ban17] Theorem 5.2 ("the round function preserves $\mathcal{L}(V)$ ... and the last round maps $\mathcal{L}(V)$ to $\mathcal{L}(W)$, no matter the round keys used"), but arrived at independently, from the S-box tables and the public M matrix alone, via the general mechanical procedure of Section 5.5, not assumed from the published specification. Where Section 7.2 checked Algorithm 1 at the S-box level against the published subspaces, this checks Algorithm 3 and Remark 20 at the round-function level against the same source.

Remark 36 (Why the deployed cipher is unaffected). Table 36 shows the *modified* (publicly deployed) S-boxes recover only the trivial subspace $\{0\}$ under Algorithm 1. Theorem 3 therefore says nothing about the deployed cipher directly. With no non-trivial $\mathbf{V}_i/\mathbf{W}_i$ pair at the S-box level, there is no $\mathbf{T}_0$ for Proposition 28 to construct in the first place, however favourably M behaves. The deployed cipher's weakness is of a different, probabilistic kind. It is the occasional agreement of the public and secret ciphers, for about 2^{-11} of the plaintexts, that Bannier's statistical attack exploits (Section 7.1), and it is independent of this section's exact, probability-1 result.

7.3.5 A Linear Layer Designed with the Partition

BEA-1's linear layer is its designers' own, and where DES's linear layer admitted Paterson's trapdoor by accident of its design criteria (Section 6.3.5), here the evidence points the other way. ShiftRows keeps every bundle in its S-box class, $\sigma(i) \equiv i \pmod 4$. Only 16 of the $8! = 40{,}320$ bundle permutations do so, and BEA-1's is one of the 15 among them that

still move bundles between the two MixColumns groups. MixColumns maps $\mathbf{W}_{04}$ onto $\mathbf{V}_{04}$ (Proposition 29). For a fixed pair of 20-dimensional subspaces, a uniformly random 40×40 binary matrix maps the first into the second with probability 2^{-400}, each of the 20 basis images landing there with probability 2^{-20}. Even allowing the designer any choice of 5-dimensional subspaces per bundle, $[10, 5]_2 = 2^{26.70}$ of them, the union bound over all $2^{213.6}$ such pairs leaves probability at most $2^{-186.4}$ that a MixColumns chosen without regard to the partition preserves one. BEA-1's MixColumns does so while attaining the maximal bundle branch number, inside the trapdoor block as well as outside it. The linear layer and the partition were designed together, consistently with BEA-1's stated purpose as a demonstration of a mathematical backdoor [BF17, Ban17]. The deployed variant keeps the same ShiftRows and MixColumns and changes only the S-boxes, so its linear layer is partition-preserving too. What it lacks is S-boxes that map cosets to cosets.

7.4 Differential Characteristics

Theorem 3 gives a probability-1 distinguisher, but only at the level of *coset* membership. It says nothing about how large the probability of a *single point-differential trail* can be. In this subsection we ask that separate question, as Section 6.6.4 did for KG-DES, and answer it exactly for up to four rounds.

A characteristic over L S-box layers (Definition 13) fixes the difference α_r entering and β_r leaving S-box layer r for $r = 0, \ldots, L-1$. Key additions leave differences unchanged, consecutive layers are linked by $\alpha_{r+1} = \mathrm{MC}(\mathrm{SR}(\beta_r))$, α_0 and β_{L-1} are free, and each active S-box contributes $\mathrm{DDT}(\alpha, \beta)/1024$. Three facts about the secret S-boxes and the linear layer settle the optimum.

1. The largest DDT entry of every secret S-box is $46/1024 = 2^{-4.476}$, and it occurs only for input differences in $\mathbf{V}_i$; with $\alpha \notin \mathbf{V}_i$ no entry exceeds $18/1024 = 2^{-5.830}$. For $\alpha \in \mathbf{V}_i$ the output difference always lies in $\mathbf{W}_i$, and the next layer's input difference in $\mathbf{V}_j$ again (Theorem 3). A characteristic that starts in the trapdoor block therefore stays there.

2. MixColumns has differential branch number 5 on bundles, the maximum for four bundles, as Bannier shows [Ban17, §5.1.2]. From the branch number he bounds ten rounds by 25 active S-boxes, ignoring ShiftRows; the count below takes ShiftRows into account and gives 35. Counting over $\mathbb{F}_2$ the differences with prescribed input and output bundle supports shows, for every pair of supports, whether a transition exists. A dynamic programme over the 2^8 bundle activity patterns of the S-box layers then gives the minimum number $A(L)$ of active S-boxes. For $L = 1, \ldots, 11$ it is $1, 5, 8, 15, 16, 20, 23, 30, 31, 35, 38$. Every characteristic therefore has weight at least $4.476\, A(L)$.

3. A characteristic with an S-box outside the trapdoor block pays at least $5.830 - 4.476 = 1.354$ bits more than that bound.

A branch-and-bound search inside the block, over activity patterns and then over exact differences, finds the optimum there. Where it is within 1.354 bits of $4.476\, A(L)$, fact 3 shows that no characteristic outside the block can beat it, and the block optimum is the global one.

Proposition 30 (Optimal differential characteristics of BEA-1)**.** *With the secret S-boxes, the best characteristic over L S-box layers has probability $2^{-4.476}$, $2^{-22.382}$, $2^{-36.068}$ and $2^{-67.672}$ for $L = 1, 2, 3, 4$; in each case it lies in the trapdoor block. Over all 11 rounds every characteristic has probability at most $2^{-38 \times 4.476} = 2^{-170.1}$.*

Proof. The block search returns the stated values with 1, 5, 8 and 15 active S-boxes. For $L = 4$ we start it at a known weight, which makes it exhaustive below that weight in a single pass. Prepending a layer to the optimal 3-layer characteristic, with the adjacent free choices optimized exactly, gives weight 67.86 (appending one gives 67.94), and the search below 67.861 finds the optimum 67.672 of Table 40. The lower bounds from fact 2 are $2^{-4.476}$, $2^{-22.382}$, $2^{-35.812}$ and $2^{-67.15}$ for $L = 1, \ldots, 4$, and any characteristic leaving the block is bounded by $2^{-5.830}$, $2^{-23.736}$, $2^{-37.165}$ and $2^{-68.50}$, in every case below the block value found. The 11-round bound is fact 2 with $A(11) = 38$. □

Table 39: An optimal 3-round differential characteristic of BEA-1 with the secret S-boxes, as bundle : difference (hex) for the active bundles; probability $2^{-36.068}$ with 8 active S-boxes. Every input difference lies in $\mathbf{V}_{j \bmod 4}$ and every output difference in $\mathbf{W}_{j \bmod 4}$.

Layer	Input difference α_r	Output difference β_r	DDT/1024
0	3: `10d`, 4: `132`, 6: `079`	3: `0c7`, 4: `243`, 6: `1f7`	46, 46, 44
1	4: `30e`, 6: `22b`	4: `05e`, 6: `13c`	44, 46
2	4: `05a`, 5: `266`, 7: `270`	4: `377`, 5: `04b`, 7: `1fb`	44, 44, 46

Table 40: An optimal 4-round differential characteristic of BEA-1 with the secret S-boxes, in the notation of Table 39; probability $2^{-67.672}$ with 15 active S-boxes, the minimum $A(4)$. Ten of them take the maximal entry 46.

Layer	Input difference α_r	Output difference β_r	DDT/1024
0	1: `165`, 3: `1d1`, 6: `079`	1: `283`, 3: `39f`, 6: `192`	46, 44, 46
1	5: `186`, 7: `1d1`	5: `2c8`, 7: `253`	46, 42
2	0: `0e3`, 1: `0d4`, 2: `0ba`, 3: `057`	0: `1fd`, 1: `17f`, 2: `0ae`, 3: `13c`	44, 40, 46, 46
3	1: `186`, 2: `318`, 3: `057`, 4: `30e`, 6: `079`, 7: `134`	1: `2c8`, 2: `253`, 3: `13c`, 4: `05e`, 6: `192`, 7: `20e`	46, 46, 46, 44, 46, 46

BEA-1's best S-box transition is much weaker than KG-DES's (46/1024 against 24/64), and branch number 5 forces at least five active S-boxes in every two consecutive rounds. The trapdoor still shapes the optimum, since every S-box of the optimal characteristics sits on a maximal or near-maximal entry of the trapdoor block. It does not rescue the point differential. Three rounds cost $2^{-36.07}$, four cost $2^{-67.67}$, and the full cipher is bounded by $2^{-170.1}$, far past the 2^{-80} that an 80-bit block could ever exploit.

Remark 37 (Why this does not contradict the probability-1 result). Theorem 3 and Proposition 30 answer different questions. The theorem says that, *summed over every branch the DDT offers at every round*, a pair whose difference lies in $\mathbf{T}_0$ keeps it there after any number of rounds with probability exactly 1. A truncated differential of this kind claims membership in a fixed subspace, never a specific point value. The proposition instead follows a single, fixed sequence of exact point differences, a strictly harder event that degrades the way any ordinary differential trail through a well-diffused SPN does. This is why the one-pair distinguisher of Remark 33 (the analogue of Corollary 3) and Bannier's statistical attack both operate on *coset membership*, never on a specific ciphertext difference. BEA-1's trapdoor hands an attacker a permanent structural distinguisher but no usable multi-round point differential. KG-DES offers both, the second because DES's linear layer has branch number 2 (Section 8.1).

7.5 Connectivity Tables of the Secret S-Boxes

A companion study [FAI+23] profiled backdoored S-boxes against 24 cryptanalytic parameters. Among them were three boomerang-family tables, the Differential-Linear Connectivity

Table (DLCT), the Boomerang Connectivity Table (BCT) and the Feistel Boomerang Connectivity Table (FBCT). The study found that the KG-DES S-boxes and several small, 4-to-6-bit backdoored constructions of Bannier's own attain the upper bound of these tables. However, it did not examine BEA-1's 10-bit secret S-boxes, and we close that gap here on the real cipher data. The tables are those of Definitions 11 and 12, applied earlier to KG-DES's non-square S-boxes in Section 6.4. Here the S-boxes are square and bijective ($m = n = 10$), so the BCT applies as well.

Proposition 31 (DLCT attains its exact upper bound on the trapdoor directions)**.** *For every $i \in \{0, 1, 2, 3\}$, every $\Delta \in \mathbf{V}_i \setminus \{0\}$ and every nonzero γ in the annihilator $\mathbf{W}_i^{\perp}$ of $\mathbf{W}_i$, which has 31 nonzero elements,* $\mathrm{DLCT}_{S_i}(\Delta, \gamma) = 2^9 = 512$*. This is the theoretical maximum for a 10-bit S-box.*

Proof. The proof of Proposition 18 applies with $m = 10$ and $\mathbf{V}_i$ in place of $\mathbf{U}$. S_i maps each coset of $\mathbf{V}_i$ into a single coset of $\mathbf{W}_i$ (Table 36), so $\gamma \cdot S_i$ is constant on the cosets of $\mathbf{V}_i$ and its Walsh support lies in $\mathbf{V}_i^{\perp} \subseteq \Delta^{\perp}$. Equation (6) and Parseval's relation then give $\mathrm{DLCT}_{S_i}(\Delta, \gamma) = 2^{-11} \cdot 2^{20} = 2^9$. □

We checked Proposition 31 exhaustively on the real secret S-box tables, over all 1023×1023 nonzero pairs (Δ, γ) per S-box and not only the $961 = 31 \times 31$ pairs that the proposition identifies. For every S_i the DLCT uniformity is 512, and it is attained at those 961 pairs and at no others. The proposition therefore accounts for every maximizer. Table 41 reports the DLCT, BCT and FBCT uniformities of all four secret S-boxes. For contrast, it also gives the same computation for the four publicly deployed S-boxes. These carry no trapdoor subspace at all (Table 36), so Proposition 31 does not apply to them.

Table 41: DLCT, FBCT, and BCT uniformity (Definitions 11, 12), computed exhaustively over all nonzero $(\Delta, \gamma)/(\Delta, \nabla)$ pairs for each of BEA-1's four secret S-boxes and, for contrast, its four publicly deployed (modified) counterparts. DLCT is at the theoretical maximum for every secret S-box (Proposition 31) and never reaches it for any public S-box. FBCT and BCT are elevated for the secret S-boxes relative to the public ones, but neither approaches its own upper bound for either variant.

S-box variant	DLCT	FBCT	BCT
Secret S_0–S_3 (backdoored)	512 (exact)	84–88	88–104
Public S_0–S_3 (deployed)	444–464	64 (all four)	72–76
Theoretical upper bound	512	1024	1024

Remark 38 (A maximal DLCT entry is a cheap detector for the trapdoor). Table 41 gives a second route to *detecting* the secret S-boxes, separate from Algorithm 1. The public S-boxes, ordinary-looking bijections in which Algorithm 1 finds only the trivial subspace, approach the bound (444–464) but never reach it. The difference between a value close to 512 and the value 512 itself is the difference between the statistical behaviour of a generic permutation and an algebraic guarantee. Moreover, it costs nothing extra to look for, since the DLCT uniformity is part of a standard hybrid-attack profile.

However, a single maximal entry certifies less than the full trapdoor. A maximal entry at (Δ, γ) means that $\gamma \cdot D_{\Delta}\pi(x) = 0$ for every x, i.e., that the image of $D_{\Delta}\pi$ lies in the hyperplane $\gamma^{\perp}$. This is the necessary condition of Lemma 1, and on its own it bounds the confinement only by dimension $n - 1$. The dimension 5 is fixed by the set of maximal entries, not by any one of them. The maximizers form the set $(\mathbf{V}_i \setminus \{0\}) \times (\mathbf{W}_i^{\perp} \setminus \{0\})$ of $961 = 31 \times 31$ pairs, so both subspaces can be read off the maximal cells. An evaluator who uses the DLCT as a detector should therefore read the whole set of maximizers, not a single cell.

Restricted to the trapdoor directions, the secret S-boxes behave in the opposite way to KG-DES's.

Proposition 32 (Second-order behaviour on BEA-1's trapdoor directions)**.** *For every secret S-box S_i:*

1. $\mathrm{FBCT}_{S_i}(\Delta, \nabla) = 0$ *for all distinct* $\Delta, \nabla \in \mathbf{V}_i \setminus \{0\}$*. Equivalently, the restriction of S_i to every coset of $\mathbf{V}_i$, a bijection onto a coset of $\mathbf{W}_i$, is an APN map between* 5*-dimensional affine spaces.*
2. $\mathrm{BCT}_{S_i}(\Delta, \nabla) \leq 46$ *for* $\Delta \in \mathbf{V}_i$ *and* $\nabla \in \mathbf{W}_i$*, which is the largest DDT entry.*

Proof. Exhaustive computation over the stated pairs for each S_i. For part 1, the four inputs x, $x \oplus \Delta$, $x \oplus \nabla$, $x \oplus \Delta \oplus \nabla$ lie in one coset of $\mathbf{V}_i$, so the count is over second derivatives of the coset maps, and a map none of whose second derivatives vanish is APN by definition. □

This explains why a trapdoor of this kind forces the DLCT to its bound but not the FBCT. The DLCT is a first-order quantity. An entry $\mathrm{DLCT}(\Delta, \gamma) = 2^{m-1}$ says only that $D_\Delta \pi$ takes all its values in the hyperplane $\gamma^\perp$, and the confinement of the first derivative to $\mathbf{W}_i$ guarantees this for both ciphers (Propositions 18 and 31). The FBCT, on the other hand, is a second-order quantity. It reaches its bound on the trapdoor directions if and only if the maps between cosets are affine. For KG-DES the cosets have four elements, every bijection between them is affine, and the FBCT is at its bound (Proposition 21). For BEA-1 the cosets have 32 elements, and the coset maps of the secret S-boxes are APN, the opposite extreme, so the FBCT on the trapdoor directions is 0. The large BCT and FBCT values of Table 41 all come from directions outside the trapdoor. With $\Delta \in \mathbf{V}_i$ and $\nabla \in \mathbf{W}_i$, the BCT gains nothing over the DDT.

7.6 Differential-Linear Trails and Boomerangs

7.6.1 A Correlation-1 Differential-Linear Trail

The decomposition $E = E_1 \circ E_m \circ E_0$ of Section 6.9.1 applies to BEA-1 unchanged, and again all three factors are exactly 1.

Proposition 33 (Correlation-1 differential-linear trails of BEA-1)**.** *With the secret S-boxes, let $\mathbf{W} = \bigoplus_{i=0}^{7} \mathbf{W}_{i \bmod 4}$. For every key, every nonzero $\Delta \in \mathbf{T}_0$ and every nonzero $\lambda \in \mathbf{W}^\perp$, the full* 11*-round cipher has $c_E(\Delta, \lambda) = 1$, and no differential-linear distinguisher has larger absolute correlation.*

Proof. Theorem 3 places every ciphertext difference in $\mathbf{W}$, so every term of the sum is $+1$. Optimality is the bound $|c_E| \leq 1$. □

Table 42 gives one explicit trail. The final ShiftRows leaves bundle 0 in place, so the mask reaches S_0 unchanged and Proposition 31 gives the maximal DLCT entry whatever input difference in $\mathbf{V}_0$ reaches it. Figure 8 draws the trail in the convention of Section 6.9.1; from X_2 on every cell is red and the probability is still 1.

The trail was verified as the KG-DES one was. We checked every layer exhaustively, i.e., the MixColumns on bases of the relevant subspaces, each S-box on all inputs and all differences in $\mathbf{V}_i$, and the DLCT entries of E_m. We also ran the full cipher with random keys, against the public S-boxes as a control (Table 43). The trail gives no capability beyond the invariant subspace. Testing every mask of $\mathbf{W}^\perp$ at once is testing membership in $\mathbf{W}$, which a random permutation passes with probability 2^{-40}. One chosen pair therefore distinguishes BEA-1 with advantage $1 - 2^{-40}$, the analogue of Corollary 3.

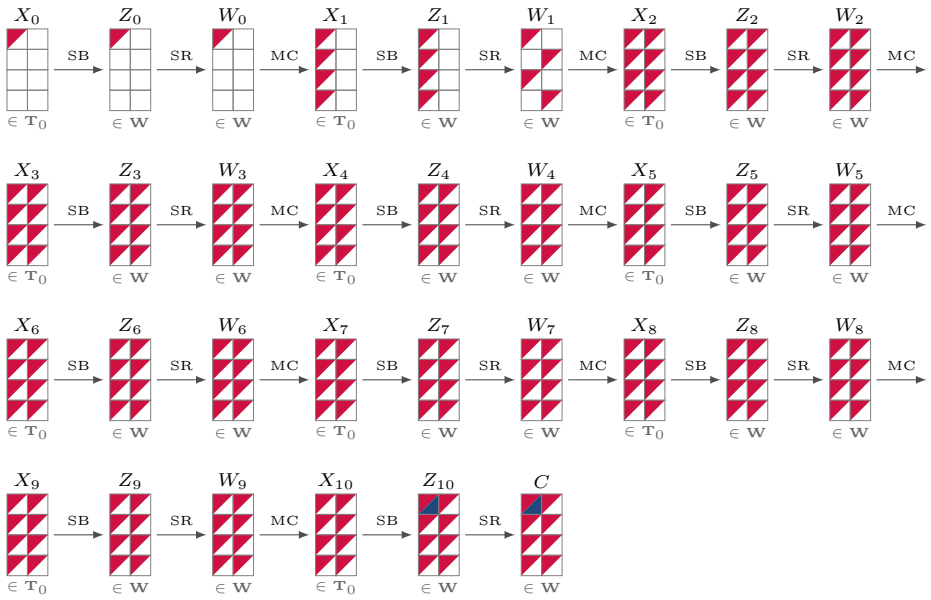

Figure 8: Correlation-1 differential-linear trail through the full 11-round BEA-1 with the secret S-boxes. Each state holds eight 10-bit bundles, bundle j in row $j \bmod 4$ and column $\lfloor j/4 \rfloor$, so each column is one MixColumns group and ShiftRows exchanges rows 1 and 3 between the columns. Red: bundle difference possibly nonzero, lying in $\mathbf{V}_{j \bmod 4}$ before SubBundles ($X_r \in \mathbf{T}_0$) and in $\mathbf{W}_{j \bmod 4}$ after it ($Z_r, W_r \in \mathbf{W}$). Blue: the output mask $\lambda =$ `0x01f` on bundle 0, carried back through the final ShiftRows to S_0; the cell of Z_{10} carrying both colours is the middle layer E_m. Key additions leave differences unchanged and are not drawn; C is W_{10} after the last key addition. RU = 10, RM = 1, RL = 0; $p = r = q = 1$.

Table 42: An explicit correlation-1 differential-linear trail of BEA-1 with the secret S-boxes. Any nonzero Δ and λ from the subspaces of Proposition 33 give the same correlation, with $2^{40} - 1$ choices of each.

Δ	bundle 0: `0x168` $\in \mathbf{V}_0$; other bundles 0
λ	`0x01f` $\in \mathbf{W}_0^{\perp}$ on ciphertext bundle 0
E_0, p	rounds 0–9 and the final key addition; difference lands in $\mathbf{T}_0$ with $p = 1$
E_m, r	final SubBundles, $\gamma =$ `0x01f` on S_0; DLCT $= 512 = 2^9$ for every $v \in \mathbf{V}_0^*$, $r = 1$
E_1, q	final ShiftRows and key addition; mask carried exactly, $q = 1$
c_E	1

Table 43: Full-cipher measurement of the trail of Table 42, over 20 random keys per row, with BEA-1's own key schedule. The control correlation is within about two standard deviations of 0.

Cipher	Pairs	$c_E(\Delta, \lambda)$	Difference in $\mathbf{W}$
BEA-1, secret S-boxes	5,000	+1.0000	5,000
BEA-1, public S-boxes (control)	5,000	+0.011	0

7.6.2 Generic Ladder Boomerangs

That the DLCT reaches its exact bound (Proposition 31) tells an *attacker* nothing the theory does not already give. It is the probability-1 coset relation of Theorem 3, seen through a different table. (Its value is to an *evaluator*, who does not yet know the trapdoor and can use the DLCT profile to find it; see Remark 38.) The BCT and FBCT describe what a boomerang switch costs, and the estimate of Section 6.9.2 applies with them.

BEA-1's coset maps are APN rather than affine (Proposition 32), so nothing like Proposition 27 holds, and the trapdoor offers no favourable switch either. With $\Delta \in \mathbf{V}_i$ and $\nabla \in \mathbf{W}_i$ the BCT never exceeds the DDT maximum, and the larger entries of Table 41 send the bottom trail out of the block. Generic boomerangs are another matter. Table 44 gives measured return probabilities on BEA-1 reduced to L S-box layers, with independent random round keys, a single-bundle top difference and a single-bundle bottom difference.

Table 44: Measured boomerangs on BEA-1 (secret S-boxes) reduced to L S-box layers, top difference `0x0b9` in one bundle and bottom difference `0x2d4` in one ciphertext bundle placed as described in the text, against the optimal differential characteristic (Proposition 30). At three and four layers we measured two placements. The bound at four layers is the 95% upper limit for no observed return in 2^{28} quartets.

Layers	Quartets returned	Boomerang	Optimal differential characteristic
2	2^{20} of 2^{20}	1	$2^{-22.38}$
3	594 and 589 of 2^{28}	$2^{-18.8}$	$2^{-36.07}$
4	0 and 0 of 2^{28}	below $2^{-26.4}$	$2^{-67.67}$

The two-layer boomerang returns always, for a reason that has nothing to do with the trapdoor. After the first layer, a single active bundle occupies only its own MixColumns group. A bottom difference placed in a bundle of the other group touches only S-boxes that the top pair leaves inactive, on the way down and on the way back. Every S-box of the switch is a ladder step, and any two S-box layers around one linear layer of this shape behave the same way. With the bottom difference moved onto the top trail's group the return rate drops to $2^{-4.96}$. At three layers the two trails can no longer be kept apart. Traced back from a single ciphertext bundle, the bottom difference fills one MixColumns

group before the last layer, and ShiftRows makes that group overlap the top trail's group in exactly two S-boxes of the middle layer, whatever the positions chosen. The return probability falls to $2^{-18.8}$. Both values are far above the optimal characteristics over the same layers. At four layers no quartet returned, which bounds the probability by $2^{-26.4}$ and leaves the comparison with the characteristic open. None of this touches the full cipher, since the trapdoor's truncated differential holds with probability 1 at every round count and remains BEA-1's best distinguisher.

7.7 Linear Characteristics

Linear characteristics behave differently in BEA-1 than in KG-DES (Section 6.8). Its linear layer diffuses far better than a Feistel round, and the best characteristic is pinned by counting active S-boxes.

Proposition 34 (Optimal linear characteristic of BEA-1)**.** *Every linear characteristic of the full* 11*-round BEA-1 has correlation at most* 2^{-76} *in absolute value, with the secret or with the public S-boxes. With the secret S-boxes the bound is attained, for every key, by a characteristic whose masks all lie in the trapdoor block, with input masks in* $\mathbf{V}_i^{\perp}$ *and output masks in* $\mathbf{W}_i^{\perp}$*.*

Proof. Every S-box of either set has linear correlation at most 2^{-2} in absolute value (the public S_3 reaches only 252/1024). MixColumns has linear branch number 5 at bundle level, as Bannier also shows [Ban17, §5.1.2]. Counting, over $\mathbb{F}_2$, the masks with prescribed input and output bundle supports shows that no nonzero mask has fewer than five active bundles on its two sides together. An exact dynamic programme over the 2^8 bundle activity patterns of the eleven S-box layers, using those exact support counts, shows that every nonzero characteristic activates at least 38 S-boxes, so $|c| \le 2^{-2\cdot 38}$. For the secret S-boxes, every Walsh coefficient with nonzero $\alpha \in \mathbf{V}_i^{\perp}$ and nonzero $\beta \in \mathbf{W}_i^{\perp}$ has correlation 0 or exactly $\pm 2^{-2}$. The characteristic of Table 45, found by searching the trapdoor block along optimal activity patterns, has 38 active S-boxes, each of correlation $\pm 2^{-2}$, and consecutive masks related by MixColumns and ShiftRows; its correlation is $\pm 2^{-76}$. □

Table 45: Optimal linear characteristic of the full 11-round BEA-1 with the secret S-boxes: input masks α_r and output masks β_r of S-box layer r, in hex per bundle $0, \ldots, 7$ ("·" for an inactive bundle). The plaintext mask is α_0; the ciphertext mask is β_{10} after the final ShiftRows. Every active S-box has correlation $\pm 2^{-2}$, and 38 are active.

r	act.	α_r									β_r					
0	3	·	01f	·	039	04d	·	·	·	·	09d	·	136	319	·	·	·
1	2	·	·	·	·	0bc	·	03b	·	·	·	·	·	2fb	·	233	·
2	3	·	·	·	·	30b	14d	03b	·	·	·	·	·	2fb	2ef	233	·
3	7	3b7	0d4	3db	05e	30b	14d	03b	·	1e2	0f1	268	20e	01f	04b	15c	·
4	3	·	24e	·	08b	132	·	·	·	·	0f1	·	2fe	25c	·	·	·
5	2	·	·	·	·	1c3	·	0a5	·	·	·	·	·	306	·	0f0	·
6	3	·	·	·	·	292	067	09e	·	·	·	·	·	306	2ef	0f0	·
7	7	3b7	0d4	3db	05e	292	067	09e	·	1e2	06c	268	057	3be	04b	298	·
8	3	·	24e	·	08b	05a	·	·	·	·	0f1	·	2fe	25c	·	·	·
9	2	·	·	·	·	1c3	·	0a5	·	·	·	·	·	306	·	0f0	·
10	3	·	·	·	·	292	067	09e	·	·	·	·	·	01f	04b	095	·

Here the trapdoor saturates the S-box spectrum without helping any single characteristic. The subspace structure leaves each secret S-box with 496 approximations of the maximal correlation 2^{-2}, where the public S-boxes have between 8 and 17, and all 496 lie in the trapdoor block. An approximation inside the block reads only the coset of $\mathbf{V}_i$ that the input lies in. The permutation that S_i induces from the 32 cosets of $\mathbf{V}_i$ to the 32 cosets of

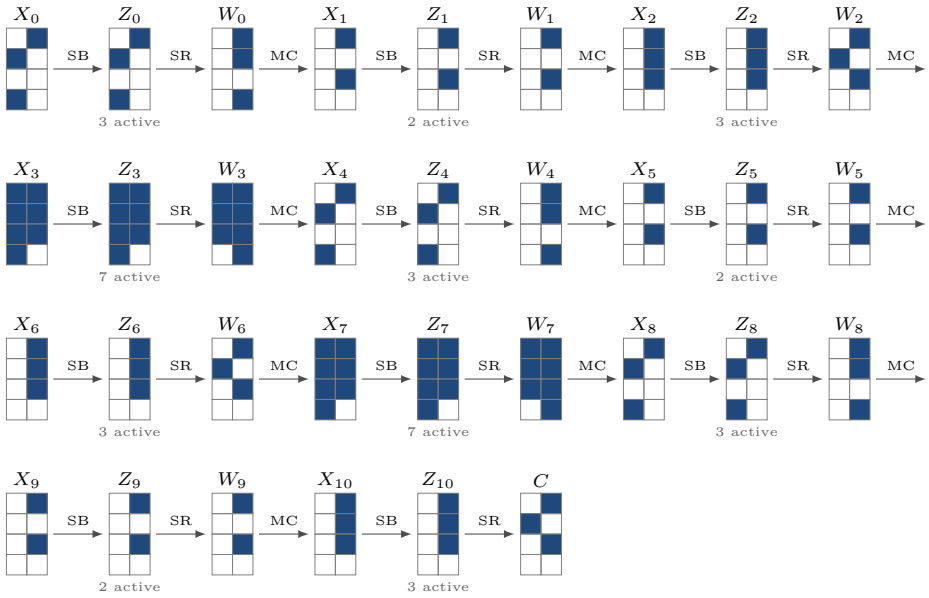

Figure 9: The characteristic of Table 45, drawn in the layout of Figure 8. Blue: nonzero bundle mask, lying in $\mathbf{V}^{\perp}_{j \bmod 4}$ before SubBundles and in $\mathbf{W}^{\perp}_{j \bmod 4}$ after it. Key additions are not drawn. 38 active S-boxes, correlation $\pm 2^{-76}$ for every key.

$\mathbf{W}_i$ is almost bent, with every Walsh coefficient equal to 0 or ± 8. An approximation inside the block is therefore either unbiased or of the maximal correlation 2^{-2}. A characteristic confined to the block can pay the minimum on every active S-box, and one exists that meets the active-S-box bound. The bound is still 2^{-76}, and a single characteristic of that correlation would need about 2^{152} known plaintexts, far beyond the 2^{80}-element codebook. For characteristics, the linear picture thus repeats the differential one. KG-DES's trapdoor opens the full cipher to a known-plaintext distinguisher, whereas BEA-1's diffusion holds every characteristic far below any usable correlation. It does not hold the linear hull. Section 7.8 shows that for every key the full BEA-1 has linear approximations of correlation at least 2^{-20}, all of them inside the trapdoor block.

7.8 The Partition in the Correlation Matrix

For the bijective secret S-boxes either half of the Fourier condition of Proposition 4 implies the other, and both hold for all four with their pairs $(\mathbf{V}_i, \mathbf{W}_i)$, while both fail for the four public S-boxes against the same subspaces. The block of C^{S_i} on $\mathbf{V}_i^{\perp} \times \mathbf{W}_i^{\perp}$ is a 32×32 orthogonal matrix with entries 0 and $\pm 1/4$, the correlation matrix of the almost bent permutation that S_i induces on the cosets (Section 7.7). Figure 10 shows the zero pattern.

At the level of the cipher, Theorem 3 says that E_K maps every coset of $\mathbf{T}_0$ into a coset of $\mathbf{W}$. Theorem 1 therefore applies with $\mathbf{T} = \mathbf{T}_0$, $\mathbf{T}' = \mathbf{W}$, $N = 80$ and $d = 40$. For every key the block of C^{E_K} on $\mathbf{T}_0^{\perp} \times \mathbf{W}^{\perp}$ is the correlation matrix of a bijection between the 2^{40} cosets, an exact 40-bit cipher inside BEA-1. Each space in the mask trail is a tensor product of one subspace per bundle, spanned by the characters with masks in $\mathbf{V}^{\perp}_{i \bmod 4}$ or $\mathbf{W}^{\perp}_{i \bmod 4}$: the subspace analogue of Beyne's rank-one approximations [Bey21, Section 6]. SubBundles maps each factor onto the next, and ShiftRows and MixColumns carry one product onto the next (Figure 11).

Corollary 2 now changes the picture of Section 7.7. Every characteristic of the full cipher has correlation at most 2^{-76} (Proposition 34), yet for every key and every nonzero input

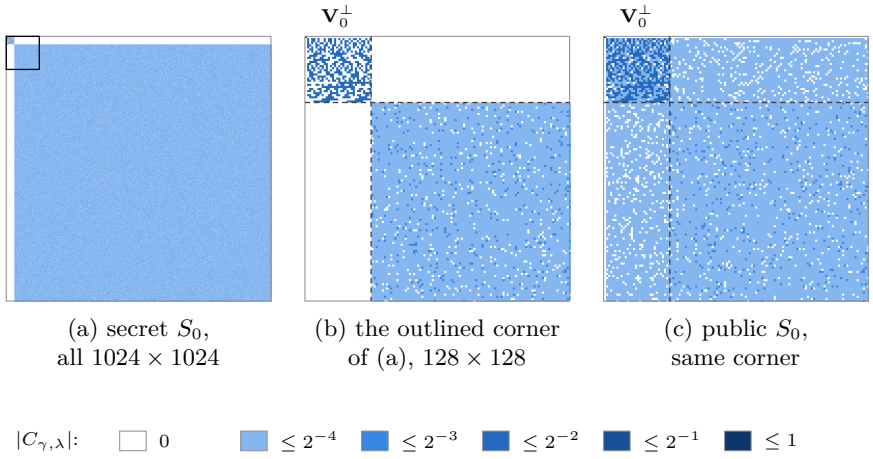

Figure 10: Correlation matrices of BEA-1's S_0 (rows: output masks γ; columns: input masks λ) with the masks of $\mathbf{W}_0^\perp$ and $\mathbf{V}_0^\perp$ listed first; white is exactly 0. (a) The secret S_0 in full: 32 rows and 32 columns are zero outside their common corner. (b) The outlined 128×128 corner of (a): the 32×32 block, with entries 0 and $\pm 1/4$, framed by zeros. (c) The public S_0 under the same ordering: no zero structure; 53,841 entries of the off-diagonal strips are nonzero. In every panel $C_{0,0} = 1$, and row and column 0 are otherwise 0 because the S-boxes are balanced.

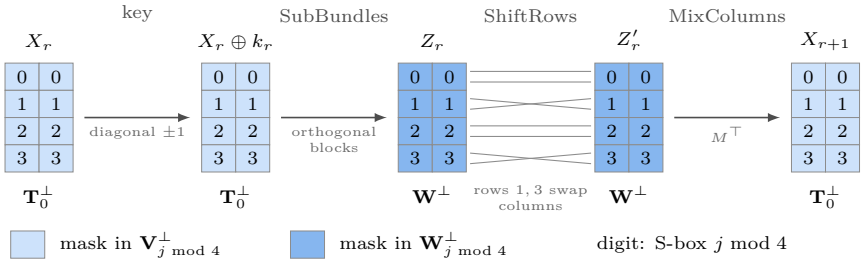

Figure 11: The mask subspace trail of BEA-1 through one round. Each cell is a 10-bit bundle, placed as in Figure 8; the whole state mask lies in $\mathbf{T}_0^\perp$ before SubBundles and in $\mathbf{W}^\perp$ after it, a product of one 5-dimensional subspace per bundle. Key addition only multiplies characters by ± 1, SubBundles connects $\mathbf{V}_i^\perp$ to $\mathbf{W}_i^\perp$ through the orthogonal blocks of Figure 10, ShiftRows keeps every bundle in its S-box class (the wires), and $M^\top$ maps $\mathbf{T}_{04}^\perp$ onto $\mathbf{W}_{04}^\perp$ in each column. The round repeats; the last one, without MixColumns, ends in $\mathbf{W}^\perp$.

mask of the block there is an output mask with correlation at least $(2^{40}-1)^{-1/2} > 2^{-20}$. The correlation is carried by the linear hull inside the block, and no single trail comes close. For masks in the block, BEA-1 with the secret S-boxes behaves like a 40-bit block cipher.

The known-plaintext distinguisher of Section 4.4 needs about $2^{20.5}$ texts for one expected pair in a common coset of $\mathbf{T}_0$, and $2^{21.5}$ for success probability about 0.98. A random permutation puts the ciphertexts of such a pair in one coset of $\mathbf{W}$ with probability 2^{-40} (Table 46). It distinguishes the full cipher from known plaintexts alone, where Bannier's attack on the deployed cipher uses chosen plaintexts.

Table 46: The known-plaintext partition distinguisher on the full BEA-1. A pair is two known plaintexts in one coset of $\mathbf{T}_0$; the expected number is $N(N-1)/2$ divided by the number of cosets, 32 per key. The last column counts pairs whose ciphertexts lie in one coset of $\mathbf{W}$. The control uses the same plaintexts, subspaces and keys.

Cipher	Texts per key	Keys	Pairs	Ciphertexts in one coset
BEA-1 secret, 11 rounds	2^{23}	4	115	115
BEA-1 public (control)	2^{23}	4	115	0

7.9 What the Analysis of BEA-1 Shows

The literature gives us the construction, Bannier's theorem that the round function preserves the partition, and his statistical attack on the deployed cipher. From the tables and the MixColumns matrix alone we reproduce the four trapdoor pairs and the invariant subspace (Sections 7.2 and 7.3). The rest is new. It comprises the optimal differential characteristics over up to four layers (Proposition 30) and the connectivity-table profile with the exact DLCT maximizers and the vanishing FBCT on the trapdoor directions (Propositions 31 and 32). It also comprises the measured boomerangs, the correlation-1 differential-linear trail and the optimal linear characteristic of correlation 2^{-76}. Finally, it comprises the perfect 40-bit approximation, with its linear approximations of correlation above 2^{-20} and its known-plaintext distinguisher.

The comparison with KG-DES is the point of this case study. The trapdoor is the same kind of object, a partition that every round preserves, and the method finds it just as reliably. What the trapdoor gives an attacker is different. BEA-1's coset maps are almost perfect nonlinear, not affine, and its MixColumns has the maximal branch number, so no point differential, characteristic or boomerang profits from the partition. The value of the trapdoor lies in coset membership and in the linear hull, not in any single trail. The limitation is that every attack here is a distinguisher.

8 The Two Partition Trapdoors Compared

The same method finds the trapdoor in KG-DES and in BEA-1, and the trapdoor has the same form in both, yet it gives an attacker very different things in the two ciphers. This is where the taxonomy is tested hardest, because placing two ciphers under one heading does not make them equally exploitable. What separates them is diffusion. In this section we compare the diffusion of the two linear layers, which decides whether a trapdoor transition can chain into a point trail. In addition, we give an independent MILP cross-check of the trails of both ciphers, and tables of the best trail of every type and of the data, time and memory of every attack.

8.1 Diffusion and the Value of a Partition

A partition trapdoor asks two things of a cipher's linear layer. It must carry the subspace in which the S-boxes leave their differences onto the subspace in which the next S-boxes expect them. For KG-DES this means $P(\bigoplus \mathbf{W}_i) \subseteq \mathbf{S}$ and $E(\mathbf{S}) \subseteq \bigoplus \mathbf{U}$ (Lemmas 5 and 4), and for BEA-1 it means $\mathrm{MC}(\mathrm{SR}(\mathbf{W})) = \mathbf{T}_0$ (Lemma 6 and Proposition 29). And, for the cipher to look sound, it must diffuse. Sections 6.3.5 and 7.3.5 showed how the first property came about in each cipher. Here we compare the second, and draw the consequence for trail search.

Branch numbers. Table 47 gives the differential and linear branch numbers (Definition 14) of both ciphers' layers. Bit-level values are the minimum distance of the code $\{(x, Mx)\}$. Every codeword has $w(x) \le t$, $w(Mx) \le t$ or weight at least $2t + 2$, so enumerating both sides to weight t gives the exact value whenever the lightest codeword found has weight at most $2t + 2$; $t = 4$ and $t = 5$ suffice for MixColumns.

Table 47: Branch numbers of the linear layers. *Attained with nothing active on the round-function side, since $E^\top$ has a 16-dimensional kernel. The bundle-level values for MixColumns are Bannier's [Ban17, §5.1.2]; the others are ours. MixColumns reaches the maximum for four bundles, $4 + 1 = 5$, on 57,288 vectors per column, and keeps it inside the trapdoor block; the bit-level minima are attained by $4 + 5$ and $4 + 7$ bits.

Cipher	Layer	Level	Differential	Linear
KG-DES, DES	P	bit, S-box	2	2
	$E \circ P$ (S-box outputs to next S-box inputs)	S-box	2	2*
BEA-1	ShiftRows	bit, bundle	2	2
	MixColumns	bundle	5	5
	MixColumns, trapdoor block	bundle	5	5
	MixColumns	bit	9	11

The two ciphers sit at opposite ends. DES's layer has the smallest branch number possible, 2, so all of its resistance must come from the S-boxes. Each S-box output still reaches six distinct S-boxes of the next round, two of its bits landing on middle input bits (one S-box each) and two on shared end bits (two S-boxes each). In the linear direction, $E^\top$ maps 48 bits onto 32 and has a 16-dimensional kernel, one vector for each duplicated input bit, covering its two copies in adjacent S-boxes. Two active S-boxes can therefore meet a zero round-function input mask. This is the $X = 0$ round of Proposition 25, and branch number 2 is what lets the optimal differential characteristics of Proposition 22 run on one active S-box per round. BEA-1's MixColumns, by contrast, is optimal at bundle level, which is where the active-S-box bounds of Propositions 30 and 34 come from.

Consequence for trail search. Every maximal DDT entry and Walsh coefficient of both ciphers' trapdoored S-boxes lies in the trapdoor block, and leaving the block costs a fixed amount per S-box, the ratio of the best entry inside to the best entry outside, which is $\log_2(24/16) = 0.585$ bits (KG-DES, DDT), $\log_2(32/24) = 0.415$ (KG-DES, Walsh), $\log_2(46/18) = 1.354$ (BEA-1, DDT) and $\log_2(256/164) = 0.642$ (BEA-1, Walsh). The partition-preserving linear layer lets a characteristic that starts in the block stay in it at every step. This suggests a search strategy that is much cheaper than an unrestricted one. Search the block first, over 3 of the 63 nonzero S-box input differences of KG-DES and 31 of the 1023 of BEA-1, and then certify. Where the block optimum lies within the exit penalty of the active-S-box bound, as for BEA-1 over up to three rounds, the bound alone certifies it. At four rounds it still certifies that the optimum lies in the

block (Proposition 30). Otherwise the block optimum serves as the starting bound of an unrestricted branch-and-bound, which then discards almost every branch that leaves the block. For KG-DES the unrestricted and restricted searches of Section 6.6.4 returned the same optima, as this predicts. In addition, for BEA-1 the branch number gives a second saving. Within a layer the two MixColumns groups are independent, so the search factors column by column. Started at the weight of a known characteristic, the block search settles the four-round case in about ten minutes (Proposition 30).

8.2 A MILP Cross-Check of the Trail Results

We found the trails of this section with searches written for each question. These were Matsui's branch and bound, block searches in the trapdoor coordinates, dynamic programming over activity patterns and exhaustive table computations. As an independent check, we encoded every trail type once more as a mixed-integer linear program, in the style of Mouha et al. [MWGP12], and solved each model with SCIP 10 [SCI] under a limit of 600 seconds. The models share the cipher definitions with the searches they check (the tables of E and P, MixColumns and ShiftRows, and the trapdoor bases) but no search code, and none with the periodic MILP of Section 6.6.3.

The models. We solved two kinds of model. *Branch-number models* give each S-box a binary activity variable. For BEA-1, MixColumns has bundle branch number 5, encoded per column as $\sum \text{in} + \sum \text{out} \geq 5d$ with every activity at most d. For KG-DES they use a truncated bit-level model of E, P and the Feistel XOR. The minimum number of active S-boxes, times the best weight of one S-box, bounds every trail. *Exact models* work inside the trapdoor block, in reduced coordinates. For KG-DES these are differences in $\mathbf{S}$ (2 bits per nibble), S-box inputs in $\mathbf{U}$ and outputs in $\mathbf{W}_i$. Similarly, for BEA-1 they are differences in $\mathbf{V}_i$ and $\mathbf{W}_i$ and masks in $\mathbf{V}_i^{\perp}$ and $\mathbf{W}_i^{\perp}$ (5 bits per bundle). Each S-box is a one-hot choice among its possible transitions, weighted by $-\log_2$ of its table entry. The linear layers become XOR constraints in the reduced coordinates, and the branch-number constraints are added as cuts. The same machinery gives the hybrid trails. A differential-linear trail is a truncated differential in the block followed by a DLCT middle layer at its bound and a linear characteristic, of weight $2w_1$ for a characteristic of weight w_1. A boomerang is a characteristic, a switch and a characteristic, of weight $2w_0 + w_{\text{sw}} + 2w_1$. The switch uses the BCT for BEA-1 and the FBCT for KG-DES, and an S-box active in only one of the two trails is free (the ladder switch).

The run. The exact KG-DES models that reached optimality took at most 463 seconds (differential, 12 rounds) and 328 seconds (linear, 9 rounds), and all eleven branch-number models of BEA-1 were solved to optimality. Two models, the KG-DES differential models over 13 and 16 rounds, stopped early, after about 171 and 131 seconds of computation. We solved them again with the same limit, and the table reports the reruns. The appendix explains what happened.

What the cross-check confirms. Table 48 lists the results. Wherever SCIP proved optimality, it reproduces the paper's values exactly. For KG-DES these are the optimal differential weights B_r for $r \leq 13$ and the optimal linear weights B_r^{L} for $r \leq 9$. For BEA-1 they are the active-S-box counts $A(L)$ for all 11 layers, the optimal characteristics over one and two layers, and the optimal linear characteristics over up to three. The cost of leaving the block certifies these two characteristics as global optima. Where SCIP hit the time limit, its values bracket the paper's and never contradict them. For KG-DES over 14, 15 and 16 rounds, the trail SCIP found has exactly the optimal weight of Table 23, although SCIP could not prove it optimal. Similarly, for BEA-1 over three and four layers, the dual

Table 48: The MILP cross-check (SCIP, 600 s per model). Weights are $-\log_2$ of a probability or of an absolute correlation; "limit" marks a model stopped at the time limit, whose value is the best trail found and, where given, the proven dual bound. The last column is the value established in this paper.

Model	Rounds / layers	MILP result (weight in bits)	This paper
KG-DES			
Min. active, differential	1–10	$0, 1, 2, 3, 4, 6$; 7 at 8; 9 at 10 (optimal)	–
	$12, 14, 16$	$\geq 7, 6, 5$ proven; $12, 13, 19$ found (limit)	–
Min. active, linear	1–12	$1, 1, 2, 2, 3, 4$; $5, 6, 8$ at $8, 10, 12$ (optimal)	–
	$14, 16$	$\geq 5, 5$ proven; $9, 11$ found (limit)	–
Differential, in the block	1–13	0, 1.415, 2.830, 4.245, 5.660, 8.490, 8.490, 9.905, 11.320, 12.735, 14.150, 16.980, 16.980 (optimal; 13 rounds rerun)	B_r, identical (Table 23)
	14, 15	18.395, 19.811 (limit)	B_r, identical
	16	21.226, dual bound 9.559 (limit; rerun)	$B_{16} = 21.23$, identical
Linear, in the block	1–9	$1, 1, 2, 3, 4, 6, 6, 8, 8$ (optimal)	B_r^{L}, identical
	10–16	$11, 31, 71, 54, 106, 110, 133$ found (limit)	B_r^{L}: 10, 10, 12, 12, 14, 14, 16 (Table 30)
Differential-linear, 16 rounds	E_1 over $0, 1, 2, 4, 8$	$0, 2, 2, 6, 16$ (optimal)	$2B^{\mathrm{L}}$ of E_1
Boomerang, FBCT switch	$4, 8, 12, 16$	5.660, 14.150, 25.471, 36.791 (optimal), with $0, 2, 2, 4$ rounds before the switch round and $3, 5, 9, 11$ after	probability 1
BEA-1, secret S-boxes			
Min. active, $A(L)$	1–11	$1, 5, 8, 15, 16, 20, 23, 30, 31, 35, 38$ (optimal)	identical
Differential, in the block	1, 2	4.476, 22.382 (optimal); leaving the block costs at least 5.830, 23.736, so both are global optima	identical
	3	36.263, dual bound 35.812 (limit); leaving the block ≥ 37.165	36.068; bound 35.812
	4	69.036, dual bound 67.147 (limit); leaving the block ≥ 68.500	67.672; bound 67.147
Linear, in the block	$1, 2, 3$	$2, 10, 16$ (optimal)	$2A(L)$, identical
	$4, 6, 8, 11$	32, 60 found at 4, 6, none at 8, 11; dual bounds $30, 40, 60, 76$ (limit)	$2A(L)$; 76 at 11
Differential-linear	E_1 over $0, 1, 2, 3$	$0, 4, 20, 32$ (optimal)	2× linear
Boomerang, BCT switch	$2, 3, 4$	8.953 (optimal), 27.341, 83.894 (limit), switch in layer 0, 1, 3	measured 0, 18.8, > 26.4

bounds 35.812 and 67.147 are the lower bounds of Proposition 30, and the characteristics found lie above the optima 36.068 and 67.672. The linear dual bounds are the active-S-box bounds $2A(L)$, which the 11-layer characteristic of Table 45 attains. The KG-DES linear models are the hardest. Beyond nine rounds the solver's incumbents, 11 to 133, stay far above B_r^{L}, and the optimality of 2^{-16} rests on Matsui's search (Proposition 25).

What it adds. The differential-linear optima are twice the linear optima of the last part, $2B^{\mathrm{L}}$ for KG-DES and $2 \times (2, 10, 16)$ for BEA-1. Within this family of trails, correlation 1 is reached only with the middle layer last, as in Proposition 26. The boomerang optima quantify the remark of Section 7.6.2 that the characteristic-based estimate is far too pessimistic. For KG-DES the best such estimate over 16 rounds is $2^{-36.79}$, worse than the best differential, while the true return probability is 1 (Corollary 4). The coset-affine structure is invisible to single characteristics. Similarly, for BEA-1 the best estimates found, $2^{-8.95}$ over two layers (optimal) and $2^{-27.34}$ over three, compare with measured probabilities of 1 and $2^{-18.8}$.

8.3 Summary: The Best Trail of Every Type

Table 49 collects the best trail of each type for both ciphers, with its status. "Optimal" means that an exhaustive search or a matching bound proves that no trail of that type does better. "Bound" means that only a bound is established, an upper bound for the 11-round BEA-1 characteristic and a lower bound for the linear approximations. Each trail is specified in full in the section listed, down to the difference or mask at every active S-box. The MILP of Section 8.2 confirms every entry it can solve to optimality. The corresponding round counts for LowMC-M, whose backdoor is of another kind, are in Table 54.

Cost of the attacks. Table 50 turns the trails into distinguishers and states their data, time and memory. The partition distinguishers use the partition, which Algorithm 1 recovers from the S-box tables, and for BEA-1 these are the secret tables. At $N = 2^{17.5}$ known plaintexts for KG-DES, or $2^{21.5}$ for BEA-1, about four pairs share a coset, so at least one does with probability $1 - e^{-4} \approx 0.98$; a random permutation puts such a pair's ciphertexts in one coset with probability only 2^{-32} or 2^{-40}. Finding the pairs means reducing each text modulo the subspace and detecting a repeat. A hash table does this in time and memory linear in N. The differential estimate uses the same criterion of four expected right pairs. The attacks on KG-DES and BEA-1 in the table are distinguishers. The key recoveries discussed in this paper are the two-stage recovery on KG-DES (Section 6.7) and Bannier's statistical attack on the deployed BEA-1, from 2^{16} chosen plaintexts (Section 7.1). The third is that of the LowMC-M backdoor (Proposition 35), which took 35 to 409 chosen-tweak queries in the seven full-size measurements of Table 55. Its two rows close the table.

Tables 51 and 52 record how the results on BEA-1 and the trail results of both ciphers were verified.

Table 49: The best trail of every type for KG-DES and BEA-1 (secret S-boxes), with DES as a reference. Differential entries are probabilities of characteristics, linear, type-II and differential-linear entries absolute correlations, boomerang entries the probability that a quartet returns (Sections 6.9.2 and 7.6.2). A linear approximation sums every characteristic with the same outer masks; its bound holds for some output mask of the trapdoor block, which depends on the key. The partition entries give the known plaintexts needed for one pair in a common coset.

Type	Cipher, rounds	Best	Status	Specified in
Differential	KG-DES, 16	$2^{-21.23}$	optimal	Table 24
	KG-DES, per round	$2^{-1.415}$ (period 12)	optimal	Proposition 22
	DES, 16 (reference)	$2^{-61.97}$	optimal	Table 23
	BEA-1, 3	$2^{-36.07}$	optimal	Table 39
	BEA-1, 4	$2^{-67.67}$	optimal	Proposition 30
	BEA-1, 11	$\leq 2^{-170.1}$	bound	Proposition 30
Linear characteristic	KG-DES, 16	2^{-16}	optimal	Proposition 25, Figure 4
	DES, 16 (reference)	$2^{-22.42}$	optimal	Table 30
	BEA-1, 11	2^{-76}	optimal	Table 45
Type-II round entry	KG-DES, 1	2^{-2}, key-dependent sign	exact, one trail	Proposition 1
	BEA-1, 1	0	exact	Proposition 1
Linear approximation (hull)	KG-DES, 16	$\geq 2^{-16}$, every key	bound	Corollary 2
	BEA-1, 11	$\geq 2^{-20}$, every key	bound	Corollary 2
Known-plaintext partition	KG-DES, 16	$2^{16.5}$ texts	measured	Table 34
	BEA-1, 11	$2^{20.5}$ texts	measured	Table 46
Differential-linear	KG-DES, any	1	optimal	Table 32, Figure 5
	BEA-1, 11	1	optimal	Table 42, Figure 8
Boomerang	KG-DES, 16	1	optimal	Corollary 4
	BEA-1, 2 / 3 / 4	1 / $2^{-18.8}$ / $< 2^{-26.4}$	measured	Table 44
S-box tables	KG-DES	DLCT 32/32 at 13 pairs, FBCT 64/64 at 6	at bound	Propositions 18, 19, 21
	BEA-1	DLCT 512/512 at 961 pairs; BCT ≤ 104, FBCT ≤ 88	DLCT at bound	Proposition 31, Table 41

Table 50: Data, time and memory of the attacks on the full ciphers (BEA-1 with the secret S-boxes; LowMC-M-128/3 with its backdoor: a full-round truncated differential, and the key recovery, with at most 64 trial encryptions after the guesses; all seven parameter sets in Table 55). CP: chosen plaintexts; CT: chosen-tweak, chosen-plaintext queries; ACC: adaptively chosen ciphertexts; KP: known plaintexts; N: number of texts. The differential-linear trails of Propositions 26 and 33 give the same one-pair distinguisher as the invariant subspace. The linear distinguisher's data is c^{-2} for $c = 2^{-16}$ and varies with the key through the linear hull (Table 31); the others hold for every key.

Distinguisher	Cipher	Data	Time	Memory	Advantage
Invariant subspace, Cor. 3	KG-DES	2 CP	2 queries	$O(1)$	$1 - 2^{-32}$
	BEA-1	2 CP	2 queries	$O(1)$	$1 - 2^{-40}$
Boomerang, Cor. 4	KG-DES	2 CP, 2 ACC	4 queries	$O(1)$	$\approx 1 - 2^{-64}$
Partition, Table 34	KG-DES	$2^{17.5}$ KP	$O(N)$	$O(N)$	≈ 0.98
Partition, Table 46	BEA-1	$2^{21.5}$ KP	$O(N)$	$O(N)$	≈ 0.98
Linear, Prop. 25	KG-DES	$\approx 2^{32}$ KP	2^{32}	$O(1)$	typical key
Differential, Prop. 22	KG-DES	$\approx 2^{24.2}$ CP	$2^{24.2}$	$O(1)$	≈ 0.98
Related-tweak, Sec. 9.5	LowMC-M	2 CT	2 queries	$O(1)$	$1 - 2^{-126}$
Key recovery, Prop. 35	LowMC-M	201 CT	$2^{9.6}$ guesses	$O(k^2)$	1, measured

Table 51: Summary of independent computational verification for the trapdoor of BEA-1 (Section 7) and for the characteristics, connectivity tables and boomerangs of both ciphers; the corresponding summary for the algorithm and KG-DES is Table 35.

Result	Verification method	Outcome
Boomerangs (Table 44, Corollary 4)	Direct measurement with random round keys: 2^{20} to 2^{28} quartets per placement; 4,000 quartets on 16-round KG-DES with a DES-S-box control	BEA-1: 1, $2^{-18.8}$, $< 2^{-26.4}$ at 2, 3, 4 layers; KG-DES: 4,000/4,000
DLCT/FBCT/BCT uniformity (Table 41, Proposition 32)	Exhaustive computation over all nonzero $(\Delta, \gamma)/(\Delta, \nabla)$ pairs, secret and public S-boxes, in two independent implementations	DLCT exact at 512/512 for all 4 secret boxes; FBCT, BCT well below bound; FBCT 0 on $\mathbf{V}_i \times \mathbf{V}_i$
Optimal differential characteristics (Proposition 30)	Branch number and minimum active S-boxes by exact support counting and dynamic programming; branch-and-bound inside the trapdoor block; bound for characteristics leaving it	Global optima for 1–4 layers ($2^{-36.068}$ at 3, $2^{-67.672}$ at 4)
Round-level candidate subspace (Proposition 28)	Direct construction (trivial preimage under identity selection)	Confirmed; $\mathbf{T}_0$ is the subspace V of [Ban17]
ShiftRows closure (Lemma 6)	Enumeration of all 8 bundle positions against σ	Confirmed, 0 violations
MixColumns closure (Proposition 29)	All 20 basis vectors of $\mathbf{W}_{04}$ mapped through M (the second group is identical)	Exact containment, 20/20
Full-round, full-cipher invariant (Theorem 3)	Direct simulation of 10,000 random triples (K, P_1, P_2) with $P_1 \oplus P_2 \in \mathbf{T}_0$ (secret S-boxes, full 11-round cipher, every layer checked; Example 6)	0 failures
Contrast with deployed cipher	200-trial simulation against the modified (public) S-boxes	0 successes, consistent with the trivial subspaces of Table 36

Table 52: Summary of independent computational verification for the differential-linear, linear and correlation-matrix results of both ciphers (Sections 6–8).

Result	Verification method	Outcome
Differential-linear trails (Propositions 26, 33)	Every layer checked exhaustively; full ciphers measured with random keys against trapdoor-free controls (Tables 33, 43)	$c_E = 1$ on every pair; controls within two standard deviations of 0
Linear characteristics (Propositions 25, 34)	Matsui's branch and bound, checked against DES; all 2^{12} key values of the round approximation; active-S-box bound and block search for BEA-1; measurement from 2^{30} to 2^{36} plaintexts	2^{-16} and 2^{-76}, both optimal; measured within a factor of about two per key (Table 31)
Fourier characterization and correlation blocks (Proposition 4, Theorem 1)	Both zero conditions on all 8 KG-DES and 8 BEA-1 S-boxes and on the DES S-boxes (Figures 6, 10)	Hold for every trapdoored S-box, fail for every control; 0 nonzero entries off the blocks
Known-plaintext partition distinguisher (Tables 34, 46)	2^{20} (KG-DES) and 2^{23} (BEA-1) texts per key, 4 keys, with trapdoor-free controls	501/501 and 115/115 pairs; 0 for the controls
Type-II round entries (Proposition 1)	All 2^{12} key values entering S_4, S_5; column 0 of all BEA-1 S-boxes; a toy Rijmen–Preneel cipher over all 2^{16} plaintexts	KG-DES $\pm 2^{-2}$ by one trail (DES $\pm 2^{-6}$); BEA-1 0; toy 9/16 for every key
MILP cross-check (Table 48)	SCIP on independent branch-number and exact block models of every trail type, 600 s per model; two models that stopped early were rerun	Every proven optimum equals the paper's value; timeouts bracket it

9 The MALICIOUS Backdoor in LowMC-M

The two trapdoors we have studied so far are partitions, and every tool of this paper finds them from the cipher alone. Algorithm 1 reads them off the S-box tables, the closure of Section 5.5 carries them through the linear layer, and the correlation matrix shows them as an orthogonal block. That raises an obvious question. Can a designer put the trapdoor somewhere these tools do not look? Peyrin and Wang's MALICIOUS framework [PW20] does that. It embeds a set of related-tweak differential characteristics of probability 1 into a tweakable block cipher. Only tweak pairs that the designer chose trigger them, and an extendable-output function (XOF) in the tweak schedule protects the pairs. The S-boxes are generic and the trapdoor sits in the linear layers. LowMC-M instantiates the framework on LowMC [ARS+15].

We treat it as the limiting case, an exact backdoor that the methods of Sections 5–7 cannot see, and against which, as Section 10.2 will show, only provenance helps. The section follows the order of the two previous case studies. We define the cipher, the trapdoor and its exploitation, and set out the published cryptanalysis. We then show that, with an ideal XOF, undetectability holds in the strongest sense and that the planted pair costs about 2^k to recover. We read the design through each tool of this paper, Algorithm 1 among them, and we measure on full-size instances, at the current round numbers, how far every trail type reaches with and without the backdoor. Finally we read the result in the correlation-matrix framework.

Notation for this section. A state $x \in \mathbb{F}_2^n$ is split as $x = (x^{(0)}, x^{(1)})$, with $x^{(0)} \in \mathbb{F}_2^s$ the *non-linear part* (the s least significant bits, as in LowMC) and $x^{(1)} \in \mathbb{F}_2^{n-s}$ the *linear part*. We write $\iota(v) := (v, 0^{n-s})$ for $v \in \mathbb{F}_2^s$, $\tau_c(x) := x \oplus c$ for a translation, and $\mathbf{Z} := \{x : x^{(0)} = 0\}$ for the linear-part subspace. A matrix $L \in \mathbb{F}_2^{n \times n}$ is partitioned into blocks $L^{00} \in \mathbb{F}_2^{s \times s}$, L^{01}, L^{10} and $L^{11} \in \mathbb{F}_2^{(n-s)\times(n-s)}$, so that L^{01} carries the linear part into the non-linear part. The S-box layer has $q = s/3$ S-boxes; the LowMC literature calls this number m, which in this paper is an S-box width. L_i is a linear layer here, not a KG-DES half.

9.1 The Cipher, the Trapdoor, and Its Exploitation

Definition 21 (MALICIOUS tweakable block cipher [PW20])**.** Fix a block size n, a non-linear size $s < n$, a key size k and a round number R. A MALICIOUS cipher is given by

- a partial S-box layer $S(x) = (\sigma_s(x^{(0)}), x^{(1)})$, with σ_s a layer of parallel S-boxes on $\mathbb{F}_2^s$;
- linear layers $L_1, \ldots, L_R \in \mathrm{GL}_n(\mathbb{F}_2)$ and constants $C_1, \ldots, C_R \in \mathbb{F}_2^n$;
- a key schedule $K \mapsto (k_0, k_1, \ldots, k_R)$ with $k_0 \in \mathbb{F}_2^n$ and $k_i \in \mathbb{F}_2^s$;
- a tweak schedule $T \mapsto (t_0, t_1, \ldots, t_R) := \mathrm{XOF}(T)$, with $t_0 \in \mathbb{F}_2^n$ and $t_i \in \mathbb{F}_2^s$.

Encryption of P under key K and tweak T sets $x_0 := P \oplus k_0 \oplus t_0$ and

$$x_i := L_i\, S(x_{i-1}) \oplus C_i \oplus \iota(k_i \oplus t_i), \qquad i = 1, \ldots, R,$$

and outputs $E_{K,T}(P) := x_R$. We write $E^{(\rho)}_{K,T}$ for the first ρ rounds.

Definition 22 (LowMC-M, version 3 [PW20])**.** LowMC-M is the MALICIOUS cipher with $s = 3q$, σ_s made of q copies of the LowMC S-box

$$\sigma(a, b, c) = (a \oplus bc,\ a \oplus b \oplus ac,\ a \oplus b \oplus c \oplus ab),$$

the linear key schedule $k_i = KL_i\,K$ with $KL_0 \in \mathbb{F}_2^{n\times k}$ and $KL_i \in \mathbb{F}_2^{s\times k}$, tweaks $T \in \mathbb{F}_2^k$ with XOF = SHAKE128 (SHAKE256 for $k = 256$), and $R = r_{\text{LowMC}} + \lceil 2k/s \rceil$ rounds, where r_{LowMC} is the round number of LowMC with the same parameters. In an *honest* instance the L_i are independent and uniform on $\mathrm{GL}_n(\mathbb{F}_2)$ and the KL_i, C_i are independent and uniform. The parameter sets attacked in [LSW+22], analyzed in Section 9.5, are $n = k = 128$ with $(s, R) = (3, 294), (6, 147), (9, 99), (30, 32)$ and $n = k = 256$ with $(s, R) = (3, 555), (9, 186), (60, 30)$. The designers also list $(90, 17)$ for $n = 128$ and $(120, 19)$ for $n = 256$.

Definition 23 (Deterministic related-tweak characteristic)**.** For tweaks $T \neq T'$ and a plaintext difference ΔP, the triple $(T, T', \Delta P)$ is a *ρ-round deterministic related-tweak characteristic* if no S-box of rounds $1, \ldots, \rho$ is differentially active for any key and plaintext. Equivalently, there is a Δ_ρ, independent of K and P, with

$$E^{(\rho)}_{K,T'} \circ \tau_{\Delta P} \;=\; \tau_{\Delta_\rho} \circ E^{(\rho)}_{K,T} \qquad \text{for every } K. \tag{8}$$

Lemma 7 (Linear-algebraic form)**.** *Let $\Delta t_i = t_i \oplus t'_i$ and $X_1 = \Delta P^{(1)} \oplus \Delta t_0^{(1)}$, and set $X_{i+1} = L_i^{11} X_i$. Then $(T, T', \Delta P)$ is a ρ-round deterministic characteristic if and only if $\Delta P^{(0)} = \Delta t_0^{(0)}$ and*

$$L_i^{01} X_i = \Delta t_i \qquad \textit{for } i = 1, \ldots, \rho - 1,$$

i.e., $A_{\rho-1} X_1 = (\Delta t_1, \ldots, \Delta t_{\rho-1})$ for the matrix $A_{\rho-1}$ with row blocks $L_i^{01} L_{i-1}^{11} \cdots L_1^{11}$. Consequently, the sub-tweak difference vectors that admit such a characteristic form the linear subspace $\mathrm{Col}(A_{\rho-1}) \subseteq \mathbb{F}_2^{s(\rho-1)}$, of dimension $\alpha = \operatorname{rank} A_{\rho-1} \leq n - s$.

Proof. The S-boxes of round $i + 1$ are inactive exactly when the non-linear part of the difference entering them is 0. The difference entering round 1 is $\Delta P \oplus \Delta t_0$, whose non-linear part vanishes iff $\Delta P^{(0)} = \Delta t_0^{(0)}$. If round i has no active S-box its input difference is $(0, X_i)$, and the difference after the round is $L_i(0, X_i) \oplus \iota(\Delta t_i) = (L_i^{01} X_i \oplus \Delta t_i,\; L_i^{11} X_i)$; keys and constants cancel. Induction on i gives the equivalence, which is the designers' own Equation (3). □

Definition 24 (The malicious backdoor)**.** A backdoored LowMC-M instance is published together with a secret

$$\mathcal{B} = \big\{ (T_1^j, T_2^j, \Delta P^j) : j = 1, \ldots, a \big\}, \qquad a = \lceil k/s \rceil,$$

of distinct, uniformly random tweak pairs such that $(T_1^j, T_2^j, \Delta P^j)$ is an $(R - j)$-round deterministic related-tweak characteristic. It is built by the designers' Algorithms 1 and 2. The pairs and the ΔP^j are fixed first. For $i = 1, \ldots, R - 2$, each row of L_i^{01} is a uniformly random solution x of $M_i x = (\Delta t_i^j[\ell])_j$, where M_i stacks the vectors $X_{i,j}$ of the $b_i \leq a$ characteristics still to be extended, and L_i^{00}, L_i^{10} and L_i^{11} are uniform. The whole draw is repeated until L_i is invertible and the new $X_{i+1,j}$ are linearly independent. The last two matrices, L_{R-1} and L_R, are uniform.

Definition 25 (Attack model)**.** The adversary holding $\mathcal{B}$ has access to $E_{K,\cdot}(\cdot)$ for an unknown key K and may choose both the tweak and the plaintext of every query (chosen-tweak, chosen-plaintext). Its goal is to recover K, and its cost is measured in queries (data) and in elementary operations (time).

The key recovery peels the cipher from the last round. In round i a guess must match an *effective key* instead of the round key k_i. With $c_{R+1} := 0$, we set

$$g_i := \big[L_i^{-1}\big(\iota(k_i) \oplus (0, c_{i+1})\big)\big]^{(0)}, \qquad c_i := \big[L_i^{-1}\big(\iota(k_i) \oplus (0, c_{i+1})\big)\big]^{(1)}.$$

The linear part c_i of the key contribution passes the S-box layer unchanged and is carried into the next round, so every g_i is an $\mathbb{F}_2$-linear function of K.

Proposition 35 (Backdoor key recovery)**.** *Assume that, for every j, each S-box of round $R-j+1$ is active in characteristic j. Then the following algorithm recovers K. For $j=1,\dots,a$, with $i=R-j+1$, query pairs (T_1^j,P), $(T_2^j,P\oplus\Delta P^j)$ and peel rounds $R,\dots,i+1$ with the effective keys already found. Keep the guesses $g\in\mathbb{F}_2^s$ for which inverting round i with g yields the difference that the characteristic predicts, and add pairs until one guess remains. This guess is g_i. Finally, solve the as linear equations $g_i(K)=g_i$ and test the $2^{k-\rho}$ solutions against one known pair, where ρ is the rank of the system. The S-boxes of round i are filtered independently. Each pair therefore costs the peeling of at most $a-1$ rounds and $8q$ guesses of one S-box each ($16q$ S-box inversions), not 2^s guesses. After the last characteristic come one linear system and at most $2^{k-\rho}$ trial encryptions. The data is two queries per filtering pair. For $q=1$ the expected number of pairs per characteristic is $7/3$, and it grows slowly with q.*

Proof. With the effective keys, peeling a round from a ciphertext gives the state up to a translation of its linear part by c_i. The translation is the same for both members of a pair, so the predicted differences are unaffected. The correct g_i always survives. A wrong guess g survives a pair only if, on every active S-box, $g\oplus g_i$ swaps the two S-box outputs of that pair. Whether it does depends on the pair, and with all S-boxes active, independent pairs eliminate every wrong guess. Since σ is APN, a pair leaves two candidates on each S-box, g_i and $g_i\oplus\beta$, where β is that S-box's output difference. By Lemma 8, β is uniform on four values. Each further pair therefore removes the wrong candidate with probability $3/4$, which gives $1+4/3$ pairs on average for one S-box. The linearity of g_i in K then gives the linear system. □

Remark 39 (The designers' accounting). The designers state that each characteristic reveals the s-bit round key k_i, for data $2(a+1)$ and time $2(a+1)+2^{\max(k-as,0)}$. We refine this in four points, none of which changes the practicality. (i) One pair per characteristic does not suffice. An APN S-box leaves two candidates per pair, and the data is about $2\cdot\frac{7}{3}\,a$ for $q=1$, i.e., 201 queries for LowMC-M-128/3 instead of 88 (Table 55). (ii) The round key k_i itself is determined only when the $s\times s$ block $(L_i^{-1})^{00}$ is invertible. For a uniform L_i this fails with probability about 0.71 (0.67 for $s=3$). The effective keys restore the count of s linear equations per characteristic. (iii) The as equations in $k\le as$ unknowns need not have rank k. A random square system over $\mathbb{F}_2$ is singular with probability about 0.71, and this adds a factor $2^{k-\rho}$, between 1 and 64 on the instances of Table 55. (iv) An S-box that is inactive in the attacked round leaves its bits undetermined. The designer can exclude this by redrawing either the secret choices or the last fitted matrix until every S-box of that round is active. The difference entering the round contains the sub-tweak difference Δt^j_{R-j}, which is uniform and enters no other constraint. The event therefore has the same probability for every choice of the published matrices, and the redraw does not bias them.

The attack is *traceable*. The queries carry the malicious tweaks, and, as the designers note, a user who records them can test pairs for a deterministic characteristic.

9.2 Relation to the Published Cryptanalysis

Three papers attack LowMC-M (Table 53). All three attack the cipher from the outside, without the backdoor. They use the tweak input to find *weak* tweak pairs, whose probability-1 related-tweak characteristics are shorter than the embedded ones, and put a key-recovery phase on top. None recovers an embedded pair, and none is practical.

The first version was thus broken at full rounds, faster than exhaustive search but far from practically. The designers answered with $\lceil 2k/s\rceil$ extra rounds and then k-bit tweaks, and the best attacks on the current round numbers stop between 4 and 17 rounds short. The rest of this section asks a different question. What is the backdoor, and what is it worth? It shares two ingredients with the published attacks. The first is Lemma 7,

Table 53: Published attacks on LowMC-M. None uses or recovers the embedded backdoor.

Work	Version attacked	Method	Best result for $n = k = 128$, $s = 3$
Beyne, Li 2020 [BL20]	v1, 208 rounds	weak pairs by collision search, probability 1 over $\lfloor (2c+n)/s \rfloor$ rounds for 2^c work; differential-linear distinguisher, difference enumeration	full rounds: one-time 2^{96} search for a weak pair, then $2^{96.8}$ time and 4 chosen plaintexts per key
Liu, Isobe, Meier, CRYPTO 2021 [LIM21]	v1, 208 rounds	the same weak pairs; difference enumeration with algebraic key recovery, negligible memory, the allowed 2^{64} data	$250 > 208$ rounds in time 2^{120} with 2^{64} data; all seven listed instances broken
Liu, Sarkar, Wang, Meier, Isobe, ASIACRYPT 2022 [LSW+22]	current round numbers, 294 rounds	algebraic meet-in-the-middle; key recovery from a trail in near-constant time	277 of 294 rounds in time $2^{124.4}$, memory $2^{111.6}$, data 2^{64}; margins of 4 to 17 rounds over all listed instances

which is the designers' own Equation (3) and underlies all three attacks. The second is the weak-pair cost after Proposition 36, which is Beyne and Li's. Several of our results are not in them. These are the undetectability theorem, the golden-collision cost of the planted pair with its dependence on the tweak length, and the effective-key accounting. They also include the correlation-matrix and translation-conjugacy statements, the S-box lemma, the full-size bounds below, and the assessment of provenance, chosen tweaks and traceability.

9.3 Undetectability and Undiscoverability

The designers argue undetectability informally, in two ways. The first is a rank test (their Proposition 1). The second defines the non-backdoored cipher by the same generation algorithm with random sub-tweak differences, which makes the two indistinguishable by definition. The comparison that matters, however, is with an honest instance, whose linear layers are uniform, and there too the two coincide up to a negligible distance.

Theorem 4 (Backdoored instances are distributed as honest ones)**.** *Model the sub-tweak differences Δt_i^j of distinct malicious pairs as uniform and independent. Then the public description of a backdoored LowMC-M instance and that of an honest one have statistical distance at most*

$$\varepsilon \;=\; \sum_{i=1}^{R-2} \left(2^{\,b_i-n+1} + 2^{\,b_i-(n-s)}\right) \;\le\; 2R \cdot 2^{\,a-(n-s)}.$$

With SHAKE in place of the ideal model, any algorithm that tells backdoored from honest instances with advantage δ distinguishes SHAKE outputs on secret uniform inputs from uniform strings with advantage at least $\delta - \varepsilon$.

Proof. Only the L_i with $i \le R-2$ are generated differently. The KL_i, C_i, L_{R-1} and L_R have the honest distribution. We compare round by round, conditioning on the history, in which the vectors $X_{i,1}, \ldots, X_{i,b}$ $(b = b_i)$ are linearly independent. For a uniform $L_i \in \mathrm{GL}_n(\mathbb{F}_2)$, the images $L_i(0, X_{i,j})$ form a uniform linearly independent b-tuple of $\mathbb{F}_2^n$. Its statistical distance from a uniform b-tuple is the probability that b uniform vectors are dependent, which is at most 2^{b-n}. Hence $Y := (L_i^{01} X_{i,j})_j$ is within 2^{b-n} of uniform on $\mathbb{F}_2^{s \times b}$. In the backdoored generation the right-hand sides Δt_i are uniform. Given them, L_i is uniform on the invertible matrices with $L_i^{01} X_{i,j} = \Delta t_i^j$ (uniform solutions for the rows of L_i^{01}, uniform other blocks, rejection until invertible), and this is the honest conditional law

given $Y = \Delta t_i$. The joint laws therefore differ by at most 2^{b-n}. The extra rejection until the $X_{i+1,j}$ are independent occurs with probability at most $2^{b-(n-s)} + 2^{b-n}$. Indeed, the $X_{i+1,j}$ are the image of the $X_{i,j}$ under the fixed linear map L_i^{11}, so they are within 2^{b-n} of a uniform independent tuple, and b uniform vectors of $\mathbb{F}_2^{n-s}$ are dependent with probability at most $2^{b-(n-s)}$. Conditioning on this event moves the law by at most that probability. The two contributions per round, 2^{b-n} and $2^{b-(n-s)} + 2^{b-n}$, sum over the rounds to ε. Since $2^{b-n} \le 2^{b-(n-s)}$, this is at most $2R \cdot 2^{a-(n-s)}$. For the second statement, consider a distinguisher between the two instance distributions. Run on instances generated from given sub-tweak differences, it distinguishes those differences from uniform, up to ε. □

For LowMC-M-128/3 ($n - s = 125$, $a = 43$, $R = 294$) the bound is below 2^{-72}, and for LowMC-M-256/120 ($n - s = 136$, $a = 3$) it is below 2^{-126}. The weakest listed set is LowMC-M-128/90, with $n - s = 38$ and a bound of about 2^{-31} on the advantage of any test on a single instance.

Corollary 5 (What does leak)**.** *(i) Reusing a tweak pair for several characteristics makes the Δt_i dependent, and the theorem no longer applies. This is the rank deficiency of the designers' Proposition 1, which the toy of Section 9.7 reproduces (*$\operatorname{rank} A_7 = 16$ *instead of* 18*). (ii) The theorem compares with honest instances whose matrices have no verifiable origin. If an instance must come with a public derivation of its matrices from a seed, the rows of L_i^{01} cannot depend on secret data, and the backdoor cannot be planted. Conversely, an instance without such a derivation is what MALICIOUS requires. (iii) The backdoor leaks when used, since the attack is traceable.*

In the terms of Table 6, MALICIOUS realizes the limiting case. From the artefact alone an evaluator cannot establish even the presence of asymmetric hidden structure, since the backdoored cipher is distributed as an honest one. The partition trapdoors, by contrast, leave exact fingerprints in every table. They give zero blocks in the correlation matrix, a DLCT at its bound on a product of subspaces, and a probability-1 truncated differential that Algorithm 1 finds in seconds.

Proposition 36 (The embedded pair is a golden collision)**.** *Let π be the quotient map $\mathbb{F}_2^{s(\rho-1)} \to \mathbb{F}_2^{s(\rho-1)}/\mathrm{Col}(A_{\rho-1})$ and, for $T \in \mathbb{F}_2^k$ with $\mathrm{XOF}(T) = (t_0, \dots, t_R)$, let $f(T) := \pi(t_1, \dots, t_{\rho-1})$. Then (T, T') triggers a ρ-round deterministic characteristic iff $f(T) = f(T')$. For the embedded characteristics of LowMC-M version 3, $s(\rho-1) - \alpha$ far exceeds $2k$. The malicious pair is therefore, with overwhelming probability, the only collision of f on $\mathbb{F}_2^k$. It is found with 2^k evaluations of f and memory 2^k by sorting, or in time about $2^{3k/2} w^{-1/2}$ with memory w by parallel collision search [vOW99].*

Proof. By Lemma 7, a characteristic exists iff the sub-tweak difference vector lies in $\mathrm{Col}(A_{\rho-1}) = \ker \pi$, iff $f(T) = f(T')$. A compatible ΔP then exists, obtained by solving $A_{\rho-1} X_1 = (\Delta t_1, \dots, \Delta t_{\rho-1})$ for X_1. It is unique iff $A_{\rho-1}$ has full column rank, and one such ΔP is all the search needs. The matrix $A_{\rho-1}$ is public. The number of pairs is 2^{2k}, and each collides with probability $2^{-(s(\rho-1)-\alpha)}$, so the expected number of accidental collisions is negligible. Exhaustive evaluation and sorting, or golden-collision search, find the planted one. □

This confirms the designers' claim that recovering the backdoor costs as much as a key search, and it locates the reason. In version 3 the binding constraint is the k-bit tweak domain, not the target-difference resistance of the XOF, whose birthday bound on $s(\rho-1) - \alpha$ bits is far above 2^k. The XOF is indispensable for another reason, the pseudorandomness of its outputs, on which Theorem 4 rests. Shorter characteristics behave differently. For ρ up to about $(2k + n)/s$ rounds, f has many collisions among 2^{2k} pairs, and parallel collision search finds one in time $2^{(s\rho-n)/2} < 2^k$. Beyne and Li [BL20] used such weak pairs, extended by differential-linear techniques, for key-recovery attacks on

most full-round instances of the first version. Together with algebraic attacks by Liu, Isobe and Meier [LIM21], this led to $\lceil 2k/s \rceil$ extra rounds in version 2. The k-bit tweaks of version 3 answer a limited-birthday argument [IPS13] of the designers' own. The embedded pair itself was not recovered.

9.4 The Chapter's Tools Applied to LowMC-M

Invariant subspaces and partitions. The backdoor is not an invariant structure of any single permutation. By Equation (8) it is a relation between two keyed permutations. Over the backdoored rounds, E_{K,T_2} is E_{K,T_1} conjugated by translations. The tools of this paper that look for a subspace or partition preserved by one cipher (Algorithm 1, the spectral detector of Section 4.3 and the round-level closure of Algorithm 3) therefore have nothing to find. However, the closure computation does apply in its affine form. Propagating the linear-part subspace $\mathbf{Z}$ with the sub-tweak differences as a parameter produces the matrix $A_{\rho-1}$ of Lemma 7, the starting point of the weak-pair attacks.

Differential cryptanalysis. The backdoor is a *point* differential of probability 1 over up to $R-1$ rounds. This is the opposite of BEA-1, where the partition gives a probability-1 truncated differential but the best point characteristic over three rounds has probability $2^{-36.07}$ (Proposition 30). MALICIOUS obtains a point differential by moving the trapdoor out of the S-boxes and into the linear layers. The S-boxes stay generic, and the linear layers absorb the sub-tweak differences.

Linear cryptanalysis and correlation matrices.

Proposition 37 (The two tweaks have the same correlation magnitudes)**.** *Over the backdoored rounds,*

$$C^{E^{(\rho)}_{K,T_2}} \;=\; D_{\Delta_\rho}\, C^{E^{(\rho)}_{K,T_1}}\, D_{\Delta P}, \qquad D_c := \operatorname{diag}\big((-1)^{c\cdot u}\big)_u,$$

so every linear approximation has the same absolute correlation under T_1 and under T_2, with signs flipped by $(-1)^{\lambda\cdot\Delta P\oplus\gamma\cdot\Delta_\rho}$.

Proof. Equation (8) reads $E_2 = \tau_{\Delta_\rho} \circ E_1 \circ \tau_{\Delta P}$. Correlation matrices are multiplicative, and a translation τ_c has the diagonal correlation matrix D_c. □

Nothing distinguishes the correlation matrix of a single tweak, since the backdoor is a relation between two of them. This is the sharpest contrast with the other trapdoors of this paper, which are visible in every single correlation matrix. Rijmen and Preneel's appears as one large entry in column 0, and the partition trapdoors as an orthogonal block with zeros around it (Theorem 1). The correlation-1 linear trails of LowMC-M, with zero mask on the non-linear part over about $(n-1)/s$ rounds, exist in every partial-S-box design. They do not involve the backdoor, because the sub-tweaks enter a linear approximation only through a sign.

Differential-linear and boomerang attacks. Over the backdoored rounds, the related-tweak differential-linear correlation $2^{-n}\sum_x (-1)^{\lambda\cdot(E_{K,T_1}(x)\oplus E_{K,T_2}(x\oplus\Delta P))}$ equals $(-1)^{\lambda\cdot\Delta_\rho}$ for every mask λ. A related-tweak boomerang ($P_2 = P_1 \oplus \Delta P$ under T_2, both ciphertexts shifted by any δ, decrypted under their own tweaks) returns with probability 1, since $E^{-1}_{T_2}(z\oplus\Delta_\rho) = E^{-1}_{T_1}(z)\oplus\Delta P$. This mirrors KG-DES, whose boomerangs return with probability 1 because the cipher is affine on cosets (Corollary 4). Here the reason is that the two permutations are translates of each other.

Rijmen–Preneel type-II relations. Proposition 1 applies with θ_{in} the identity, if we treat the linear part as 1-bit identity S-boxes. Every S-box is balanced, so column 0 of the round function's correlation matrix is trivial. LowMC-M therefore carries no Rijmen–Preneel relation at any level.

S-box-level detectors. The LowMC S-box σ is a quadratic APN permutation, with differential uniformity 2 and maximal $|\widehat{\sigma}| = 4$. The candidate scan of Algorithm 1 returns the empty set. No difference is confined to a proper subspace, and there is no partition trapdoor. The DLCT needs care, and the following lemma describes it fully.

Lemma 8 (Derivatives of the LowMC S-box)**.** *For every nonzero $d \in \mathbb{F}_2^3$, the derivative $D_d\sigma$ takes each value of the affine hyperplane $H_d := \{v \in \mathbb{F}_2^3 : d \cdot v = 1\}$ exactly twice. Consequently, for nonzero d, e, γ: (i) $\mathrm{DDT}(d, v) = 2$ for $v \in H_d$ and 0 otherwise; (ii) $\mathrm{DLCT}(d, \gamma) = -4$ if $\gamma = d$ and 0 otherwise; (iii) $\mathrm{FBCT}(d, e) = 0$ for $e \neq d$; (iv) $\mathrm{BCT}(d, v) = \mathrm{DDT}(d, v)$.*

Proof. Every component $\gamma \cdot \sigma$ is a balanced quadratic function of three variables. Its bilinear form has rank 2, so its radical is $\{0, \Delta_\gamma\}$, and balance forces Δ_γ to be a 1-linear structure, by the argument of Proposition 19. For this S-box $\Delta_\gamma = \gamma$. Since σ is quadratic, $D_d\sigma$ is affine, and since σ is APN, it is two-to-one, so its image is an affine plane. As $d \cdot D_d\sigma(x) = 1$ for every x, the image is the four-element set H_d, which gives (i). For (ii), $\gamma \cdot v$ is constant on H_d when $\gamma \in \{0, d\}$ and balanced otherwise. For (iii), $D_e D_d\sigma$ is a constant, and it vanishes only if $D_d\sigma(x) = D_d\sigma(x \oplus e)$, and this happens only for e in the kernel $\{0, d\}$ of the linear part of $D_d\sigma$. All four tables were also checked exhaustively, including (iv). □

Hence $|\mathrm{DLCT}| = 4 = 2^{m-1}$ at seven pairs, all with sign $-$, and the signed DLCT uniformity of Definition 11 is 0. A detector that reads the signed maximum, as that definition does, raises no alarm. One that reads $|\mathrm{DLCT}|$ would flag every quadratic S-box. The MALICIOUS trapdoor is in the linear layers, where S-box tools are blind by construction.

9.5 Bounds at the Current Round Numbers

In this subsection we take the seven parameter sets and round numbers R of Liu, Sarkar, Wang, Meier and Isobe [LSW+22], from their Table 5, which gives the round numbers of version 3. On one full-size backdoored instance per set, we determine how far each trail type reaches. The instances are generated as in Definition 24, with $a = \lceil k/s \rceil$ characteristics, SHAKE128 or SHAKE256, and the designer's redraw of Remark 39(iv). We compute every structure below exactly, by linear algebra on the instance, and then confirm it by encryption under random keys, tweaks and plaintexts.

Trails of a single tweak. Under a single tweak, keys, constants and sub-tweaks cancel in differences and only change the signs of linear approximations. Let $\mathcal{D}_t$ be the set of plaintext differences that activate no S-box in rounds $1, \ldots, t$; for an end round e, let Λ_u^e be the set of masks on x_{e-u} with a correlation-1 trail through rounds $e-u+1, \ldots, e$ (zero mask on every S-box), and ∇_u^e the set of differences on x_{e-u} reached by decrypting rounds $e, \ldots, e-u+1$ from a difference on x_e without an active S-box. All three are linear subspaces.

Proposition 38 (Deterministic structures of one tweak)**.** *The following hold for every key and tweak.*

(i) A probability-1 differential over rounds $1, \dots, t$ exists iff $\mathcal{D}_t \neq \{0\}$, and a correlation-1 linear trail over rounds $e-u+1, \dots, e$ iff $\Lambda_u^e \neq \{0\}$. Over those rounds the cipher maps the cosets of $(\Lambda_u^e)^\perp$ onto the cosets of a fixed subspace for every key. This is a preserved partition with $2^{\dim \Lambda_u^e}$ blocks.

(ii) Let $\Delta \in \mathcal{D}_t$ have difference δ after round t, and let $\mu \in \Lambda_{r-t-1}^r$ end in the mask λ on x_r. If $\lambda' := L_{t+1}^\top \mu$ satisfies $\lambda'_j \in \{0, \delta_j^{(0)}\}$ on every S-box j, then $\lambda \cdot \big(E^{(r)}(x) \oplus E^{(r)}(x \oplus \Delta)\big)$ is constant. This is a deterministic differential-linear trail over r rounds, or equivalently a probability-1 truncated differential into a hyperplane.

(iii) For $\Delta \in \mathcal{D}_t$ and $\nabla \in \nabla_{r-t}^r$, $\nabla \neq 0$, every r-round boomerang quartet returns.

Proof. (i) is Lemma 7 with $\Delta t_i = 0$ and its dual for masks. For the partition, each mask $\mu \in \Lambda_u^e$ and its end mask μ_e satisfy $\mu_e \cdot x_e = \mu \cdot x_{e-u} \oplus c$ with c independent of the state, so two inputs in one coset of $(\Lambda_u^e)^\perp$ give outputs that agree on every end mask, i.e., lie in one coset of the annihilator of the end masks. The rounds form a permutation and the two cosets have equal size, so the map is onto. For (ii), the correlation-1 trail gives $\lambda \cdot x_r = \mu \cdot x_{t+1} \oplus c$ with c independent of the plaintext, so the parity equals $\lambda' \cdot \big(S(y) \oplus S(y \oplus \delta)\big) = \bigoplus_j \lambda'_j \cdot D_{\delta_j}\sigma(y_j) \oplus \lambda'^{(1)} \cdot \delta^{(1)}$, with y the state after round t. By Lemma 8 each term is constant, equal to 1 when $\lambda'_j = \delta_j \neq 0$. (iii) is the boomerang argument of Section 9.4 with two deterministic halves. □

Each round imposes s linear conditions, and so does the middle round of (ii) when $\lambda'^{(0)}$ is required to be 0 or $\delta^{(0)}$, which is a joint condition on (Δ, μ). Generically, therefore, $\dim \mathcal{D}_t = n - st$, and the longest trails are $r_d, r_l \approx \lfloor (n-1)/s \rfloor$, $r_B \approx 2\lfloor (n-1)/s \rfloor$ and $r_{\mathrm{DL}} \approx \lfloor (2n-1)/s \rfloor$ rounds. The tail can differ. Once a subspace has fewer than s dimensions, the next conditions form a small system that is singular with constant probability, and the trail survives one more round. The counts must therefore be computed per instance. Table 54 gives them; they exceed the generic values by at most three rounds, the case of r_B for LowMC-M-128/3.

Counting estimate. For random linear layers the expected number of r-round characteristics with exactly x active S-boxes is about $\binom{qr}{x} 2^{n-3qr+5x}$, counting 2^n input differences, a factor 2^{-3} for each inactive S-box and four output differences for each active one (Lemma 8). The smallest x for which this reaches 1, $x_{\min}(r)$, is the number of active S-boxes the best characteristic should be expected to have. It then has probability about $2^{-2x_{\min}}$, and the best linear trail has correlation $2^{-x_{\min}}$, since the nonzero Walsh coefficients of σ are ± 4. Over the full R rounds, $x_{\min}$ lies between 98 and 200. Under the 2^{64} data limit, characteristics and trails are usable over at most r_{64} rounds, 129 of 294 for LowMC-M-128/3. This is an estimate, not a bound. It counts characteristics, not differentials, and ignores their clustering, which the published attacks exploit by difference enumeration.

Related tweaks. Weak tweak pairs give probability-1 related-tweak characteristics over $\lfloor (2c+n)/s \rfloor$ rounds for 2^c work (Proposition 36 and the discussion after it), at most $\lfloor (2k+n)/s \rfloor$ rounds at $c = k$. The tail of Proposition 38(ii) adds $r_{\mathrm{DL}} - r_d$ rounds. Even at the cost of an exhaustive key search, these deterministic structures cover between 50 and 62 percent of R. The first phase of Liu et al.'s attacks, 124, 62, 41, 12, 253, 84 and 12 rounds, matches $\lfloor (2c+n)/s \rfloor$ with c between 116 and 122 for $k = 128$, and between 232 and 252 for $k = 256$. Probabilistic difference enumeration and algebraic key recovery then take their attacks to 84–98 percent of R.

Type-II relations and invariants. The round function is a permutation, so column 0 of its correlation matrix is the unit vector, and there is no Rijmen–Preneel relation at any

number of rounds. An invariant subspace or coset partition of the whole cipher would need a perfect approximation over all R rounds, and the correlation-1 trails end after r_l rounds.

Table 54: Rounds covered by each trail type of LowMC-M at the current round numbers R. r_d, r_l, r_{DL}, r_B: longest probability-1 differential, correlation-1 linear trail (and preserved partition), deterministic differential-linear trail and probability-1 boomerang of a single tweak, on the generated instance (Proposition 38). $x_{\min}(R)$: expected fewest active S-boxes over R rounds; r_{64}: rounds within the 2^{64} data limit. Weak pair: $\lfloor (2k+n)/s \rfloor$, then with the differential-linear tail. Published: rounds attacked without the backdoor. The backdoor covers all R rounds in every set.

		single tweak, exact				counting		weak pair, 2^k		published	
n/s	R	r_d	r_l	r_{DL}	r_B	$x_{\min}(R)$	r_{64}	prob. 1	+ tail	[LIM21]	[LSW+22]
128/3	294	43	43	85	87	98	129	128	170	250	277
128/6	147	21	21	42	42	98	64	64	85	125	138
128/9	99	14	14	29	28	99	43	42	57	83	92
128/30	32	4	4	8	8	109	12	12	16	25	27
256/3	555	85	85	171	171	182	177	256	342	496	546
256/9	186	28	28	57	56	183	59	85	114	164	181
256/60	30	4	4	8	8	200	8	12	16	24	26

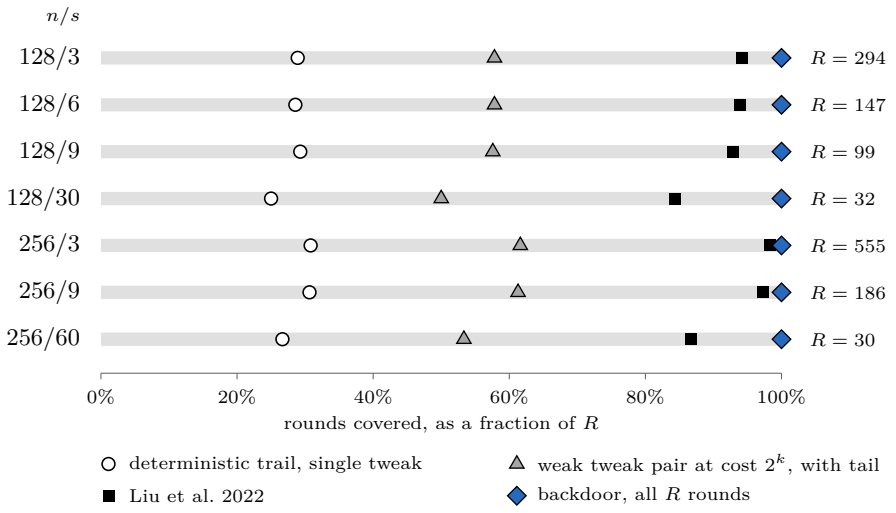

Figure 12: Rounds of LowMC-M covered at the current round numbers, as a fraction of R (Table 54).

The backdoor at full size. On every instance the first embedded characteristic holds with probability 1 over $R-1$ rounds. By Lemma 8 the last round sends its fixed input difference into a known affine space of dimension $2q$, the product of the hyperplanes H_{d_j} mapped by L_R. One pair of queries (T_1, P), $(T_2, P \oplus \Delta P)$ is therefore a full-round related-tweak distinguisher, with false-positive probability 2^{2q-n}. Equivalently, every mask orthogonal to that space, a subspace of dimension $n-2q$, has differential-linear correlation ± 1 over all R rounds. A boomerang whose returning difference $\nabla = L_R(0, z)$ activates no S-box of the last round also returns with probability 1. All three were confirmed on every instance, and the key recovery of Proposition 35 succeeded on every instance (Table 55). The published attacks cost $2^{124.4}$ to $2^{253.3}$ time and 2^{61} to 2^{64} data and stop 4 to 17 rounds short of R.

In these seven measurements the holder of the backdoor needed 35 to 409 queries and seconds of computation for the full cipher.

Table 55: The LowMC-M backdoor at full size, measured. $2q$: dimension of the affine space of the full-round truncated differential (false positives 2^{2q-n}). Queries: chosen-tweak queries of the key recovery, including one known pair for the final test. S-box guesses: 8 per S-box per filtering pair. Equations: the as linear equations in the k key bits and their rank ρ; $2^{k-\rho}$ candidates remain for trial encryption.

n/s	a	$2q$	queries	pairs per char.	S-box guesses	equations, rank	trial enc. ($2^{k-\rho}$)	key
128/3	43	2	201	2.33 (max 7)	800	129, 122	64	found
128/6	22	4	119	2.68 (max 4)	944	132, 125	8	found
128/9	15	6	75	2.47 (max 5)	888	135, 128	1	found
128/30	5	20	35	3.40 (max 4)	1360	150, 128	1	found
256/3	86	2	409	2.37 (max 6)	1632	258, 252	16	found
256/9	29	6	171	2.93 (max 6)	2040	261, 254	4	found
256/60	5	40	39	3.80 (max 4)	3040	300, 256	1	found

9.6 The Backdoor in the Correlation Matrix

Proposition 39 (Correlation matrix of a LowMC-M round)**.** *Write masks as* $u = (u^{(0)}, u^{(1)})$*. The S-box layer has*

$$C^S_{v,u} = \delta_{v^{(1)},u^{(1)}}\, C^{\sigma_s}_{v^{(0)},u^{(0)}},$$

a block-diagonal matrix with one copy of C^{σ_s} *for each* $u^{(1)}$*, and the round* $x \mapsto L_i S(x) \oplus c$ *has correlation matrix* $D_c P_{L_i} C^S$*, where* $(P_L)_{v,u} = [u = L^\top v]$ *permutes the masks and* D_c *is diagonal with entries* ± 1*.*

*(i) The correlation-*1 *trails of Proposition 38 give an orthogonal block of* $C^{E^{(r)}}$*, a signed permutation matrix of size* $2^{\dim \Lambda}$*, with zeros in the rest of its rows and columns. It has* 2^{n-s} *rows after one round, loses a factor* 2^s *with each round, and is trivial after about* $\lfloor (n-1)/s \rfloor + 1$ *rounds.*

(ii) For $X \in \mathbb{F}_2^{n-s}$*,* $D_{(0,X)}$ *commutes with* C^S*, since translations by vectors of* $\mathbf{Z}$ *commute with the S-box layer. A round turns* $\tau_{(0,X_i)}$ *into* $\tau_{L_i(0,X_i)}$*, and the sub-tweak difference returns it to* $\mathbf{Z}$ *iff* $L_i^{01} X_i = \Delta t_i$*.*

Proof. The S-box layer is σ_s on $x^{(0)}$ times the identity on $x^{(1)}$, and the correlation matrix of a product of maps on independent coordinates is the tensor product of theirs; the identity has the identity matrix. A linear map L sends the character of mask v to that of $L^\top v$, and a translation by c multiplies it by $(-1)^{v \cdot c}$. For (i), a row with an entry ± 1 has no other nonzero entry by Parseval, and likewise for columns; the dimension count is that of Proposition 38. For (ii), $D_{(0,X)}$ has entries $(-1)^{u^{(1)} \cdot X}$, constant on each diagonal block of C^S. □

Statement (ii) is the backdoor in geometric form (Figure 14). Over the backdoored rounds the translation between the two tweaks never leaves $\mathbf{Z}$, the diagonal sign matrices telescope, and Proposition 37 follows. Statement (i) is the contrast with the partition trapdoors (Figure 13, beside Figures 6 and 10). In KG-DES and BEA-1 the orthogonal block is the same at every round, so it survives the whole cipher, and it is the trapdoor. In LowMC-M the block is a generic feature of a partial S-box layer. It shrinks by 2^s per round and has disappeared long before R. The backdoor is not a block of any single

correlation matrix. It is an identity between two of them, $C^{T_2} = D_\Delta C^{T_1} D_{\Delta P}$. On a toy with $n = 9$ and $s = 3$ and a planted two-round characteristic, the identity holds entrywise over the two rounds. With the third round, whose S-box is active, the pair has four output differences and the identity fails. On the same toy, the perfect approximations of 1, 2 and 3 rounds number 64, 8 and 1 and form subspaces, as (i) predicts. The toy has no others.

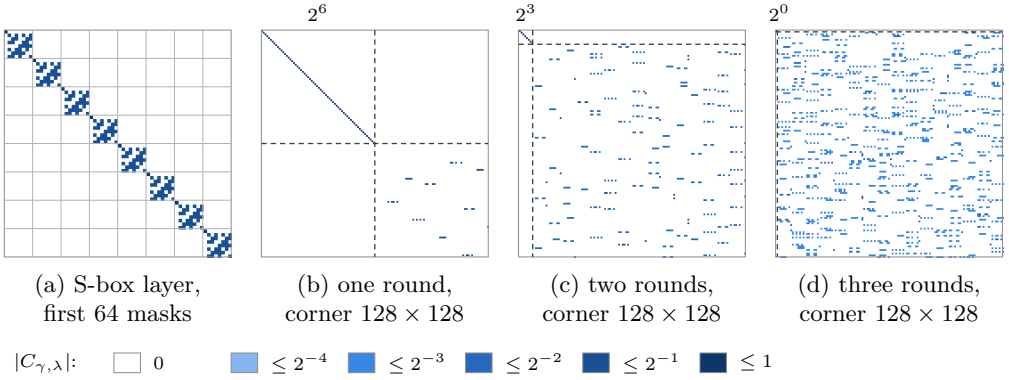

Figure 13: Correlation matrices of a toy LowMC-M with $n = 9$, $s = 3$. (a) The S-box layer, first 64 masks in natural order: one copy of the S-box's 8×8 matrix per value of $u^{(1)}$. (b)–(d) One, two and three rounds, with the masks of the perfect approximations first (rows) and their images first (columns), top-left 128×128 corner. The dashed lines bound the orthogonal block of size 2^6, 2^3 and 2^0; the rest of its rows and columns is 0.

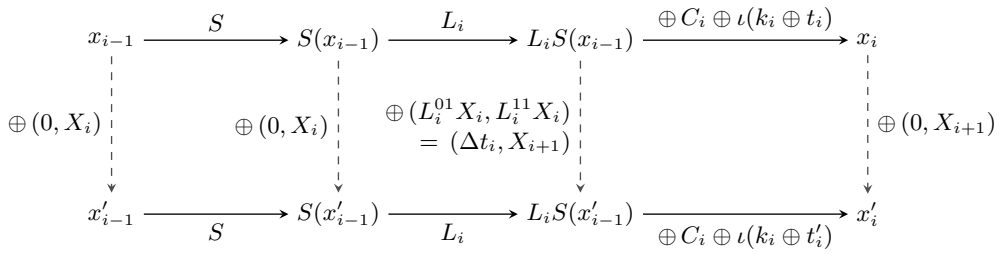

Figure 14: One backdoored LowMC-M round under the tweaks T_1 (top) and T_2 (bottom). The difference stays in $\mathbf{Z} = \{x : x^{(0)} = 0\}$: the S-box layer commutes with the translation, and the sub-tweak difference cancels the non-linear part $L_i^{01} X_i$ that L_i creates.

9.7 Verification

We checked every construction step on a toy LowMC-M built as specified, with $n = 24$, $s = 6$, $k = 24$, $R = 12$ and $a = 4$, SHAKE128 as tweak schedule and the designers' Algorithms 1 and 2. The full-size instances of Section 9.5 and the geometric toy were checked as described there. Table 56 collects the outcomes.

Table 56: Summary of computational verification for Section 9.

Result	Verification method	Outcome
Embedded characteristics (Definition 24)	Toy: 300 random (K, x) per characteristic; full size: every set	Toy: 11, 10, 9, 8 rounds, constant output difference; full size: $R-1$ rounds, constant
Related-tweak boomerang	Toy: 300 quartets; full size: 8 to 24 quartets per set	All return
Key recovery (Proposition 35)	Toy and all seven full-size sets, random keys	Toy: 21 queries, rank 22 of 24, 4 trial encryptions; full size: Table 55; key found every time
Rank test of the designers' Proposition 1	20 honest and 20 backdoored toys; one pair reused three times	rank $A_7 = 18$ for both; 16 with the reused pair
Uniformity (Theorem 4)	First row of L_1^{01}, 64 bins, 3,200 toy instances	$\chi^2 = 53.4$ honest, 43.0 backdoored (mean 63, s.d. 11.2)
LowMC S-box (Lemma 8)	Exhaustive DDT, LAT, DLCT, BCT, FBCT; Algorithm 1	All four statements hold; APN; no candidate subspace
Single-tweak trails (Proposition 38)	Exact linear algebra on each full-size instance, then encryption with random keys, tweaks and plaintexts	All four structures confirmed on all seven sets
Full-round distinguishers	Truncated differential, DL parity and boomerang on the full cipher	Every pair and quartet as predicted
Correlation matrix (Proposition 39)	All 2^{18} entries of the toy's 1-, 2-, 3-round and S-box-layer matrices; planted characteristic under both tweaks	Blocks of size 64, 8, 1; $C^{T_2} = DC^{T_1}D'$ entrywise

9.8 Comparison and Assessment

Of the four constructions compared here, MALICIOUS is the only one whose concealment reduces to a standard primitive, and Theorem 4 shows that its undetectability is stronger than the designers argue, since a backdoored instance is distributed as an honest one. Its limits lie elsewhere.

1. *The security notions.* The designers' second undetectability argument defines the honest cipher by the backdoor's own generation procedure, which makes indistinguishability true by definition. The meaningful comparison is with uniform linear layers, as we make it here.

2. *What undiscoverability rests on.* In version 3 it is the 2^k tweak domain (Proposition 36), not target-difference resistance. The backdoor's security equals the key's, as intended, with no margin.

3. *Exploitability needs tweak control.* The attack requires chosen tweaks equal to the malicious ones, uniformly random k-bit strings. Where the tweak is a counter, a sector number or a nonce, the designer cannot inject them. The backdoor is usable only where the adversary controls the tweak input, and each use is traceable.

4. *Provenance.* An instance whose matrices come with a verifiable derivation from a public seed cannot carry the backdoor. MALICIOUS depends on users accepting unexplained matrices (Corollary 5).

5. *The cipher needed two revisions.* The tweak input gives an outsider the same freedom that the designer uses. Weak tweak pairs for characteristics of about $(2k+n)/s$ rounds are found below 2^k, and full-round attacks on most first-version instances followed [BL20, LIM21]. The fix, $\lceil 2k/s \rceil$ extra rounds and k-bit tweaks, keeps the embedded characteristics beyond reach. The security of the undisturbed cipher now

Table 57: The four constructions side by side. CP and KP: chosen and known plaintexts.

	Rijmen–Preneel	KG-DES (Paterson)	BEA-1 (Bannier–Filiol)	LowMC-M (MALICIOUS)
Hidden object	biased mask β in the S-box	partition preserved by every round	partition, dimension 5 per bundle	malicious tweak pairs; linear layers fitted to their XOF differences
Located in	S-box table	S-boxes, fitted to DES's E, P	S-boxes and MixColumns	linear layers and tweak schedule; S-box generic
Probability	$c_T < 1$	1, truncated	1, truncated	1, point, related-tweak
Correlation matrix	one entry in column 0	orthogonal block, zeros around it	orthogonal block, zeros around it	nothing in a single tweak; $C^{T_2} = D\,C^{T_1}\,D'$
Exploitation	linear cryptanalysis, part of the key	distinguisher from 2 CP; $2^{17.5}$ KP	distinguisher from 2 CP; $2^{21.5}$ KP	full-round distinguisher from 2 chosen-tweak queries; full key from 35 to 409 queries in the seven measurements of Table 55
Detectable from the cipher	yes, by linear algebra [WBDY98]	yes, Algorithm 1	yes, from the secret tables	no (Theorem 4)
Concealment rests on	noise rate (fails)	nothing	nothing	XOF pseudorandomness; 2^k tweak domain
Traceable	no	no	no	yes

rests on a margin that later work has cut to between 4 and 17 rounds [LSW+22], while its own deterministic trails stop near $2n/s$ rounds, 25 to 31 percent of R (Table 54).

6. *Key-recovery accounting.* The designers' accounting is correct in substance, with the refinements of Remark 39.

10 Backdoors Across Primitives

The four constructions we have studied are one region of a larger map. All of them are design-level backdoors. They are planted in the mathematics of a primitive and not in its implementation, and all of them sit in the S-boxes or the linear layers of a symmetric cipher. In this section we place them beside the backdoors found in other primitives, put the detection question for all of them into one table, and say what is still open.

10.1 Beyond Block Ciphers

Shared structural objects. The subspaces and partitions we detect are the objects that ordinary cryptanalysis studies from the other side, and the tools carry over in both directions. An invariant-subspace attack [LAAZ11, LMR15] finds an affine subspace that the round function maps to itself for some keys, and this is what broke PRINTcipher. Subspace-trail cryptanalysis [GRR16] chains such subspaces across rounds and gives distinguishers on round-reduced AES. The nonlinear-invariant attack [TLS16] replaces the subspace by a Boolean function that the round preserves, and breaks full SCREAM and Midori64 for a class of weak keys. A partition backdoor is a preserved structure of the same kind, planted on purpose and holding for every key. Finding it is therefore the same computation as attacking it. The S-box detector of Section 5 and the subspace-trail search of Leander, Tezcan and Wiemer [LTW18] approach one problem from two directions, and the primitivity tests of Caranti et al. [CDVS09] approach it from a third, group-theoretic one.

Key generation in public-key cryptography. The public-key counterpart is kleptography. Young and Yung's SETUP attacks [YY96, YY97] modify a key generator, for RSA, DSA or Diffie–Hellman, so that the public key or a first protocol message carries an encryption of the user's secret under the saboteur's own public key. Everyone else sees an honest key, and an analyst who reverse-engineers the device recovers only the saboteur's public key, never the trapdoor that unlocks it. Young and Yung call this universal protection. A worked demonstration in the same spirit is the trapdoored Diffie–Hellman prime. Fried et al. [FGHT17] showed that a 1024-bit prime can be chosen so that a hidden special structure makes discrete logarithms modulo it feasible for whoever knows the trapdoor. The hidden object is a trapdoor in a number-theoretic problem and not a subspace, but the aim is the same as in the symmetric case, to leak a secret in a way that survives black-box inspection. As with Dual EC, only a verifiable derivation of the public parameters excludes it.

Randomness. The Introduction described Dual EC DRBG. If $Q = dP$ for a secret scalar d, about thirty bytes of output are enough to recover the internal state and predict all later output [SF07]. Checkoway et al. [CFN+14] carried the attack out against TLS libraries, so it was more than a theoretical possibility. Dodis et al. [DGG+15] give the general theory. A backdoored generator is equivalent to public-key encryption, so it cannot exist without a genuine trapdoor, and it can be immunized by post-processing its output. In the terms of Table 58, such a backdoor cannot be detected without the secret and is stopped only by immunization or provenance.

Hashing. A backdoor can also sit in the round constants of a hash function. Albertini et al. [AAE+14] built a "Malicious SHA-1" whose constants differ from the standard ones in about forty bits and admit collisions between meaningful, distinct files. Nobody else's security is affected. The backdoor is the collision that the constants were chosen to allow. Like Dual EC, it is a backdoor in constants. It cannot be seen in the specification, and only a rigid public derivation of the constants excludes it. This is why "nothing-up-my-sleeve" constant generation matters, and why it matters that such generation can itself be manipulated [BCC+15].

Substitution and defences. Another line of work subverts the implementation and leaves the design alone. In an algorithm-substitution attack [BPR14] an encryption routine is replaced by one that behaves identically for the users but leaks the key through a hidden channel. For a randomized scheme this can be done by biasing the ciphertext or the IV, so that a few intercepted messages reveal key bits to whoever knows the bias. Bellare, Paterson and Rogaway also show where the attack stops working, since schemes with unique ciphertexts resist the strongest such attacks. The defences are a field of their own. A cryptographic reverse firewall [MSD15] re-randomizes a party's messages so that a subverted implementation cannot leak through them, and cliptography [RTYZ16] uses a watchdog that tests each component before it is used. These defences protect against a swapped implementation, not against a poisoned specification, so they lie outside the scope of this paper. They do complete Table 58, because they are what catches the kleptographic column.

10.2 What Detection Can Reach

Table 58 puts the detection question for all these constructions in one place. It extends the table of detection techniques in Section 5 (Table 12) beyond the symmetric constructions.

The table is uneven. A statistical battery catches only the crudest backdoor, the biased approximation, and it does so because bias is a deviation from randomness. Structural

Table 58: Which detection technique catches which backdoor family. Y means that the technique catches the backdoor from the specification alone, ∼ that it does so only under a condition, and a dash that it stays silent. For the partition, the differential uniformity of KG-DES's S-boxes gives them away (Remark 31), whereas BEA-1's deployed S-boxes carry no partition to find. Provenance catches any family whose constants can be re-derived. To the right of the partition column, and without provenance, detection reduces to a hard problem: the discrete logarithm for Dual EC, a collision for malicious hashing, and inverting the XOF for MALICIOUS.

Detection technique	RP (biased)	Partition (KG-DES/BEA-1)	MALICIOUS (LowMC-M)	Const. DL (Dual EC)	Mal. hash (SHA-1)	Klepto (SETUP)
Statistical battery	Y	∼	—	—	—	—
Structural search	—	Y	—	—	—	—
Primitivity test	—	Y	—	—	—	—
Correlation-matrix screen	Y	—	—	—	—	—
Provenance / rigidity	∼	∼	Y	Y	Y	∼
Watchdog / firewall	—	—	∼	∼	—	Y

search and primitivity tests catch exact partitions in the S-boxes, and the algorithm reads them straight off the tables, but they go silent as soon as the structure leaves the substitution layer. Every backdoor that lives in unexplained constants, or in a linear layer behind a pseudorandom function, is caught by nothing except a verifiable account of where the constants came from. That covers MALICIOUS, Dual EC and malicious hashing. Kleptographic backdoors in key generation need a watchdog or a reverse firewall instead. No single test covers the whole table.

A recommendation follows. Constants that are derived verifiably from a public seed rule out every backdoor in the right half of the table at once. MALICIOUS depends on users accepting unexplained matrices (Section 9.8), so this is the defence it cannot survive. It is also the defence to which the debates over Dual EC and over the provenance of standard curves have pointed.

10.3 A Research Agenda

The open problems cluster on the boundary that the table draws.

1. *Probabilistic partitions.* Our detector looks for an exact relation. A partition that holds only with high probability, Bannier's probabilistic relaxation [Ban17, Definition 3.11] of Harpes and Massey's partitioning cryptanalysis [HM97], leaves no exact fingerprint, and we know of no test in the left half of Table 58 that would find one. This is the largest gap on the detectable side.

2. *Partial activation.* Every structural test in this paper assumes that the whole S-box layer carries the trapdoor. A backdoor placed on a chosen subset of the S-boxes, or in a partial activation pattern, would escape all of them, and searching over which S-boxes are active is an open problem.

3. *Constant provenance at scale.* Rigidity is the only general defence on the right of the table, yet nothing-up-my-sleeve constants can themselves be rigged, and real designs rarely come with a verifiable derivation. An auditable standard for deriving a cipher's constants and matrices from a public seed, together with tools to check it, would close most of the gap.

4. *Differentials, not characteristics.* Our bounds, and the reach of the MALICIOUS trails, are stated for single characteristics. Clustering into differentials, which the published LowMC-M attacks exploit, is bounded for none of the three ciphers.

5. *One detector for the taxonomy.* No single technique covers the whole table. Whether the structural, spectral, primitivity and provenance tests can be combined into one pipeline, and how such a pipeline would be audited, is the question this paper raises and does not answer.

11 Conclusion

A block cipher is trusted once it resists the published attacks and its S-boxes pass the standard tests. A design-level backdoor is built to survive this evaluation. Its designer hides a structure in the mathematics of the cipher and keeps the knowledge of it to themselves. Four such constructions have been published for block ciphers since 1997, and each comes with an account of what it gives its designer. Two questions had not been answered for all four. Does the hidden structure also help an outsider who uses ordinary cryptanalysis? Can an evaluator who holds only the specification find it? The available tools could not settle the second question. Generic subspace-trail search either needs the hidden subspace as input or assumes that the S-boxes have no linear structures, and a trapdoor is such a structure. Our own earlier detector returned candidates without certifying them. Consequently, an S-box that passed the usual tests was trusted by default, and an evaluator could not say which backdoors a review of the specification would catch.

In this paper we put both questions to all four constructions with one set of tools. We placed the constructions in one taxonomy and described each in Beyne's correlation-matrix framework. We corrected and proved our detection algorithm, and followed the trapdoor pairs it finds through the cipher-design equivalences and the linear layer. We then put KG-DES, BEA-1 and LowMC-M through the same battery of attacks, from exact trail searches to key recovery, and checked the trail optima with an independent MILP model. For the partition trapdoors the first question took a sharper form as the work went on. Once a trapdoor's probability-1 guarantee has been proved, how much of it survives contact with an actual attack?

The answers turn on two properties of a trapdoor, how exact its relation is and where in the cipher it sits. Exactness comes first. Rijmen and Preneel hid a biased linear approximation behind S-boxes too wide to scan. The bias makes the trapdoor usable, but it also gives the trapdoor away. Every row of the published S-box table is a noisy linear equation in the secret mask, and the linear algebra of Wu et al. recovers the mask. Balanced S-boxes rule the construction out altogether, because the planted quantity sits in the zero-input-mask row of the linear approximation table. In the correlation-matrix framework this trapdoor is a single imperfect entry, which a CAST-type round passes on by one key-independent trail.

The partition trapdoors of Paterson and of Bannier and Filiol take the opposite course, because their relation is exact. The derivative $D_u\pi$ is confined to a proper subspace. This single property of the S-box layer propagates unchanged through the round function. As Paterson and Bannier showed for their ciphers, it gives an invariant subspace of the whole cipher that holds for every key. It also gives an exact distinguisher from two chosen plaintexts, with advantage $1-2^{-32}$ for KG-DES and $1-2^{-40}$ for BEA-1. In the correlation matrix such a partition is a perfect block. Hence, for every key the cipher contains an exact smaller cipher that acts on the cosets.

What this exact relation gives an ordinary attacker has no single answer. It depends on the rest of the cipher, and our two case studies differ sharply. On KG-DES the trapdoor favours the attacker at every turn. The optimal differential characteristic has probability $2^{-21.23}$ over all sixteen rounds, against $2^{-61.97}$ for DES. It is built entirely from trapdoor transitions, and two independent methods agree on the periodic optima. Similarly, a two-round iterative linear characteristic reaches correlation 2^{-16}, against $2^{-22.42}$ for DES. This gives a known-plaintext distinguisher from about 2^{32} texts for a typical key, and it

needs no knowledge of the partition. The reason is that the trapdoor puts every maximal S-box approximation on the input bits that the expansion shares between neighbouring S-boxes. The coset maps are affine, so sixteen-round KG-DES is affine on every coset of its trapdoor subspace. Consequently, a boomerang with both differences in that subspace returns with probability 1.

The optimal characteristic also leads to the key. Peeling the last round fixes the 32 coset-label bits of the last-round subkey across the eight S-boxes. The bits that the trapdoor hides are not needed to peel that round, so a second stage fixes the coset-label bits of the round before. Together the two stages pin 49 master-key bits, and a search over the remaining 2^{7} returns the full 56-bit key. The connectivity tables tell the same story. The DLCT reaches its exact theoretical maximum, as we prove through the Walsh spectrum of the trapdoor components, and the FBCT sits at its bound. The remaining DLCT maxima, and a second, weaker near-trapdoor direction, come from one quadratic trapdoor component per S-box, and Paterson's criteria S2 and S6 force at least one such component. Interestingly, DES's expansion even gives KG-DES a Rijmen–Preneel-type relation. It sits in the round function and not in the S-boxes, and its sign depends on the key.

BEA-1 behaves differently, because there the trapdoor opens no single trail. The optimal characteristic costs $2^{-36.07}$ over three rounds and $2^{-67.67}$ over four, and every characteristic over the full eleven rounds is bounded by $2^{-170.1}$. The DLCT again reaches its exact maximum, but one derivative further the two ciphers part ways. BEA-1's coset maps are almost perfect nonlinear, and its FBCT vanishes on the trapdoor directions. Its boomerangs owe nothing to the trapdoor, since generic ladder switches beat the optimal characteristics over two and three rounds and no quartet returned at four. Similarly, the best linear characteristic lies wholly inside the trapdoor subspaces, and the branch number of MixColumns still holds it to 2^{-76}. In both ciphers a differential-linear trail reaches correlation 1 over the full cipher, the best any such trail can do. However, it restates the invariant-subspace distinguisher and adds no new attack.

The reason for the difference lies outside the trapdoor. Both trapdoors are the same kind of object, a partition that every round preserves. In KG-DES the cosets have four elements, so the coset maps are affine, and DES's linear layer, of branch number 2, lets single trapdoor transitions chain from round to round. The trapdoor then opens point differentials, linear characteristics and boomerangs. In BEA-1 the coset maps are almost perfect nonlinear and MixColumns has the maximal branch number, so none of these attacks profits.

What neither design can remove is the partition itself, and the correlation-matrix framework shows what the characteristics miss. For every key the partition leaves an exact 32-bit cipher on the cosets of KG-DES and an exact 40-bit cipher on those of BEA-1. In effect it halves the block size. As a result, BEA-1 has linear approximations of correlation at least 2^{-20}, however weak its characteristics. Both ciphers also fall to known-plaintext distinguishers at the birthday bound on the cosets, from about $2^{16.5}$ and $2^{20.5}$ texts. Table 50 gives the data, time and memory of every attack. The cheapest distinguishers need two chosen plaintexts and one comparison. The known-plaintext ones need $2^{17.5}$ and $2^{21.5}$ texts for success probability 0.98, with time and memory linear in the data. Apart from the KG-DES key recovery, none of these attacks recovers a key. For comparison, Bannier's key recovery on the deployed BEA-1 uses 2^{16} chosen plaintexts.

An independent MILP model of every trail type supports these numbers. It reproduces each optimum it can prove and brackets the others. Its boomerang optima also show how far single characteristics fall short of the true return probabilities. For KG-DES they put the sixteen-round boomerang at $2^{-36.79}$, although it returns with probability 1.

The second question, what an evaluator who holds only the specification can find, turns on where the trapdoor sits. For a trapdoor in the S-boxes the answer is encouraging.

Exactness removes the noise that hid the Rijmen–Preneel trapdoor, and with it much of the concealment. Our corrected algorithm recovers a trapdoor pair from the S-box tables alone, and we proved it sound and, below an explicit cap on its candidate set, complete. A spectral variant reads the support of the Walsh spectrum and needs no cap. It is exact whenever the masks it flags belong to a single trapdoor, and it can check this condition. The algorithm found Paterson's trapdoor in KG-DES without being told where to look. On BEA-1 it recovered all four 5-dimensional trapdoors from the secret S-boxes and none from the deployed ones. A closure computation then decides whether the pairs survive the linear layer, and for BEA-1 it derives the invariant subspace from the MixColumns matrix alone. The same machinery works unchanged for a Feistel cipher and an SPN. A trapdoor pair survives inversion and affine equivalence. Extended affine equivalence preserves it only under a condition we identify, and an extended affine, hence CCZ, equivalent S-box can carry no trapdoor at all. The method sees what it was built to see, exact coset partitions in single S-boxes. Probabilistic partitions, trapdoors on a subset of the S-boxes and structure outside the S-boxes are beyond it, and no analysis of the primitive can decide intent.

A trapdoor outside the S-boxes is a different matter, and the MALICIOUS backdoor in LowMC-M is the limiting case. Its relation is just as exact, a related-tweak characteristic of probability 1 over all but the last round. However, it holds between the permutations of two tweaks and is planted in the linear layers, so every tool that finds a partition is blind to it. In the correlation matrix it is not a block but an identity between the matrices of two tweaks. The only block, which comes from the partial S-box layer, shrinks by 2^s per round and is gone long before the end. With an ideal XOF a backdoored instance is distributed as an honest one, and the planted pair costs about 2^k to recover. Without that pair an attacker gains nothing from the backdoor. At full size and at the current round numbers, the deterministic trails of a single tweak cover 25 to 31 percent of the rounds, and the best published attacks 84 to 98 percent. The holder of the backdoor, by contrast, distinguishes the full cipher from two queries. In the seven full-size measurements the holder recovered the key from 35 to 409 chosen-tweak queries. The limits of the backdoor are practical. It needs control of the tweak, every use is traceable, and a verifiable derivation of the matrices from a public seed excludes it.

We began with the belief that a malicious designer leaves some trace of the backdoor. The case studies show where this belief holds and where it fails. A trapdoor that lives in the S-boxes and holds with certainty leaves an exact trace in the difference spans, the Walsh support, the connectivity tables and the correlation matrix, and a sound algorithm finds it. A biased trapdoor that enters the published tables linearly is found by linear algebra. However, a trapdoor planted in unexplained linear layers and protected by a pseudorandom function is found by no test of the primitive. Against it only a verifiable account of how the constants were chosen helps. Section 10 carries the same boundary across other primitives. The backdoors that an evaluator can catch from the specification fall on one side of it. Those that only provenance excludes, such as MALICIOUS, Dual EC and malicious hashing, fall on the other (Table 58).

These findings bear directly on how new designs should be evaluated. Structural tests of the S-boxes are cheap, and they find any exact partition that the S-boxes carry, as they found Paterson's in KG-DES and the four in BEA-1's secret S-boxes. They should therefore be a routine part of every evaluation. However, they cannot reach a backdoor in the linear layers or in the constants. For that part of a design, the only safeguard we know is to require that every constant and matrix be derived verifiably from a public seed (Section 10.2).

Limitations. Our results have four limits. The algorithm detects exact coset partitions in single S-boxes and nothing else. Our trail bounds hold for single characteristics, not for differentials. The key recovery is a chosen-plaintext attack on KG-DES. At sixteen rounds

it needs about 2^{26} pairs for each of the eight S-boxes, and $2^{24.8}$ more for each in the second stage. Finally, KG-DES, BEA-1 and LowMC-M were all built as demonstrations. What we have measured is therefore what such backdoors can do, not what any deployed cipher contains.

Open questions. Three gaps remain. First, the spectral detector is exact only when the masks it flags belong to a single trapdoor (Proposition 6), and on the counterexample S-box it misses the trapdoor. Whether a spectral method can be made complete without the subset search of Algorithm 1 is open. Second, BEA-1's optimal characteristic is known exactly only up to four rounds, and beyond that the active-S-box bound is all we have. That bound is tight for one and two layers but not for three and four. There the optimum lies 0.26 and 0.53 bits above it, because not every active S-box can take the maximal entry 46. Third, for LowMC-M only the deterministic trails are computed exactly. The reach of probabilistic trails rests on a counting estimate that ignores how characteristics cluster into differentials, and this clustering is what the published attacks exploit. More generally, bounding differentials, and not only characteristics, remains open for all three ciphers.

Future directions. Several directions follow from these limits. The linear hull of KG-DES's iterative approximation deserves a closer look. Its key-by-key spread, measured in Section 6.8, is a factor of about two, and characterizing the keys for which the hull cancels would sharpen the known-plaintext distinguisher. Algorithm 1 admits two natural extensions. One is the detection of partitions that hold only with high probability, and the other is the search for the subset of the S-box layer that carries a trapdoor. Beyond the ciphers studied here, two tools would together cover most of Table 58 (Section 10.3). The first is an auditable standard for deriving a cipher's constants from a public seed. The second is a single pipeline that combines the structural, spectral, primitivity and provenance tests. More broadly, our analysis can serve as a template for testing the hybrid attack vectors of current automated tools against a realistic backdoor. Such tools are usually built and tested on small curated S-boxes.

In the longer term we expect this work to change where the effort of an evaluation goes. Checking the S-boxes for an exact trapdoor becomes a routine computation on the published tables, and attention can move to what no such computation reaches, the provenance of the constants. The case studies also carry one caution. A backdoor's probability-1 guarantee and its practical bite can be very different numbers, and only testing both on real cipher data shows which of them an adversary actually gets.

A Methodology and Reproducibility

Every computational statement in the paper was obtained by a program in the accompanying repository, whose reference outputs are stored beside the code. Here we record the software, the data, the program behind each result, the treatment of randomness, and the history of the one long computation, the MILP cross-check of Section 8.2.

Software. The analysis programs are written in Python 3 (version 3.9 or later) and, where speed matters, in C compiled with `gcc`, optionally with OpenMP. The periodic MILP of Section 6.6.3 is solved with CBC through the PuLP modelling library, and the MILP cross-check of Section 8.2 with SCIP 10 through PySCIPOpt 6.2.1. Reported running times are indicative and refer to a single desktop computer.

Data. The KG-DES S-boxes are taken from Paterson's paper [Pat99]; the secret and public BEA-1 S-boxes and its MixColumns matrix from Appendix A of Bannier's thesis [Ban17];

and the DES S-boxes, used only as a control, from the DES specification. Each DES and KG-DES row is checked on import to be a permutation of $\{0, \ldots, 15\}$, and the BEA-1 tables were checked against an independent implementation of the deployed cipher byte for byte.

Randomness. Every random baseline (random S-boxes and row maps, random keys for the measured trails, boomerangs and distinguishers, and the random instances of LowMC-M) is drawn from a generator with a fixed seed, so reruns give identical numbers. The plaintext P_1 of the BEA-1 trace in Section 7.3 is drawn with seed 2026.

Programs. Table 59 maps the results to the programs that produce them. Programs in `code/` write their output to `code/results/`; the programs at the repository root cover the earlier BEA-1 experiments.

Independent checks. The results that rest on one search were recomputed along a second route wherever one exists, as the four verification tables record (Tables 35, 51, 52 and 56). The branch-and-bound searches were first run on the DES S-boxes and reproduce the published optima of Biham and Shamir and of Matsui. Algorithm 1 was tested on Paterson's published subspaces before it was run on BEA-1, and its BEA-1 output coincides with the subspaces Bannier publishes [Ban17, Section 5.2.1]. The implementation of Bannier's key recovery [Ban17, Algorithm 8] uses only the public S-boxes at attack time, and recovers a full key from 2^{16} chosen plaintexts.

The MILP run. The limit of 600 seconds per model is on SCIP's wall clock, and the machine was suspended five times during the run, for 17 hours 35 minutes in total. The run therefore took 25 hours 13 minutes of elapsed time but only about 7 hours 38 minutes of computation. A suspension can only shorten a model, never extend its search, because the wall clock runs on and SCIP stops at its next check after resuming. Two models were cut short in this way. The KG-DES differential models over 13 and 16 rounds were running during the two long suspensions, of 16 hours 45 minutes and 47 minutes. They stopped on resuming after about 171 and 131 seconds of computation, and their elapsed times of 60,485 and 2,950 seconds are almost entirely suspension. We solved both again with the same model and the same limit, without interruption. The 13-round model then reaches the optimum 16.980 with a proof, in 346 seconds. The 16-round model stops at the limit with a trail of weight 21.226, the optimum B_{16}, and a dual bound of 9.559. The three short suspensions, of 13, 17 and 146 seconds, fell inside other models and took at most that much from their budgets. Table 48 reports the reruns for the two affected models.

References

[AÅBL12] Mohamed Ahmed Abdelraheem, Martin Ågren, Peter Beelen, and Gregor Leander. On the distribution of linear biases: Three instructive examples. In *Advances in Cryptology – CRYPTO 2012*, volume 7417 of *LNCS*, pages 50–67. Springer, 2012.

[AAE+14] Ange Albertini, Jean-Philippe Aumasson, Maria Eichlseder, Florian Mendel, and Martin Schläffer. Malicious hashing: Eve's variant of SHA-1. In *Selected Areas in Cryptography (SAC) 2014*, volume 8781 of *LNCS*, pages 1–19. Springer, 2014.

[ARS+15] Martin R. Albrecht, Christian Rechberger, Thomas Schneider, Tyge Tiessen, and Michael Zohner. Ciphers for MPC and FHE. In *Advances in Cryptology –*

Table 59: Programs behind the computational results.

Result	Program
Rijmen–Preneel screen, Table 17; Proposition 2	`code/rp_screen.py`
Proposition 1 and its toy cipher, Table 2	`code/rp_correlation.py`
Figures 1, 6, 10 and the zero counts in their captions; the KG-DES mask trail of Figure 7	`code/corr_figures.py`
Algorithm 1 on the KG-DES and BEA-1 S-boxes, Tables 13 and 36	`preservable_subspaces_v2.c`
Spectral variant, Proposition 3, Table 9	`code/spectral_detector.py`, `code/spectral_detector_bea1.py`
The fallback on the counterexample, Table 8; the condition of Proposition 6 on every S-box; Lemma 2 and Proposition 14	`code/trapdoor_search.py`
Balance, uniformity bounds, the equivalence counterexamples	`code/verify_claims.py`
Connectivity tables of KG-DES, Propositions 18 and 19, the 5/9 baseline	`code/dlct_quadratic.py`
Proposition 20: the row maps that S2 and S6 allow, and the sample of S-boxes meeting Paterson's criteria	`code/kgdes_quadratic_rank.py`
Periodic MILP, Tables 20, 21 and 22	`code/kgdes_milp.py`, `code/milp_experiments.py`, `code/verify_milp_trail.py`
Optimal differential characteristics of KG-DES, Tables 23 and 24; connectivity tables of BEA-1, Table 41; BEA-1 boomerangs, Table 44	`code/verify_trails.py`, with `feistel_diff_search.c` and `sbox_tables.c`
Optimal differential characteristics of BEA-1, Proposition 30, Tables 39 and 40	`code/bea1_diff_trails.py` (four layers: arguments `4 67.861`)
Linear characteristics, Tables 30, 31 and 45, Figures 4 and 9, the key-by-key hull spread	`code/linear_trails.py`, with `feistel_linear_search.c`
Differential-linear trails, Tables 32, 33, 42 and 43, and their figures	`code/dl_trails.py`, `code/dl_trail_figures.py`
Measured boomerangs of KG-DES and of reduced BEA-1	`code/boomerang_check.py`, with `bea1_boomerang.c`
Fourier zero pattern, Proposition 4; known-plaintext distinguishers, Tables 34 and 46	`code/partition_distinguisher.py`, with `bea1_partition_kp.c`
Branch numbers, Table 47; Proposition 17	`code/branch_numbers.py`
Last-round key recovery, Proposition 23, Table 29; the per-S-box characteristics; the full master-key recovery, and its second stage (Proposition 24, option `--two-stage`)	`code/kgdes_key_recovery.py`, `.c`, `code/kgdes_trail_cover.py`, `code/kgdes_master_key_recovery.c`
MILP cross-check, Table 48	`code/milp_trails.py`, `code/milp_rerun_kgdes_diff.py`
BEA-1 probability-1 truncated differential and trace, Tables 37 and 38	`bea1_truncated_differential.py`
Bannier's key recovery on the deployed BEA-1	`bea1_cryptanalysis.c`
LowMC-M toy instance, Section 9.7	`code/lowmc_m_toy.py`
LowMC-M bounds, Tables 54 and 55; Lemma 8	`code/lowmc_m_bounds.py`
Figures 12 and 13	`code/lowmc_m_figures.py`

EUROCRYPT 2015, Part I, volume 9056 of *LNCS*, pages 430–454. Springer, 2015.

[Ban17] Arnaud Bannier. *Combinatorial Analysis of Block Ciphers With Trapdoors*. PhD thesis, École Nationale Supérieure d'Arts et Métiers, 2017. HAL Id: tel-03125786.

[BBF16] Arnaud Bannier, Nicolas Bodin, and Éric Filiol. Partition-based trapdoor ciphers, 2016. IACR Cryptology ePrint Archive, Report 2016/493.

[BCC+15] Daniel J. Bernstein, Tung Chou, Chitchanok Chuengsatiansup, Andreas Hülsing, Eran Lambooij, Tanja Lange, Ruben Niederhagen, and Christine van Vredendaal. How to manipulate curve standards: A white paper for the black hat. In *Security Standardisation Research (SSR 2015)*, volume 9497 of *LNCS*, pages 109–139. Springer, 2015.

[Bey18] Tim Beyne. Block cipher invariants as eigenvectors of correlation matrices. In *Advances in Cryptology – ASIACRYPT 2018, Part I*, volume 11272 of *LNCS*, pages 3–31. Springer, 2018.

[Bey21] Tim Beyne. A geometric approach to linear cryptanalysis. In *Advances in Cryptology – ASIACRYPT 2021*. Springer, 2021. Full version: Cryptology ePrint Archive, Paper 2021/1247.

[BF17] Arnaud Bannier and Éric Filiol. Mathematical backdoors in symmetric encryption systems – proposal for a backdoored AES-like block cipher, 2017. arXiv preprint arXiv:1702.06475.

[BHL+20] Hamid Boukerrou, Paul Huynh, Virginie Lallemand, Bimal Mandal, and Marine Minier. On the Feistel counterpart of the boomerang connectivity table: Introduction and analysis of the FBCT. *IACR Transactions on Symmetric Cryptology*, 2020(1):331–362, 2020.

[BL20] Tim Beyne and Chaoyun Li. Cryptanalysis of the MALICIOUS framework, 2020. IACR Cryptology ePrint Archive, Report 2020/1032.

[Bla94] Matt Blaze. Protocol failure in the escrowed encryption standard. In *Proceedings of the 2nd ACM Conference on Computer and Communications Security (CCS)*, pages 59–67. ACM, 1994.

[BLN16] Daniel J. Bernstein, Tanja Lange, and Ruben Niederhagen. Dual EC: A standardized back door. In *The New Codebreakers*, volume 9100 of *LNCS*, pages 256–281. Springer, 2016.

[BODKW19] Achiya Bar-On, Orr Dunkelman, Nathan Keller, and Ariel Weizman. DLCT: A new tool for differential-linear cryptanalysis. In *Advances in Cryptology – EUROCRYPT 2019*, volume 11476 of *LNCS*, pages 313–342. Springer, 2019.

[BPR14] Mihir Bellare, Kenneth G. Paterson, and Phillip Rogaway. Security of symmetric encryption against mass surveillance. In *Advances in Cryptology – CRYPTO 2014*, volume 8616 of *LNCS*, pages 1–19. Springer, 2014.

[BS93] Eli Biham and Adi Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer-Verlag, 1993.

[CCZ98] Claude Carlet, Pascale Charpin, and Victor Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Designs, Codes and Cryptography*, 15(2):125–156, 1998.

[CDVS09] A. Caranti, F. Dalla Volta, and M. Sala. On some block ciphers and imprimitive groups. *Applicable Algebra in Engineering, Communication and Computing*, 20(5–6):339–350, 2009.

[CDVSV06] A. Caranti, F. Dalla Volta, M. Sala, and F. Villani. Imprimitive permutation groups generated by the round functions of key-alternating block ciphers and truncated differential cryptanalysis, 2006. arXiv preprint math/0606022.

[CFN+14] Stephen Checkoway, Matthew Fredrikson, Ruben Niederhagen, Adam Everspaugh, Matthew Green, Tanja Lange, Thomas Ristenpart, Daniel J. Bernstein, Jake Maskiewicz, and Hovav Shacham. On the practical exploitability of Dual EC in TLS implementations. In *23rd USENIX Security Symposium*, pages 319–335. USENIX Association, 2014.

[CHP+18] Carlos Cid, Tao Huang, Thomas Peyrin, Yu Sasaki, and Ling Song. Boomerang connectivity table: A new cryptanalysis tool. In *Advances in Cryptology – EUROCRYPT 2018*, pages 683–714. Springer, 2018.

[Cop94] Don Coppersmith. The Data Encryption Standard (DES) and its strength against attacks. *IBM Journal of Research and Development*, 38(3):243–250, 1994.

[DGG+15] Yevgeniy Dodis, Chaya Ganesh, Alexander Golovnev, Ari Juels, and Thomas Ristenpart. A formal treatment of backdoored pseudorandom generators. In *Advances in Cryptology – EUROCRYPT 2015*, volume 9056 of *LNCS*, pages 101–126. Springer, 2015.

[DGV95] Joan Daemen, René Govaerts, and Joos Vandewalle. Correlation matrices. In *Fast Software Encryption (FSE 1994)*, volume 1008 of *LNCS*, pages 275–285. Springer, 1995.

[Fah24] Shah Fahd. *On the Implementation of Mathematical Backdoors in Cryptographic Algorithms and Protocols*. PhD thesis, Military College of Signals, National University of Sciences and Technology (NUST), Islamabad, Pakistan, 2024.

[FAI+23] Shah Fahd, Mehreen Afzal, Waseem Iqbal, Dawood Shah, and Ijaz Khalid. The reality of backdoored S-boxes – an eye opener, 2023. IACR Cryptology ePrint Archive, Report 2023/1073.

[FASA23] Shah Fahd, Mehreen Afzal, Dawood Shah, and Yawar Abbas. Detection of non-trivial preservable quotient spaces in S-box(es). *Neural Computing and Applications*, 2023. Published origin of the trapdoor-subspace detection algorithm and its behaviour under EA/CCZ equivalence.

[FGHT17] Joshua Fried, Pierrick Gaudry, Nadia Heninger, and Emmanuel Thomé. A kilobit hidden SNFS discrete logarithm computation. In *Advances in Cryptology – EUROCRYPT 2017*, volume 10210 of *LNCS*, pages 202–231. Springer, 2017.

[GRR16] Lorenzo Grassi, Christian Rechberger, and Sondre Rønjom. Subspace trail cryptanalysis and its applications to AES. In *IACR Transactions on Symmetric Cryptology*, volume 2016(2), pages 192–225, 2016.

[Har96] Carlo Harpes. *Cryptanalysis of Iterated Block Ciphers*. PhD thesis, ETH Zürich, 1996. ETH Series in Information Processing, vol. 7.

[HDE24] Hosein Hadipour, Patrick Derbez, and Maria Eichlseder. Revisiting differential-linear attacks via a boomerang perspective with application to AES, Ascon, CLEFIA, SKINNY, PRESENT, KNOT, TWINE, WARP, LBlock, Simeck, and SERPENT. In *Advances in Cryptology – CRYPTO 2024*, volume 14923 of *LNCS*, pages 38–72. Springer, 2024. doi:10.1007/978-3-031-68385-5_2.

[HM97] Carlo Harpes and James L. Massey. Partitioning cryptanalysis. In *Fast Software Encryption (FSE 1997)*, pages 13–27. Springer, 1997.

[IPS13] Mitsugu Iwamoto, Thomas Peyrin, and Yu Sasaki. Limited-birthday distinguishers for hash functions: Collisions beyond the birthday bound can be meaningful. In *Advances in Cryptology – ASIACRYPT 2013, Part II*, volume 8270 of *LNCS*, pages 504–523. Springer, 2013.

[LAAZ11] Gregor Leander, Mohamed Ahmed Abdelraheem, Hoda AlKhzaimi, and Erik Zenner. A cryptanalysis of PRINTcipher: The invariant subspace attack. In *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *LNCS*, pages 206–221. Springer, 2011.

[LH94] Susan K. Langford and Martin E. Hellman. Differential-linear cryptanalysis. In *Advances in Cryptology – CRYPTO 1994*, volume 839 of *LNCS*, pages 17–25. Springer, 1994.

[LIM21] Fukang Liu, Takanori Isobe, and Willi Meier. Cryptanalysis of full LowMC and LowMC-M with algebraic techniques. In *Advances in Cryptology – CRYPTO 2021*, 2021. IACR Cryptology ePrint Archive, Report 2020/1034.

[LMR15] Gregor Leander, Brice Minaud, and Sondre Rønjom. A generic approach to invariant subspace attacks: Cryptanalysis of Robin, iSCREAM and Zorro. In *Advances in Cryptology – EUROCRYPT 2015*, volume 9056 of *LNCS*, pages 254–283. Springer, 2015.

[LSW+22] Fukang Liu, Santanu Sarkar, Gaoli Wang, Willi Meier, and Takanori Isobe. Algebraic meet-in-the-middle attack on LowMC. In *Advances in Cryptology – ASIACRYPT 2022*, 2022. IACR Cryptology ePrint Archive, Report 2022/019.

[LTW18] Gregor Leander, Cihangir Tezcan, and Friedrich Wiemer. Searching for subspace trails and truncated differentials. *IACR Transactions on Symmetric Cryptology*, 2018(1):74–100, 2018.

[Mat94a] Mitsuru Matsui. The first experimental cryptanalysis of the data encryption standard. In *Advances in Cryptology – CRYPTO '94*, volume 839 of *LNCS*, pages 1–11. Springer, 1994.

[Mat94b] Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology – EUROCRYPT 1993*, volume 765 of *LNCS*, pages 386–397. Springer, 1994.

[Mat95] Mitsuru Matsui. On correlation between the order of S-boxes and the strength of DES. In *Advances in Cryptology – EUROCRYPT 1994*, volume 950 of *LNCS*, pages 366–375. Springer, 1995.

[MSD15] Ilya Mironov and Noah Stephens-Davidowitz. Cryptographic reverse firewalls. In *Advances in Cryptology – EUROCRYPT 2015*, volume 9057 of *LNCS*, pages 657–686. Springer, 2015.

[MWGP12] Nicky Mouha, Qingju Wang, Dawu Gu, and Bart Preneel. Differential and linear cryptanalysis using mixed-integer linear programming. In *Information Security and Cryptology – Inscrypt 2011*, volume 7537 of *LNCS*, pages 57–76. Springer, 2012.

[Nat94] National Institute of Standards and Technology. Escrowed encryption standard (EES). FIPS PUB 185, 1994.

[Nat06] National Institute of Standards and Technology. Recommendation for random number generation using deterministic random bit generators. NIST Special Publication 800-90A, 2006.

[Pat99] Kenneth G. Paterson. Imprimitive permutation groups and trapdoors in iterated block ciphers. In *Fast Software Encryption (FSE 1999)*, pages 201–214. Springer, 1999.

[PW20] Thomas Peyrin and Haoyang Wang. The MALICIOUS framework: Embedding backdoors into tweakable block ciphers. In *Advances in Cryptology – CRYPTO 2020, Part III*, volume 12172 of *LNCS*. Springer, 2020. Revised version (LowMC-Mv3): IACR Cryptology ePrint Archive, Report 2020/986. doi:10.1007/978-3-030-56877-1_9.

[RP97] Vincent Rijmen and Bart Preneel. A family of trapdoor ciphers. In *Fast Software Encryption (FSE 1997)*, volume 1267 of *LNCS*, pages 139–148. Springer, 1997.

[RTYZ16] Alexander Russell, Qiang Tang, Moti Yung, and Hong-Sheng Zhou. Cliptography: Clipping the power of kleptographic attacks. In *Advances in Cryptology – ASIACRYPT 2016*, volume 10032 of *LNCS*, pages 34–64. Springer, 2016.

[SCI] SCIP optimization suite, version 10.0, via PySCIPOpt 6.2.1. https://www.scipopt.org. Solver used for the MILP cross-check.

[SF07] Dan Shumow and Niels Ferguson. On the possibility of a back door in the NIST SP800-90 Dual Ec Prng. CRYPTO 2007 rump session, 2007.

[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 *LNCS*, pages 3–33. Springer, 2016.

[vOW99] Paul C. van Oorschot and Michael J. Wiener. Parallel collision search with cryptanalytic applications. *Journal of Cryptology*, 12(1):1–28, 1999.

[WBDY98] Hongjun Wu, Feng Bao, Robert H. Deng, and Qin-Zhong Ye. Cryptanalysis of Rijmen-Preneel trapdoor ciphers. In *Advances in Cryptology – ASIACRYPT 1998*, volume 1514 of *LNCS*, pages 126–132. Springer, 1998.

[YY96] Adam Young and Moti Yung. The dark side of "black-box" cryptography, or: Should we trust capstone? In *Advances in Cryptology – CRYPTO '96*, volume 1109 of *LNCS*, pages 89–103. Springer, 1996.

[YY97] Adam Young and Moti Yung. Kleptography: Using cryptography against cryptography. In *Advances in Cryptology – EUROCRYPT '97*, volume 1233 of *LNCS*, pages 62–74. Springer, 1997.