Which Masks Are Worth Testing? Key-Dependent Correlation and Symmetry in Lightweight Block Ciphers

Alireza Gholizadeh Shahrbejari[1] and Reza Ebrahimi Atani[1]

[1]Department of Computer Engineering, University of Guilan, Rasht, Iran

Abstract. The quality of a differential–linear candidate and the cost of selecting one are separate questions. We study both in reduced-round lightweight block ciphers. First, for 11-round Craft, we compare an alternative output mask with the published mask of Zhou et al. at the same input difference. In a paired study with 128 master keys, the alternative increases the observed mean signed correlation by factors of 1.69 to 1.86 across four tweak settings. Approximate simultaneous 95% confidence intervals for the four mean differences lie above zero. A separate held-out study of ranking rules selects two distinct masks and gives inconclusive simultaneous intervals for detection-rate gains at a fixed data budget. Second, for restricted two-round linear sections, we use cell permutations to identify masks with the same distribution of absolute correlation over keys. We observe reductions in the number of evaluated masks in Midori64 and Qarma64. For three symmetric Qarma64 cases, 81 masks reduce to 21, 45, and 36 representatives. Exact evaluation agrees with exhaustive search for RMS, stability, lower-tail, and threshold coverage objectives. The tested coordinate-permutation family gives only the identity in the selected Skinny-64-64 section.

Keywords: Differential–linear cryptanalysis, key-dependent correlation, mask ranking, cell permutation symmetry, lightweight block ciphers

1 Introduction

A differential–linear distinguisher examines a fixed output parity on pairs of plaintexts with a fixed input difference. If that parity agrees more or less often than it would under a balanced reference model, the discrepancy can supply a distinguishing signal. Finding such a pair of difference and mask is a search problem. Quantifying its benefit requires specifying what is averaged over keys, what data budget is available, and how the statistic is used.

The search and evaluation problems interact. Many candidates look promising under a trail decomposition, in which the differential and linear ends are joined through a middle section. A model of the connection can be an effective filter even when it differs from the observed correlation after all rounds. The output mask can also matter when the input difference and round count are held fixed. Comparing two masks on the same keys, tweaks, and base plaintexts isolates the empirical effect of that choice.

The distribution over keys gives a second reason to compare candidates carefully. Two masks may have similar root-mean-square correlations and differ in lower-tail performance. A score that rewards a few strong keys and a score that rewards more consistent behavior are answering different questions. Estimated correlations on a finite discovery sample

E-mail: gholizadeh.a2000@gmail.com (Alireza Gholizadeh Shahrbejari), rebrahimi@guilan.ac.ir (Reza Ebrahimi Atani)

add another source of variation. We therefore separate the choice of a candidate from its evaluation on fresh keys, and report the mean signed correlation and fixed-budget rejection rate as different empirical quantities.

The cost of ranking can be studied separately on sections where correlations are exactly computable. Suppose two masks have identical distributions of absolute correlation over effective keys. Any score that depends only on that distribution assigns them the same value. If a permutation of cipher cells proves that equality, we can compute the distribution of one mask and reuse it for its entire equivalence class. Whether this reduction is possible depends on the diffusion layer, the fixed input mask, and the distribution of the effective key. It is therefore useful to include both sections with nontrivial symmetries and a section where the tested family gives only the identity.

1.1 Related work and positioning

The differential–linear connectivity table links differential and linear behavior, and subsequent work refined the treatment of dependencies across middle rounds [4, 7]. Yan et al. and Zhou et al. provide concrete differential–linear instances for Craft [11, 12]. We use the 11-round mask from Zhou et al. as a remeasured reference. Their multi-cipher automatic search framework and longer-round Craft results address a broader search problem.

Röck and Nyberg showed how a key-mask basis and a Walsh transform can recover correlation classes for a linear hull [10]. Kranz et al. studied the influence of key schedules and tweaks, including variation across keys [9]. LibFWHT applies distribution-based ranking to exact per-key spectra and analyzes key dependence in reduced-round distinguishers [8]. Our Craft experiment uses sampled correlations for an equal-budget comparison among ranking rules. For the linear sections, the histograms are exact over the specified effective key spaces. Alfarano et al. exploit cell-permutation equivalences in the design of ShiftRows alternatives [1]. Here the cipher and the input mask are fixed, and the equivalences act on the output masks being evaluated.

Table 1: Relationship to selected prior work. Each row identifies the closest point of contact with one of the experiments.

Work	Main point of contact	Role in this study
Bar-On et al. and Hadipour et al. [4, 7]	Connectivity and dependencies in differential–linear analysis	Middle-section scores guide discovery. Direct measurements evaluate the selected full 11-round interval.
Yan et al. and Zhou et al. [11, 12]	Published differential–linear instances for Craft	Zhou's 11-round mask is remeasured at the same difference, key, tweak, and plaintext budget as the alternative.
Röck and Nyberg and Kranz et al. [10, 9]	Key-dependent linear hulls and correlation distributions	Walsh evaluation supplies the exact local histograms used by the symmetry experiment.
Hadipour et al. (LibFWHT) [8]	RMS, stability, and lower-tail ranking using per-key correlations	Three score definitions are used alongside threshold coverage under a shared sampled budget in Craft.
Alfarano et al. [1]	Cell equivalence in searches for suitable design permutations	We evaluate one output mask per orbit for a fixed two-round section and fixed input mask.

1.2 Contributions and organization

Section 2 reports a paired comparison of two 11-round CRAFT masks at one fixed difference. It then tests whether four ranking rules select the same mask and compares their frozen selections on fresh keys. Section 3 gives a sufficient condition for exact mask equivalence and measures its consequences on local sections of three ciphers. The QARMA64 experiment explicitly checks that this equivalence preserves the optimum across all evaluated ranking objectives. Together, the experiments address the quality of mask selection and the cost of evaluating the candidate set.

2 Candidate selection and validation for 11-round CRAFT

Statistic and endpoints. Let $E^{(11)}_{K,T}$ apply rounds indexed 0 through 10 under master key K and tweak T. For a fixed difference δ and output mask β, define

$$c_{K,T}(\delta,\beta) = \mathbb{E}_X(-1)^{\langle \beta, E^{(11)}_{K,T}(X)\oplus E^{(11)}_{K,T}(X\oplus\delta)\rangle}.$$

The expectation is over uniform base plaintexts. We estimate it by the mean of the observed signs. The conventional probability bias relative to $1/2$ is $c_{K,T}/2$. We use the cell order of Zhou et al. and the round function of CRAFT [6, 12]. Both masks use input difference `2000220000002000`. The reference output mask is `a0000000000a0000` from Table 10 of [12]. The alternative output mask is `5000000000050000`.

Discovery. The initial search kept the $2 + 8 + 1$ round split and fixed the active cell at each boundary of the eight-round middle section. It screened 225 combinations of nonzero middle difference and mask nibbles on 24 keys at zero tweak, with 2^{18} samples per middle difference and key. A model based on the differential transition table for the prefix and the linear approximation table for the suffix supplied the exterior terms. The product of those terms and the measured middle correlation served as a ranking proxy. The reference used middle difference 2 and mask 5. The selected alternative used middle difference 2 and mask `E`. Subsequent measurements evaluate the full 11 rounds. They estimate the combined correlation directly.

From the middle score to an external mask. The prefix comparison estimates how frequently the input difference reaches the selected middle difference. The middle experiment measures the signed response of the output mask to this difference, and the suffix score weights the linear extension. This decomposition reduces the screening burden. Dependence on other middle transitions and on the key makes direct validation of the external endpoints necessary. Once the two endpoints are fixed, the paired experiment estimates their difference in full-window correlation without conditioning on a particular intermediate transition.

Paired validation across tweaks. The masks were fixed before measuring a separate set of 128 pseudorandom master keys. Each key was evaluated at zero tweak, two fixed nonzero tweaks, and one uniformly sampled nonzero tweak, using 2^{21} pairs per key and tweak setting. Both masks share the same base plaintexts. The two fixed nonzero tweaks are `0000000000000001` and `0123456789abcdef`. The four tweak settings also share the master keys. The same tweak is used for both members of each plaintext pair. A paired difference is formed within each key and setting. The intervals use a Student t approximation with 127 degrees of freedom and a Bonferroni adjustment for eight comparisons.

The family consists of the four gains, three differences in gain relative to zero tweak, and the average gain across the three nonzero settings.

Table 2: Paired 11-round comparison. Means, differences, and interval limits are scaled by 10^4. The ratio is a point estimate.

Tweak	Reference	Alternative	Ratio	Gain	95% interval
Zero	5.061	9.403	1.86	4.342	[2.014, 6.670]
Fixed 1	4.910	9.154	1.86	4.243	[1.826, 6.661]
Fixed `0123456789abcdef`	5.403	9.145	1.69	3.742	[1.295, 6.189]
Random nonzero	5.458	9.979	1.83	4.521	[1.901, 7.142]

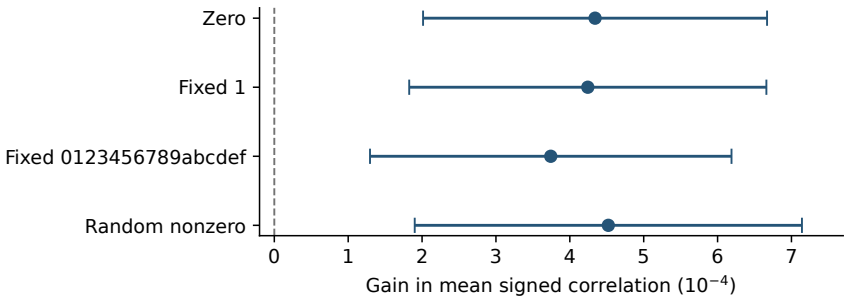


Figure 1: Paired gains in mean signed correlation. Lines show approximate simultaneous 95% confidence intervals.

Table 2 gives the primary comparison. Resampling whole key records in a sensitivity analysis also yielded positive intervals for the four gains. The intervals for changes in gain relative to zero tweak included zero. A separate zero-tweak holdout with 128 keys and 2^{22} pairs per key gave a mean gain of 4.515×10^{-4} with an individual approximate 95% interval $[3.329, 5.701] \times 10^{-4}$. The experiments estimate the mean over the sampled key and tweak regimes. They leave per-key magnitudes and the data complexity of a particular attack as separate quantities.

Zhou et al. give $2^{-14.03}$ as the model estimate and $2^{-10.92}$ as the experimental correlation of their 11-round candidate [12, Tables 1 and 10]. Our zero-tweak reference mean, $2^{-10.95}$, is close to their experimental value. Our reported improvement uses the paired comparison with that remeasured reference under identical conditions.

Ranking with a common discovery budget. We fixed the input difference and enumerated the 225 masks with nonzero nibbles at positions 0 and 11. For each tweak regime (zero and random nonzero), we ranked masks using a shared sample of 32 master keys and 2^{22} pairs per key and regime. Let a_i be the absolute measured correlation for discovery key i. The four rules maximize the root mean square, the mean minus the sample standard deviation, the mean of the smallest $\lceil 0.25K_d \rceil$ values with $K_d = 32$, and the fraction strictly above $1/1024$, respectively. The first three follow the score definitions in [8, Section 6.3]. Here correlations are estimated from sampled pairs. Each rule sees the same measurements, and ties select the numerically smallest mask.

We froze those selections and evaluated them with the two fixed masks on 128 fresh master keys, with $N = 2^{22}$ pairs per key and regime. The primary endpoint is the probability that the sign sum S satisfies

$$|S| \geq \left\lceil \sqrt{2N \log(200)} \right\rceil .$$

Under a null of independent balanced signs this threshold gives a false rejection probability of at most 1% by Hoeffding's inequality. The measured rate incorporates both key and plaintext variation. A larger mean signed correlation can occur without a conclusive increase in this rate. The rate is determined by how often the measured magnitude reaches the rejection threshold for the chosen sample size, whereas the mean includes all signed measurements. We therefore interpret this stage as a separate fixed-budget check on the masks selected by the discovery rules. Paired rate differences use intervals formed by subtracting Clopper–Pearson intervals for the probabilities of the two types of discordant outcomes. The adjustment covers both probabilities in each of the ten prespecified contrasts.

Table 3: Frozen selections and observed rejections on 128 validation keys per regime. Mask shorthand `ab` denotes nibbles `a` and `b` at positions 0 and 11. Identical masks share outcomes.

	Zero tweak		Random nonzero tweak	
Rule	Mask	Reject	Mask	Reject
Reference	`aa`	1	`aa`	1
Fixed alternative	`55`	10	`55`	16
RMS	`55`	10	`55`	16
Stable	`77`	6	`55`	16
Lower quarter	`55`	10	`55`	16
Threshold coverage	`77`	6	`77`	9

RMS and lower-tail ranking recover the fixed alternative `55` in both tweak regimes. Threshold coverage selects `77`, which has fewer observed rejections than `55` in both regimes. For `55` versus the reference, the observed rate differences are 7.03 and 11.72 percentage points at zero and random nonzero tweak. The simultaneous intervals are $[-4.36, 17.58]$ and $[-1.51, 23.58]$ percentage points. All prespecified contrasts include zero. The data establish the observed selections for this discovery sample. The repeatability of the rankings under resampling of discovery keys remains an open empirical question.

3 Symmetry and exact ranking of linear masks

Model. Let F_u be a local two-round map on four cells, with u drawn uniformly from its effective key space U. It has the form $S \circ L \circ T_u \circ S$ or $S \circ T_u \circ L \circ S$, where S acts with the same S-box on each cell, L is linear, and T_u adds the effective key. Linear layers at the physical boundaries can be absorbed into the masks by transposition. When a boundary contributes a key-dependent common sign to the correlation, taking its absolute value removes that sign.

For cell permutations P, Q satisfying $QL = LP$, S commutes with each permutation. In the first form above, $F_{Pu}(Px) = QF_u(x)$. In the second form, the same identity holds with Qu in place of Pu. The effective key relabeling must preserve its probability law. An invariant key subspace with uniform sampling suffices. A fixed offset is also harmless when the relevant relabeling preserves the resulting affine key space. We restrict to permutations that preserve the input mask α.

Distribution invariance. Let φ be the resulting measure-preserving key relabeling and $q_u(\alpha, \beta) = \mathbb{E}_X(-1)^{\langle \alpha, X \rangle + \langle \beta, F_u(X) \rangle}$. Substitution of PX for X shows

$$|q_{\varphi(u)}(\alpha, Q\beta)| = |q_u(\alpha, \beta)|.$$

Here cell permutations act on masks by the same coordinate relabeling, and the formula also holds if a boundary introduces a common sign. Consequently the absolute-correlation histogram of $Q\beta$ equals that of β over uniform effective keys. For an output family closed under the retained permutations, choose the lexicographically smallest mask in each orbit. Any score that is a function of the histogram is constant on the orbit, so maximizing over these representatives gives the same score and smallest-mask tie break as exhaustive search.

Search guarantee and cost. Write B for the candidate family and G for its retained group of output permutations. Canonicalizing every element of B costs a pass through the candidate family and the retained permutations. The number of expensive histogram evaluations drops from $|B|$ to $|B/G|$. The savings in elapsed time also depend on the ranks of the representatives' key-mask spaces and on setup costs. This is why the experiment reports both mask counts and measured time. For the tested detector, trying each input/output permutation pair is practical because the local section has only four cells.

Exact evaluation. For each endpoint, the signed correlation is a sum of character terms over effective key masks. The implementation combines terms with equal characters and applies an integer Walsh transform on independent key parities [10]. The resulting counts are weighted by the sizes of the parity classes. Each local model uses the effective key distribution induced by the cipher under the specified master-key sampling. All cases use the full output family $\{\mathtt{3},\mathtt{5},\mathtt{A}\}^4$ and evaluate the strict coverage thresholds 1/64, 1/32, and 1/16.

Cipher instances. In Midori64 [3], the tested column diffusion is $J + I$ over $\mathbb{F}_2$, with J the all-ones matrix. Its coordinate permutations commute with diffusion. With fixed input mask, permutations within positions bearing the same input nibble are retained. In original Qarma64 with S-box σ_1 and diffusion $M_{4,2}$ [2], the four-cell column has a circulant rotation-valued diffusion. The retained permutations are obtained by testing the matrix identity with its rotation operators intact and then fixing the local input mask. The effective four-cell tweakey is uniform over 16 bits for a fixed tweak and uniform core key. Direct local codebook checks at two key/tweak settings per case verify the modeled correlations at a fixed audit mask.

For Skinny-64-64 [5], the chosen column uses the actual two-cell key injection (256 effective key assignments). We tested all 24^2 input/output pairs of four-cell permutations. The selected diffusion admits only the identity pair even before imposing input-mask and key-support constraints. This result concerns that permutation family and section.

Comparison with exhaustive search. Table 4 and Figure 2 report the reduction in evaluated masks. Each experiment compares the two methods with the same evaluator. The measured speedups are medians of paired elapsed-time ratios, with six repetitions and alternating or rotated method order. The timings characterize the implementations and tested domains. For the Skinny-64-64 cases, detection falls back to the streaming evaluator.

For the four Qarma64 patterns, we separately recomputed all 81 histograms and all representative histograms and checked equality for every mask and its representative. We ranked both sets by exact RMS, a population version of the stability score with standard deviation taken over the entire effective key distribution, the mean of the lowest quarter of that distribution, and the three strict threshold coverage scores. For each pattern and each of the six objectives, the exhaustive and orbit-based searches returned the same smallest maximizing mask.

In Midori64, the all-equal input gives the largest candidate reduction among the reported patterns, from 81 masks to 15. The counts increase to 30, 36, and 54 as the

equality pattern is relaxed. In QARMA64, `55EE` and `5E5E` have the same multiplicities of input nibbles but different arrangements. Their representative counts (45 and 36) show that position matters for the tested diffusion. The all-distinct input leaves 81 masks in both ciphers. The SKINNY-64-64 cases also retain all 81 masks, reflecting the restriction of the diffusion itself.

Table 4: Mask counts and median paired speedup against exhaustive streaming evaluation. The setup is included in the timing. For cases with 81 representatives, no transform is saved.

Cipher	Input	Representatives	Speedup
Midori64	`5555`	15	2.32×
	`555E`	30	1.70×
	`55EE`	36	1.63×
	`55EA`	54	1.30×
	`5E7A`	81	0.91×
QARMA64	`5555`	21	3.79×
	`55EE`	45	1.71×
	`5E5E`	36	2.20×
	`5E7A`	81	1.05×
SKINNY-64-64	`5555`	81	0.95×
	`5E7A`	81	1.02×

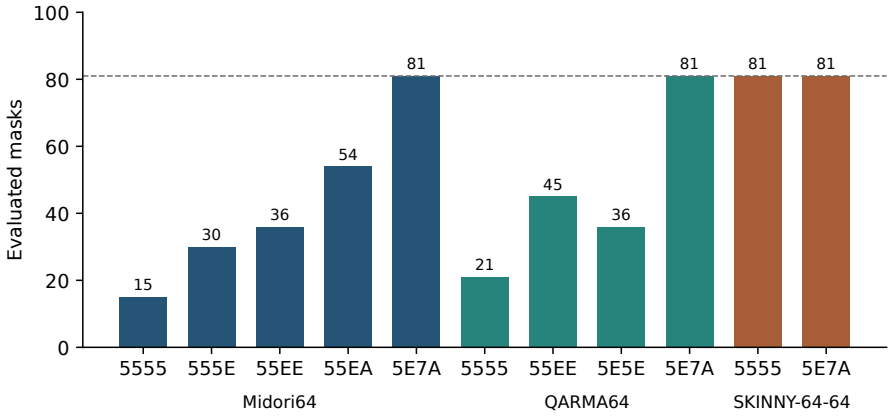


Figure 2: Number of masks evaluated in the two-round experiments. The dashed line is the exhaustive count of 81.

Table 5: Lexicographically smallest winning local output masks in the QARMA64 audit. All six columns agree with exhaustive enumeration.

Input	RMS	Stable	Lower quarter	1/64	1/32	1/16
`5555`	`5555`	`3355`	`3355`	`5555`	`3A5A`	`3333`
`55EE`	`3333`	`3333`	`3333`	`3333`	`5555`	`3333`
`5E5E`	`5555`	`3353`	`55AA`	`5555`	`5555`	`3333`
`5E7A`	`AAA5`	`A533`	`AAAA`	`AAA5`	`AAA5`	`3333`

4 Discussion and limitations

Interpreting the CRAFT comparison. The paired experiment measures a gain in average signed correlation over the stated sample of master keys and tweak settings. The agreement across settings strengthens the empirical comparison with the published endpoint. Key-level variability, however, means that a ratio of average correlations cannot be squared and interpreted as a data saving for every key. The fixed-budget experiment measures a quantity closer to the behavior of a specific test. Its simultaneous intervals remain wide enough to include zero, even where the observed rejection count is higher. Repeated discovery samples would be needed to compare the stability of the ranking rules themselves.

Interpreting the symmetry reduction. Histogram equality proves that the selected orbit representatives suffice for the tested scores. The largest observed reductions occur where the chosen input pattern is invariant under more of the diffusion-compatible permutations. The number of eligible permutations alone does not determine the exact candidate count, because different masks can have different orbit sizes. Likewise, a reduction in mask count and a reduction in running time need not have the same factor. Keeping the two figures separate makes the cost of setup and the variation in evaluation cost visible.

The connection between the two experiments is the use of key-dependent scores for mask selection. In Craft, the score is computed from sampled correlations and may change the selected external mask. In the local linear sections, exact distributions permit a guarantee that a particular symmetry reduction preserves the optimum. A finite discovery sample would require additional care about how keys are relabeled or paired across an orbit. Extending the guarantee to an entire differential–linear construction would require a symmetry of the whole construction, including the fixed input difference and the relevant key and tweak distributions.

5 Conclusion

For the fixed input difference of Zhou et al.'s 11-round Craft instance, the alternative output mask increases observed mean signed correlation by a factor of 1.69 to 1.86 across the four tested tweak settings. The separate fresh-key experiment recovers this mask under both RMS and lower-quarter ranking, while threshold coverage selects a different mask. The uncertainty in the fixed-budget rejection rates limits the performance conclusion to the reported correlation comparison.

For two-round linear sections, the same distribution over effective keys can make distinct output masks interchangeable for several ranking objectives. In the evaluated Qarma64 cases, orbit representatives retain the exhaustive optimum for every objective. The Midori64 and Skinny-64-64 comparisons show how the reduction depends on the compatible permutation structure. These results motivate testing larger candidate families with a matched optimized evaluator and repeating the discovery stage on independent key samples. The present comparisons remain tied to the stated domains, estimands, and published Craft reference.

References

[1] Gianira N. Alfarano, Christof Beierle, Takanori Isobe, Stefan Kölbl, and Gregor Leander. ShiftRows alternatives for AES-like ciphers and optimal cell permutations for Midori and Skinny. *IACR Transactions on Symmetric Cryptology*, 2018(2):20–47, 2018.

[2] Roberto Avanzi. The QARMA block cipher family. *IACR Transactions on Symmetric Cryptology*, 2017(1):4–44, 2017. https://eprint.iacr.org/2016/444.

[3] Subhadeep Banik, Andrey Bogdanov, Takanori Isobe, Kyoji Shibutani, Harunaga Hiwatari, Toru Akishita, and Francesco Regazzoni. Midori: A block cipher for low energy. In *ASIACRYPT*, pages 411–436, 2015.

[4] Achiya Bar-On, Orr Dunkelman, Nathan Keller, and Ariel Weizman. DLCT: A new tool for differential-linear cryptanalysis. In *EUROCRYPT*, pages 313–342, 2019.

[5] Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. The SKINNY family of block ciphers and its low-latency variant MANTIS. In *CRYPTO*, 2016. https://eprint.iacr.org/2016/660.

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

[7] Hosein Hadipour, Patrick Derbez, and Maria Eichlseder. Revisiting differential-linear attacks via a boomerang perspective. In *CRYPTO*, 2024. https://eprint.iacr.org/2024/255.

[8] Hosein Hadipour, Saleh Khalaj Monfared, Jens Alich, and Jan Vorloeper. LibFWHT: From exact walsh spectra to key dependence in differential-linear correlations, 2026. Cryptology ePrint Archive, Report 2026/1997, https://eprint.iacr.org/2026/1997.

[9] Thorsten Kranz, Gregor Leander, and Friedrich Wiemer. Linear cryptanalysis: key schedules and tweakable block ciphers. *IACR Transactions on Symmetric Cryptology*, 2017(1):474–505, 2017. https://eprint.iacr.org/2017/154.

[10] Andrea Röck and Kaisa Nyberg. Exploiting linear hull in matsui's algorithm 1, 2011. Extended version of WCC 2011, https://research.ics.aalto.fi/publications/arock/linHullEV.pdf.

[11] Mei Yan, Siwei Chen, Zejun Xiang, Shasha Zhang, and Xiangyong Zeng. A novel method for finding differential-linear distinguishers: application to Midori64, CRAFT, and Skinny64. In *CANS*, pages 215–239, 2024. https://doi.org/10.1007/978-981-97-8016-7_10.

[12] Yanyan Zhou, Lin Jiao, Senpeng Wang, Yunong Wu, Bin Hu, Tairong Shi, and Kai Zhang. An improved automatic framework for searching for differential-linear distinguishers with applications to SPN and Feistel block ciphers. *Designs, Codes and Cryptography*, 94:153, 2026. https://doi.org/10.1007/s10623-026-01867-1.