

Improved Related-Key Boomerang Distinguisher for Full-Round FUTURE

Guoqiang Liu , Suping Liu and Wuyou Zhang

College of Science, National University of Defense Technology, Changsha, China

Abstract

FUTURE is a lightweight block cipher with a 64-bit block, a 128-bit key and 10 rounds, proposed at AFRICACRYPT 2022 for low-latency hardware. This study analyzes FUTURE in the related-key setting with a bit-level constraint model that carries the exact weights of the differential distribution table and the exact entries of the boomerang connectivity table, and whose objective function $2w_0 + 2w_1 + w_{\text{bct}} = -\log_2(p^2 q^2 r)$ optimizes both sub-ciphers and the middle layer of the sandwich framework together. The model returns a full-round distinguisher whose objective function value 36 is optimal over all switching rounds and whose probability is $P = \hat{p}^2 \bar{r} \hat{q}^2 = 11 \cdot 2^{-39} \approx 2^{-35.54}$, at a cost of $2^{37.54}$ queries under four related keys and $2^{37.54}$ XOR operations. Running it on the full cipher over $2^{44.34}$ quartets returns 341 right quartets and an experimental probability of $2^{-35.92}$, in 63.9 hours on an ordinary personal computer; at the previous full-round probability $2^{-45.8}$ the same computer would take about 6.8 years. The distinguisher is therefore a practical one, and improves the best previously known full-round related-key boomerang distinguisher of FUTURE by a factor of $2^{10.26}$ in probability, in data and in time. An 8-round distinguisher placed into the unified key recovery framework further gives a full-round attack with $2^{51.05}$ data, 2^{64} time and 2^{64} memory, whose time complexity attains the optimum of the framework at any memory within the codebook and improves the best previously known attack by a factor of 2^6 .

Keywords: block cipher, FUTURE, related-key boomerang attack, differential cluster

1 Introduction

Lightweight block ciphers built for low-latency hardware keep the round function shallow, so that one round fits into a single clock cycle, and reduce the key schedule to rotations of the master key. Both choices are visible to a related-key adversary. A round key difference that equals the state difference cancels it, and the round then contributes no active S-box at all, so under a related key a cipher of this kind can pass whole rounds for free. The related-key model is therefore the setting in which the security margin of such a design is thinnest. It is also the setting a block cipher has to survive before it can be used in a hashing mode whose message block enters through the key input.

The resistance of a new design to boomerang attacks [Wag99] is nevertheless often argued at design time from a count of active S-boxes in the single-key model. The probability of a boomerang quartet is written as $\hat{p}^2 \hat{q}^2$, and each factor is bounded by the

This work was supported by the National Cyber Security-National Science and Technology Major Project (No. 2026ZD1500700) and the National Natural Science Foundation of China (No. 61702537).

Emails: liuguoqiang87@hotmail.com (Guoqiang Liu), liusuping24@163.com (Suping Liu), zhang-wuyou@nudt.edu.cn (Wuyou Zhang)

Date of this document: 2026-09-04.

minimum number of active S-boxes of the corresponding sub-cipher; the design is then declared secure once the product is below 2^{-n} for an n -bit block. Such an argument carries over to the related-key model only if the two counts agree, and for a key schedule made of rotations of the master key they need not.

FUTURE [GPS22] is a lightweight block cipher of exactly this kind. It has a 64-bit block, a 128-bit key and 10 rounds, and is an AES-like substitution-permutation network over $\text{GF}(2^4)$ whose state is a 4×4 array of 4-bit cells (nibbles). Each round applies SubCell, MixColumn, ShiftRow and AddRoundKey, and the round keys are rotations of the two halves of the master key. Its designers analyze the resistance of FUTURE to boomerang attacks in Section 5.3 of [GPS22]. They bound $\hat{p}^2$ and $\hat{q}^2$ by the minimum number of active S-boxes of the two sub-ciphers, observe from their Table 4 that any split of eight rounds carries at least 32 active S-boxes in total, and conclude on that basis that the full cipher is secure against boomerang attacks. Table 4 of [GPS22] gives 9, 25 and 26 active S-boxes over 3, 4 and 5 rounds, so a single-key characteristic over 4 rounds has weight at least 50. The optimal related-key weights obtained in Section 3 over the same numbers of rounds are 2, 4, 14 and 24; over 4 rounds the optimum drops from 50 to 4, a factor of 2^{46} . This gap is the margin that the related-key model removes, and locating a distinguisher inside it calls for an automatic search.

Automatic search for boomerang distinguishers has been developed in several directions. Liu et al. [LS19] build the BCT into an automatic search and search the middle layer together with the related-key differential characteristics of the two sub-ciphers, and obtain 19-round distinguishers of GIFT-64 and GIFT-128. Hadipour et al. [HBS21] give a MILP model of the rectangle attack that takes the clustering effect of the two sub-ciphers into account. Bariant et al. [BL23] search truncated boomerang characteristics of AES-based ciphers. Delaune et al. [DDV20] give a constraint model for boomerang characteristics with a proved optimal objective, and Derbez et al. [DEFN22] revisit related-key boomerang attacks on AES with a computer-aided tool. Delobel et al. [DDG⁺24] instantiate truncated characteristics with constraint programming. Zeng et al. [ZTX25] improve the search of boomerang distinguishers for generalized Feistel ciphers, and Wu et al. [WZZ⁺26] give search models for related-key boomerang distinguishers of ciphers with a bit-level key schedule. In most of these models the quantity minimized is the number of active S-boxes, and the middle layer of the sandwich framework $E = E_1 \circ E_m \circ E_0$ of Dunkelman et al. [DKS14] is treated separately from the two sub-ciphers. The model of [LS19] instead places DDT and BCT weights in a single objective, and it is that line we follow here. Two quantities are then open to improvement. The first is the quantity that is minimized, which is not the boomerang probability itself. The second is the probability finally reported for the selected characteristic, which is an estimate: either the single-characteristic value $p^2 q^2 r$ or the product $\hat{p}^2 \hat{q}^2$ of the two differential effects.

FUTURE itself has been analyzed by several techniques. İlter et al. [İS23] give MILP-aided differential and linear results, and Das et al. [DG26] model the cipher at the nibble level and place the cost of enumerating the last-round subkey in the objective of the trail search; both work in the single-key setting. Schrottenloher et al. [SS23] include FUTURE in a simplified model of meet-in-the-middle attacks, and Lin et al. [LZL23] apply the differential meet-in-the-middle attack to it. Mondal et al. [MRSA24] give a yoyo analysis, Roy et al. [RDMA24] and de Carvalho et al. [dCK26] give biclique attacks on the full cipher, and Xu et al. [XCHW25] give an integral attack on the full cipher. In the related-key model, Jana et al. [JDC⁺26] give a boomerang distinguisher over all 10 rounds with probability $2^{-45.8}$, which is the best previously known result in the setting of this paper.

In this study, we analyze FUTURE in the related-key setting and improve on the state of the art in five respects. Firstly, the search is carried out on the probability itself rather than on a count of active S-boxes. A bit-level constraint model is built in which the

S-box layers are constrained cell by cell by the 96 entries of the support of the differential distribution table (DDT), each with its exact weight. The middle layer is constrained by the 149 entries of the support of the boomerang connectivity table (BCT) of Cid et al. [CHP⁺18], again with their exact weights, and the round key differences by the rotation amounts of the key schedule. The objective function is $2w_0 + 2w_1 + w_{\text{bct}}$, which is exactly $-\log_2(p^2 q^2 r)$, so that the two sub-ciphers and the middle layer are optimized jointly. Secondly, the cluster correction of the distinguisher is evaluated in closed form. The probability equals $\hat{p}^2 \bar{r} \hat{q}^2$, where $\bar{r}$ is the average of the connection probability over the differential clusters of the two sub-ciphers. For the distinguisher found here the upper cluster holds eight differences of probability 2^{-14} each, and the union of their supports meets the support of the lower middle-layer difference in a single cell. The whole cluster correction is therefore determined by a single column of the BCT, which gives $\bar{r} = 11/64$ and a rational value for the probability. Thirdly, the distinguisher is optimal in two precisely stated senses: its objective function value 36 is optimal over all switching rounds, and, with the upper sub-cipher fixed, its lower middle-layer difference maximizes the probability over all 64-bit differences.

Furthermore, the distinguisher is executed on the full cipher itself. Over $2^{44.34}$ quartets under uniformly drawn related keys, some $2^{8.80}$ times the data the distinguisher requires, it returns 341 right quartets and an experimental probability of $2^{-35.92}$. The probability is therefore measured on all ten rounds, and not extrapolated from a reduced number of rounds. The improvement in probability is what makes such a measurement practical: those 341 right quartets are collected in 63.9 hours of run time on an ordinary personal computer, where the same count at the probability $2^{-45.8}$ of [JDC⁺26] keeps the same computer running for about 6.8 years. Finally, an 8-round boomerang distinguisher with output difference constrained to two columns is built into the unified framework of [SZY⁺22, YSZ⁺24]. It leads to a full-round key recovery attack with $2^{51.05}$ data, 2^{64} time and 2^{64} memory; the time complexity 2^{64} attains the optimum of the framework at any memory within the codebook, improving the best previously known attack by a factor of 2^6 .

The resulting distinguisher covers all 10 rounds of FUTURE with probability

$$P = \hat{p}^2 \bar{r} \hat{q}^2 = 11 \cdot 2^{-39} \approx 2^{-35.54},$$

against 2^{-64} for a random permutation, and uses $2^{35.54}$ quartets, that is $2^{37.54}$ chosen plaintexts and adaptively chosen ciphertexts under four related keys, $2^{37.54}$ XOR operations and negligible memory. It improves the probability of the best previously known full-round related-key boomerang distinguisher of FUTURE, and likewise its data and time complexity, by a factor of $2^{10.26}$; of that factor, $2^{9.8}$ comes from the exact search and $2^{0.46}$ from the cluster weighting. The comparison is shown in Table 1. Its second row carries the probability $2^{-35.54}$ of the distinguisher in the sandwich framework, and the note below the table carries the experimental probability $2^{-35.92}$ obtained by running the same distinguisher on the full cipher; the two values agree to within a factor of $2^{0.38}$, which Section 6.2 resolves into its two parts. Table 2 compares the key recovery attacks on FUTURE. A quartet consists of two encryption queries and two decryption queries, and is formed and tested with four 64-bit XOR operations. The data complexity of a row is therefore four times its number of quartets, and its time complexity is four times that number of XOR operations. The two columns carry the same exponent in different units, as in [JDC⁺26]. The memory requirement of both distinguishers is negligible. Reference [JDC⁺26] states its data complexity as 2^{48} , with its quartet count rounded up to 2^{46} . Applied to the parameters of the two 5-round characteristics of [JDC⁺26], the model of Section 3 returns optimal weights 14 and 16 and so reproduces the probabilities 2^{-14} and 2^{-16} reported there, so the rows of Table 1 are computed in the same setting.

This paper is organized as follows. Section 2 fixes the notation, the specification of FUTURE and the sandwich framework. Section 3 describes the exact bit-level model of

Table 1: Comparison with previous full-round distinguisher

Rounds	Probability	Quartets	Data (queries)	Time (XORs)	Ref.
10	$2^{-45.8}$	$2^{45.8}$	$2^{47.8}$	$2^{47.8}$	[JDC ⁺ 26]
10	$11 \cdot 2^{-39} = 2^{-35.54}$ [†]	$2^{35.54}$	$2^{37.54}$	$2^{37.54}$	Sections 4–5

[†] Experimental probability $2^{-35.92}$, from 341 right quartets over $2^{44.34}$ quartets on the full cipher (Section 6).

Table 2: Comparison with previous key recovery attacks

Rounds	Type	Data (queries)	Time	Memory	Ref.
9	Related-key	2^{59}	2^{69} Enc.	2^{59}	[JDC ⁺ 26]
10	Related-key	2^{48}	2^{112} Enc./Dec.	Negligible	[JDC ⁺ 26]
10	Related-key	2^{18}	2^{70} Enc./Dec.	2^{64}	[JDC ⁺ 26]
10	Related-key	$2^{51.05}$	2^{64} [†] Enc./Dec.	2^{64}	Section 6.3

[†] The optimum of the framework of [SZY⁺22, YSZ⁺24] at a memory within the codebook (Section 6.3).

related-key differential characteristics and gives the optimal weights over 3 to 6 rounds. Section 4 gives the full-round distinguisher and its optimality over all switching rounds. Section 5 computes its probability in closed form and establishes the optimality over all lower middle-layer differences. Section 6 measures that probability by running the distinguisher on the full cipher and gives the key recovery attack on all ten rounds. Finally, conclusions are drawn in Section 7.

2 Preliminaries

2.1 Notation

We consider four related keys

$$K^A, \quad K^B = K^A \oplus \Delta k, \quad K^C = K^A \oplus \nabla k, \quad K^D = K^A \oplus \Delta k \oplus \nabla k,$$

where $\Delta k = (\Delta k_0, \Delta k_1)$ and $\nabla k = (\nabla k_0, \nabla k_1)$ are the key differences of the two sub-ciphers. A 64-bit value is a 4×4 array of 4-bit cells and is written as 16 hexadecimal digits; cell n , $0 \leq n \leq 15$, lies in column $\lfloor n/4 \rfloor$ and row $n \bmod 4$, and the digits are listed in the order $n = 0, 1, \dots, 15$. For a difference δ we write δ_n for its n -th cell and $\text{supp } \delta = \{n : \delta_n \neq 0\}$ for the set of its active cells.

The full-round encryption is split as $E = E_1 \circ E_m \circ E_0$. Following the boomerang literature, Δ_{in} and Δ_{out} denote the input and output differences of E_0 and ∇_{in} and ∇_{out} those of E_1 , so that Δ_{out} is the upper boundary difference of the middle layer and ∇_{in} its lower boundary difference. As is standard in differential cryptanalysis, the probabilities below are expected probabilities under the assumption of independent random round keys, so that the probability space carries the four related keys along with the message. The dependence of a characteristic on the key is treated by Peyrin et al. [PT22], the passage from the fixed-key to the expected differential probability by Boura et al. [BDG25], and the fixed-key boomerang probability by Chang et al. [CHH⁺25]. For fixed $(\Delta k, \Delta_{\text{in}})$ and

$(\nabla k, \nabla_{\text{out}})$ define

$$p(\delta) = \Pr_{x,K} \left[E_0^{K^A}(x) \oplus E_0^{K^B}(x \oplus \Delta_{\text{in}}) = \delta \right], \quad (1)$$

$$q(\varepsilon) = \Pr_{y,K} \left[E_1^{K^A}(y) \oplus E_1^{K^C}(y \oplus \varepsilon) = \nabla_{\text{out}} \right], \quad (2)$$

and call $\text{supp } p$ and $\text{supp } q$ the *differential clusters* of the upper and the lower sub-cipher. Taking the clustering effect into account, the differential effects of the two sub-ciphers are

$$\hat{p}^2 = \sum_{\delta \in A} p(\delta)^2, \quad \hat{q}^2 = \sum_{\varepsilon \in B} q(\varepsilon)^2, \quad (3)$$

where $A \subseteq \text{supp } p$ and $B \subseteq \text{supp } q$ collect the members of the two clusters that are identified. The terms of (3) are non-negative, so any such choice gives a lower bound on the differential effect of the sub-cipher it belongs to. The probabilities of a single characteristic are written without a hat, as p and q . The *connection probability* of the middle layer is

$$r(\delta, \varepsilon) = \Pr_x \left[E_m^{-1}(E_m(x) \oplus \varepsilon) \oplus E_m^{-1}(E_m(x \oplus \delta) \oplus \varepsilon) = \delta \right]. \quad (4)$$

The sandwich framework treats E_0 , E_m and E_1 as independent, and under that treatment the probability of the distinguisher is $p^2 q^2 r$ for a single characteristic and $\hat{p}^2 \hat{q}^2 \bar{r}$ once the clusters are taken into account, with $\bar{r}$ as defined in Theorem 1. A characteristic is also called a trail, and its weight is $-\log_2$ of its probability.

2.2 Specification of FUTURE

FUTURE encrypts a 64-bit plaintext P under a 128-bit key $K = (K_0, K_1)$ in 10 rounds, numbered 0 to 9. Writing $x \lll s$ for the rotation of a 64-bit word to the left by s bits, the round keys are

$$RK[r] = \begin{cases} K_0 \lll 5(r/2), & r \text{ even,} \\ K_1 \lll 5((r-1)/2), & r \text{ odd,} \end{cases} \quad 0 \leq r \leq 10. \quad (5)$$

The whitening key $RK[0] = K_0$ is added to the plaintext. Round r then applies SubCell, MixColumn, ShiftRow and AddRoundKey with $RK[r+1]$, followed by the round constant $RC[r+1]$. Round 9 omits MixColumn, and $RK[10] = K_0 \lll 25$ is its round key. The four layers are as follows.

- **SubCell** applies the 4-bit S-box $S = (1, 3, 0, 2, 7, \text{e}, 4, \text{d}, 9, \text{a}, \text{c}, 6, \text{f}, 5, 8, \text{b})$ to every cell.
- **MixColumn** multiplies each column, that is each quadruple of cells $(s_{4c}, s_{4c+1}, s_{4c+2}, s_{4c+3})$, by the MDS matrix

$$M = \begin{pmatrix} 8 & 9 & 1 & 8 \\ 3 & 2 & 9 & 9 \\ 2 & 3 & 8 & 9 \\ 9 & 9 & 8 & 1 \end{pmatrix}$$

over $\text{GF}(2^4) = \mathbb{F}_2[x]/(x^4 + x + 1)$, a cell $\sum_i b_i 2^i$ standing for the field element $\sum_i b_i x^i$.

- **ShiftRow** permutes the cells by $s'_i = s_{13i \bmod 16}$, $0 \leq i \leq 15$; in terms of the array this rotates row j by j positions towards increasing column index.
- **AddRoundKey** adds the round key, after which the round constant is added.

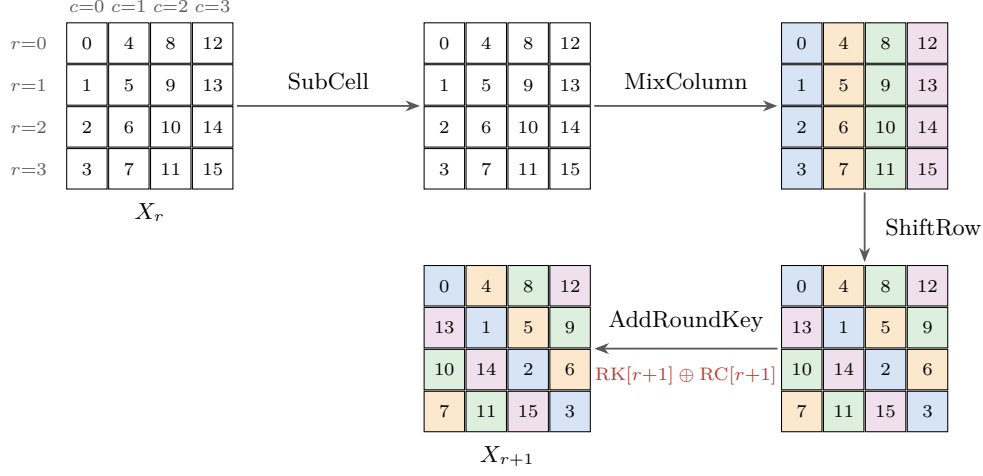


Figure 1: One round of FUTURE

Figure 1 shows one round. The state is a 4×4 array of 4-bit cells, with cell n in column $\lfloor n/4 \rfloor$ and row $n \bmod 4$, so four consecutive cells form one column. ShiftRow sends the content of cell n to cell $5n \bmod 16$, since $5 \cdot 13 \equiv 1 \pmod{16}$.

The rotation amounts in (5) are cumulative multiples of 5, and the round constant is added after AddRoundKey, as in the reference implementation of [GPS22]. Round constants do not affect the probability of an XOR differential characteristic and play no further role.

Two structural facts are used later. Round 9 has no MixColumn, so a ciphertext difference determines the output difference of the last SubCell layer directly through ShiftRow^{-1} . By (5) the relative rotation amounts of $RK[1], \dots, RK[4]$ and of $RK[7], \dots, RK[10]$ are both 0, 5, 5, 10, so rounds 6 to 9 are equivalent to rounds 0 to 3 under $\Delta k_0 \mapsto \Delta k_0 \lll (-15)$, $\Delta k_1 \mapsto \Delta k_1 \lll (-15)$.

2.3 Related-Key Boomerang Attacks and the Sandwich Framework

In the related-key boomerang attack [Wag99, BDK05] the adversary asks for the encryption of a plaintext pair with difference Δ_{in} under (K^A, K^B) . It shifts both ciphertexts by ∇_{out} and asks for the decryption of the shifted pair under (K^C, K^D) . The quartet is right when the two new plaintexts differ by Δ_{in} . For a random permutation this happens with probability 2^{-n} with $n = 64$.

The sandwich framework of Dunkelman et al. [DKS14] splits $E = E_1 \circ E_m \circ E_0$ and writes the probability of a right quartet as $p^2 q^2 r$, where p and q are the probabilities of the two characteristics and r is the connection probability (4) of the middle layer. The middle layer is where the two characteristics interact, and the dependency between them at the switching round is what r measures.

Definition 1 (Boomerang connectivity table [CHP⁺18]). For a permutation S of $\mathbb{F}_2^m$ and $a, b \in \mathbb{F}_2^m$, the boomerang connectivity table of S is

$$\text{BCT}_S(a, b) = \# \{x \in \mathbb{F}_2^m : S^{-1}(S(x) \oplus b) \oplus S^{-1}(S(x \oplus a) \oplus b) = a\}.$$

Table 3: Optimal weights of related-key differential characteristics of FUTURE

Rounds R	3	4	5	6
$W(R)$	2	4	14	24

When E_m is a single SubCell layer the connection probability factorizes over the cells,

$$r(\delta, \varepsilon) = \prod_{n=0}^{15} \frac{\text{BCT}_S(\delta_n, \varepsilon_n)}{16}. \quad (6)$$

Since $\text{BCT}_S(a, 0) = \text{BCT}_S(0, b) = 16$ for all a and b , a cell in which at most one of the two differences is active contributes a factor 1; in particular $r(\delta, \varepsilon) = 1$ whenever $\text{supp } \delta$ and $\text{supp } \varepsilon$ are disjoint, which is the ladder switch. Refinements of the BCT that describe switches over several rounds are given by Wang et al. [WP19] and by Song et al. [SQH19]; the distinguisher below uses a middle layer of one SubCell layer, for which (6) is exact.

3 Exact Model of Related-Key Differential Characteristics

3.1 Bit-Level Constraint Model with True DDT Weights

A related-key differential characteristic of FUTURE over R consecutive rounds is encoded in a constraint model whose variables are the 64 bits of each state difference, the 128 bits of $(\Delta k_0, \Delta k_1)$, and one weight variable per cell of each SubCell layer.

The SubCell layer is constrained cell by cell by a table of allowed assignments. The table lists the 96 entries of the support of the DDT of S together with the exact weight of each, namely $-\log_2$ of the transition probability. Over the DDT of S these weights take the two values 2 and 3, with 24 entries of weight 2 and 72 of weight 3. MixColumn and ShiftRow are expanded bitwise into XOR constraints. The round key difference of round r is the rotation of Δk_0 or Δk_1 by the amount given by (5), so that all rounds share the same 128 bits of key difference and the rotation amounts are the real ones. The objective function is the sum of the weight variables, that is the exact weight $-\log_2 p$ of the characteristic, and not the number of active S-boxes. The model is solved by CP-SAT of OR-Tools [PF25].

Let $W(R)$ denote the minimum of this objective function over R consecutive rounds, over all $(\Delta k, \Delta_{\text{in}})$ and all intermediate differences. Since a characteristic over $R+1$ rounds contains one over R rounds, W is non-decreasing:

$$W(R) \leq W(R+1) \quad \text{for all } R \geq 1. \quad (7)$$

3.2 Optimal Weights over Three to Six Rounds

Table 3 gives $W(R)$ for $3 \leq R \leq 6$. Each of the four values is returned by the solver with status OPTIMAL and a lower bound equal to the optimum, for both starting rounds $r_0 = 0$ and $r_0 = 1$. The round-by-round numbers of active cells and weights of the optimal characteristics are 0/0, 1/2, 0/0, 1/2 over 4 rounds; 3/6, 0/0, 2/5, 0/0, 1/3 over 5 rounds; and 4/8, 0/0, 2/4, 0/0, 3/6, 3/6 over 6 rounds. The rounds with no active cell are those in which the state difference equals the round key difference, which is what a related key makes available. A 4-round window admits two such rounds, and this is the reason for the jump from $W(4) = 4$ to $W(5) = 14$.

The same model applies to a characteristic whose key, input and output differences are prescribed. On the parameters of the two 5-round characteristics of [JDC⁺26] it returns optimal weight 14 for the upper characteristic, with round-by-round weights 6/0/6/0/2, and

optimal weight 16 for the lower one, with round-by-round weights 8/0/5/0/3. It therefore reproduces the probabilities 2^{-14} and 2^{-16} reported there. On the lower characteristic, with the full triple of key, plaintext and ciphertext differences fixed, the model also reproduces the output difference 1000000000000000 of that work. The weight 14 of the upper characteristic equals $W(5)$, so that characteristic is optimal over 5 rounds without any constraint on the key or the plaintext difference.

The weights 14 and 16 were verified experimentally by counting right pairs under the four related keys. Monte Carlo counting over 2^{22} plaintexts per key gives $2^{-14.130}$ and $2^{-16.046}$ for the two characteristics, and $2^{-15.923}$ over 2^{24} plaintexts for the lower one with the full triple of differences fixed, in agreement with the predicted 2^{-14} and 2^{-16} .

4 Full-Round Related-Key Boomerang Distinguisher

4.1 Joint Optimization of Both Sub-Ciphers and the Middle Layer

Let m be the switching round, so that

$$\begin{aligned} E_0 &= \text{rounds } 0 \text{ to } m - 1, \\ E_m &= \text{the SubCell layer of round } m, \\ E_1 &= \text{the rest of round } m \text{ and rounds } m + 1 \text{ to } 9. \end{aligned}$$

Then E_0 contains m SubCell layers and E_1 contains $9 - m$ of them. Writing w_0 and w_1 for the weights of the two trails and $w_{\text{bct}} = -\log_2 r$, the probability of the sandwich distinguisher satisfies

$$2w_0 + 2w_1 + w_{\text{bct}} = -\log_2(p^2 q^2 r), \quad (8)$$

so minimizing the left-hand side maximizes $p^2 q^2 r$.

The model of Section 3 is extended to the whole of E at once. The two sub-ciphers carry independent key differences Δk and ∇k , as the four related keys allow. The middle layer is constrained by a table of allowed assignments listing the 149 entries of the support of the BCT of S with their weights $-\log_2(\text{BCT}_S(a, b)/16)$, which range over 0.00 to 3.00. The identities $\text{BCT}_S(a, 0) = \text{BCT}_S(0, b) = 16$ are entries of that table, so the ladder switch is available to the solver rather than imposed by hand. The last round is treated separately because round 9 has no MixColumn and carries the whitening key $RK[10] = \nabla k_0 \lll 25$. The objective function is (8).

Figure 2 shows the resulting structure at the optimum $m = 5$, which is established in Section 4.2. The upper sub-cipher E_0 covers rounds 0 to 4 under the key difference Δk , the middle layer E_m is the SubCell layer of round 5, and the lower sub-cipher E_1 covers the rest of round 5 and rounds 6 to 9 under the key difference ∇k . The dashed arrow in the figure is the event whose probability is computed in Corollary 1.

4.2 Optimality over All Switching Rounds

In Table 4 the rows $m = 4$ and $m = 5$ are solved to optimality with lower bounds equal to the optima, and the other rows are bounded below by Table 3 together with the monotonicity (7).

Table 4 settles every switching round. For $m = 4$ and $m = 5$ the joint model is solved to optimality, giving 40.00 and 36.00. For the remaining values of m the bound follows from $w_{\text{bct}} \geq 0$ and Table 3. If $m \in \{3, 6\}$ then one sub-cipher covers 3 rounds and the other 6, so the objective function is at least $2W(3) + 2W(6) = 2 \cdot 2 + 2 \cdot 24 = 52$. If $m \leq 2$ or $m \geq 7$ then one sub-cipher covers at least 7 rounds, so by (7) the objective function is at least $2W(7) \geq 2W(6) = 48$. All of these exceed 36. Hence 2^{-36} is the optimum of $p^2 q^2 r$ over all switching rounds for a middle layer of one SubCell layer, and this uses only the

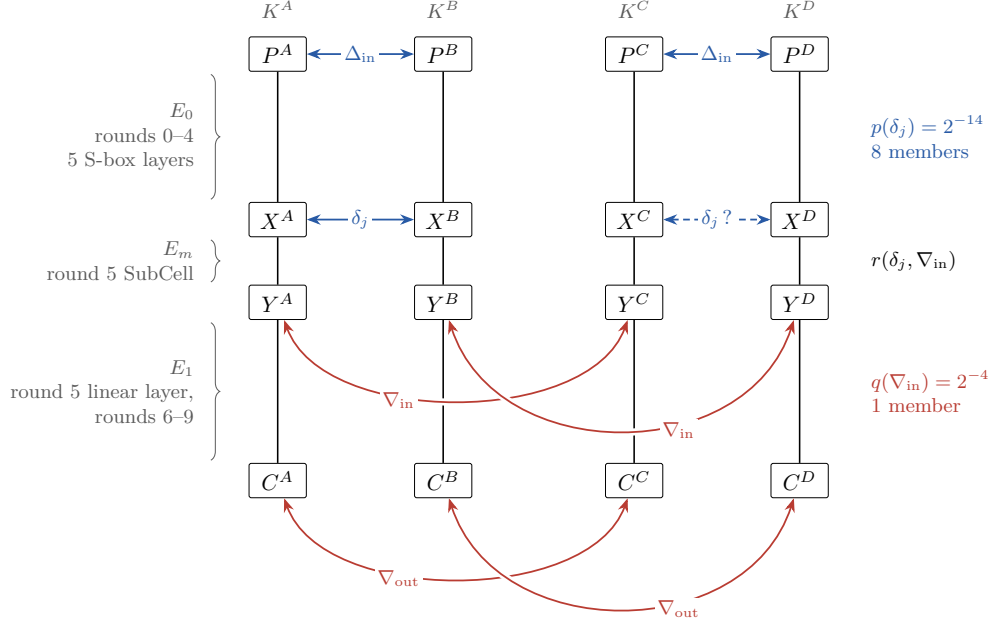


Figure 2: Full-round related-key boomerang distinguisher of FUTURE

Table 4: Objective function value for each switching round

m	Rounds of E_0	Rounds of E_1	$2w_0 + 2w_1 + w_{\text{bct}}$	Obtained from
≤ 2	≤ 2	≥ 7	≥ 48	$2W(7) \geq 2W(6)$
3	3	6	≥ 52	$2W(3) + 2W(6)$
4	4	5	40.00	OPTIMAL, bound 40.00
5	5	4	36.00	OPTIMAL, bound 36.00
6	6	3	≥ 52	$2W(6) + 2W(3)$
≥ 7	≥ 7	≤ 2	≥ 48	$2W(7) \geq 2W(6)$

optimality of $W(6) = 24$ and the monotonicity of W , not any search over seven or more rounds.

The weights are represented in the solver by integers, and the representation is chosen so that the value 36.00 is exact. Every DDT weight is 2 or 3 and is therefore represented without error. Every BCT weight is rounded down, so the modeled objective function never exceeds the true one and a lower bound on the modeled minimum is a lower bound on the true one. At a scale of 10^6 the rounding removes at most 8.0×10^{-6} across all 16 cells of the middle layer. In this representation the row $m = 5$ returns 36.00 with a matching lower bound, so every candidate has true objective function value at least 36. The parameters of Section 4 have $w_{\text{bct}} = 0$ exactly and attain 36. Every other row of Table 4 is at least 40 and so stands away from 36 by at least 4, which exceeds the rounding by six orders of magnitude.

4.3 Parameters of the Distinguisher

The optimum at $m = 5$ is attained by the following parameters.

key difference			state difference	
E_0	Δk_0	0034000000000000	Δ_{in}	003400000000a011
	Δk_1	0300004000018000	Δ_{out}	70010e0006500000
E_1	∇k_0	0006000011000040	∇_{in}	0c6000000000228c
	∇k_1	0000000002000000	∇_{out}	0022100080000c00

The weights are $w_0 = 14$ and $w_1 = 4$, so $p = 2^{-14}$ and $q = 2^{-4}$; the round-by-round states of both characteristics are listed in Appendix A. The active cells of Δ_{out} are $\{0, 3, 5, 9, 10\}$ and those of ∇_{in} are $\{1, 2, 12, 13, 14, 15\}$; the two sets are disjoint, so all sixteen factors in (6) are of the form $\text{BCT}_S(a, 0)$ or $\text{BCT}_S(0, b)$ and $r = 1$. The solver selects an upper trail of weight 14 over 5 rounds in preference to one of weight 4 over 4 rounds precisely because the objective function (8) charges w_{bct} in the same units as the trails. The single-characteristic probability is therefore

$$p^2 q^2 r = 2^{-28} \cdot 2^{-8} \cdot 1 = 2^{-36}.$$

5 Probability of the Distinguisher

The value $p^2 q^2 r$ counts one characteristic on each side, whereas a right quartet only requires the two pairs to carry the same difference at each boundary of the middle layer. Summing over the two clusters of boundary differences therefore refines it, and the clustering effect is well documented for other ciphers [DDV20, SWW18, LPS21, CMY⁺24]; here the members that dominate it are identified and their contribution is computed exactly. The evaluation below is the one supplied by the sandwich framework, in which the three parts of the cipher are treated as independent; Section 6 measures the probability of the same distinguisher directly and reports the agreement.

5.1 Probability Identity and Tightness Criterion

Theorem 1. *Let $\mu(\delta) = p(\delta)^2 / \hat{p}^2$ on A and $\nu(\varepsilon) = q(\varepsilon)^2 / \hat{q}^2$ on B , and let $\bar{r} = \mathbb{E}_{\mu \otimes \nu}[r]$. Then the probability of the related-key sandwich distinguisher over the two clusters,*

$$P = \sum_{\delta \in A} \sum_{\varepsilon \in B} p(\delta)^2 r(\delta, \varepsilon) q(\varepsilon)^2, \quad (9)$$

equals $\hat{p}^2 \bar{r} \hat{q}^2$. Consequently

- (a) $P \leq \hat{p}^2 \hat{q}^2$, with equality if and only if $r \equiv 1$ on $A \times B$;
- (b) $P \geq p(\delta^*)^2 r(\delta^*, \varepsilon^*) q(\varepsilon^*)^2$ for every $(\delta^*, \varepsilon^*) \in A \times B$;
- (c) $\hat{p}^2 \hat{q}^2 r(\delta^*, \varepsilon^*) > P$ if and only if $r(\delta^*, \varepsilon^*) > \bar{r}$, and in that case $\hat{p}^2 \hat{q}^2 r(\delta^*, \varepsilon^*)/P = r(\delta^*, \varepsilon^*)/\bar{r}$.

Proof. Writing $p(\delta)^2 q(\varepsilon)^2$ as $\hat{p}^2 \hat{q}^2 \mu(\delta) \nu(\varepsilon)$ turns (9) into $P = \hat{p}^2 \hat{q}^2 \sum_{\delta, \varepsilon} \mu(\delta) \nu(\varepsilon) r(\delta, \varepsilon) = \hat{p}^2 \hat{q}^2 \mathbb{E}_{\mu \otimes \nu}[r]$. Item (a) follows from $0 \leq r \leq 1$ and the monotonicity of the expectation, the equality condition being that $r = 1$ almost everywhere for $\mu \otimes \nu$. Item (b) is the restriction of the sum (9) to one term. Item (c) follows from $\hat{p}^2 \hat{q}^2 r(\delta^*, \varepsilon^*) - P = \hat{p}^2 \hat{q}^2 (r(\delta^*, \varepsilon^*) - \bar{r})$ together with $P = \hat{p}^2 \hat{q}^2 \bar{r}$. $\square$

Theorem 1 isolates the difference between $\hat{p}^2 \hat{q}^2 r$ and P : taking the connection probability to be the single value $r(\delta^*, \varepsilon^*)$ of the characteristic itself is the same as holding r constant on the whole cluster. When the characteristic sits at a position where $r = 1$, that constant gives $\hat{p}^2 \hat{q}^2$, which by (c) exceeds P by the factor $1/\bar{r}$.

Lemma 1. *Let $(X_1, K_1), \dots, (X_N, K_N)$ be independent, each a uniform plaintext together with a uniform quadruple of related keys, and set $D_t = E_0^{K_t^A}(X_t) \oplus E_0^{K_t^B}(X_t \oplus \Delta_{\text{in}})$ and $c_\delta = \#\{t : D_t = \delta\}$. Then for every weight function w ,*

$$\mathbb{E} \left[\frac{1}{\binom{N}{2}} \sum_{\delta} \binom{c_\delta}{2} w(\delta) \right] = \sum_{\delta} p(\delta)^2 w(\delta).$$

Proof. $\sum_{\delta} \binom{c_\delta}{2} w(\delta) = \sum_{s < t} w(D_s) \mathbf{1}[D_s = D_t]$. The D_t are independent and each distributed as p , so every one of the $\binom{N}{2}$ terms has expectation $\sum_{\delta} p(\delta)^2 w(\delta)$. Take expectations and use linearity. $\square$

With $w \equiv 1$ Lemma 1 gives an unbiased estimator of $\hat{p}^2$, and with $w = r(\cdot, \varepsilon)$ an unbiased estimator of the inner sum of (9). A collision pair is itself a sample of $p(\delta)^2$, so neither quantity requires the differential cluster to be enumerated or truncated, and the collision probability 2^{-64} of a random permutation does not interfere.

5.2 Differential Cluster of the Upper Sub-Cipher

Proposition 1. *For the parameters of Section 4, that is $\Delta k_0 = 0034000000000000$, $\Delta k_1 = 0300004000018000$ and $\Delta_{\text{in}} = 003400000000a011$, the differences that the upper sub-cipher reaches at weight 14 or less are exactly the eight of Table 5, each by a single characteristic, so each of them contributes 2^{-14} to $p(\delta_j)$; and E_1 maps $\nabla_{\text{in}} = 0c6000000000228c$ to ∇_{out} with a characteristic of weight 4. With A the eight differences of Table 5 and $B = \{\nabla_{\text{in}}\}$,*

$$\hat{p}^2 = 8 \cdot 2^{-28} = 2^{-25}, \quad \hat{q}^2 = 2^{-8}.$$

Proof. Fix Δk_0 , Δk_1 and Δ_{in} in the model of Section 3 over rounds 0 to 4 and leave Δ_{out} free. The solver returns optimal weight 14 with status OPTIMAL and lower bound 14, so no difference is reached at weight less than 14. Adding the constraint $w_0 \leq 14$ and excluding each solution in turn by a constraint on Δ_{out} terminates with status INFEASIBLE after eight solutions, which are the eight differences of Table 5. For each of them, excluding solutions by a constraint on the S-box output differences of the five rounds terminates with status INFEASIBLE after one solution. Every δ_j therefore carries a single characteristic of weight 14, which contributes 2^{-14} to $p(\delta_j)$, and the eight terms of (3) give $\hat{p}^2 = 8 \cdot 2^{-28} = 2^{-25}$. Applied to E_1 with ∇k_0 , ∇k_1 , ∇_{in} and ∇_{out} fixed, the same model returns optimal weight 4 with status OPTIMAL and lower bound 4, so $\hat{q}^2 = q(\nabla_{\text{in}})^2 = 2^{-8}$ for $B = \{\nabla_{\text{in}}\}$. $\square$

Table 5: Members of the differential cluster of the upper sub-cipher

j	δ_j	$\text{supp } \delta_j$	$\delta_j[15]$	$r(\delta_j, \nabla_{\text{in}})$
1	f0010d0006700009	$\{0, 3, 5, 9, 10, 15\}$	9	0
2	d001090006e00006	$\{0, 3, 5, 9, 10, 15\}$	6	1/8
3	8001030006200005	$\{0, 3, 5, 9, 10, 15\}$	5	1/8
4	2001040006900003	$\{0, 3, 5, 9, 10, 15\}$	3	0
5	300106000640000d	$\{0, 3, 5, 9, 10, 15\}$	d	1/8
6	60010c000680000e	$\{0, 3, 5, 9, 10, 15\}$	e	0
7	70010e0006500000	$\{0, 3, 5, 9, 10\}$	0	1
8	a001070006b0000a	$\{0, 3, 5, 9, 10, 15\}$	a	0

The last two columns of Table 5 record the value of cell 15 of each member and the connection probability against $\nabla_{\text{in}} = 0c600000000228c$, both of which are used in Section 5.3; member 7 is the characteristic of Section 4 itself.

Remark 1. Proposition 1 was confirmed by the collision counting of Lemma 1. Over 2^{26} plaintexts, namely 2^{24} under each of four keys, the differences δ occurring more than once are exactly the eight of Table 5, eight under each of the four keys separately. Together they account for 7.958×2^{-14} of the sample against an expectation of 8×2^{-14} with standard deviation 0.044, a deviation of 0.96 standard deviations. The remaining members of the two clusters occur at most once each in the 2^{26} samples and so contribute no collision pair at that sample size. The terms they add to (9) are positive. Section 6 measures the eight probabilities individually at a sample size at which each of them is resolved.

Proposition 2. *The union of the supports of the eight members is $S = \{0, 3, 5, 9, 10, 15\}$. The support of ∇_{in} is $\{1, 2, 12, 13, 14, 15\}$, which meets S only in cell 15, and $\nabla_{\text{in}}[15] = c$. Consequently, by (6),*

$$r(\delta_j, \nabla_{\text{in}}) = \frac{\text{BCT}_S(\delta_j[15], c)}{16} \quad (j \neq 7), \quad r(\delta_7, \nabla_{\text{in}}) = 1. \quad (10)$$

Proof. Outside cell 15 the active cells of δ_j and of ∇_{in} are disjoint, so the corresponding factors in (6) are all $\text{BCT}_S(a, 0) = 16$ or $\text{BCT}_S(0, b) = 16$. Member 7 is inactive in cell 15 as well, and $\text{BCT}_S(0, c) = 16$. $\square$

The entire cluster correction is therefore determined by a single column of the BCT, namely

$$\text{BCT}_S(a, c) = 16, 0, 2, 0, 4, 2, 2, 2, 4, 0, 0, 2, 4, 2, 0, 0 \quad \text{for } a = 0, \dots, f.$$

The seven members other than member 7 carry $\delta_j[15] \in \{9, 6, 5, 3, d, e, a\}$, of which $\{6, 5, d\}$ fall on the value 2 and $\{9, 3, e, a\}$ on the value 0.

Figure 3 shows the middle layer together with the cluster. Its upper boundary ranges over the eight members, whose supports are contained in S , and its lower boundary is the single difference ∇_{in} with support $\{1, 2, 12, 13, 14, 15\}$. The asterisk marks cell 15, the only cell active on both sides and hence the only cell whose factor in (6) differs from 1. That cell is active for the seven members $j \neq 7$ and inactive for the characteristic itself, which has $\delta_7[15] = 0$. The factor it contributes is the entry $\text{BCT}_S(\delta_j[15], c)$ of the column displayed above.

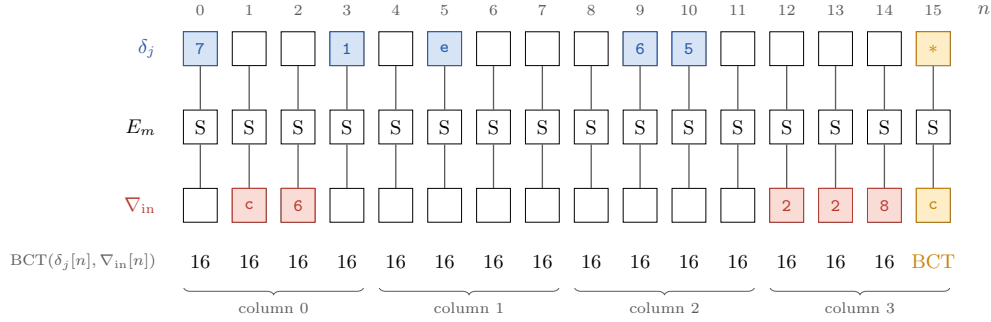


Figure 3: Middle layer of the distinguisher at round 5

5.3 Value of the Probability and the Data Complexity

Algorithm 1 states how the distinguisher is used, with Δ_{in} , ∇_{out} and the four related keys as in Section 4.

Algorithm 1 Related-key boomerang distinguisher for full-round FUTURE

Input: Oracle access to E^{K^A} , E^{K^B} , $(E^{K^C})^{-1}$ and $(E^{K^D})^{-1}$, where the four keys are related by $\Delta k_0, \Delta k_1$ and $\nabla k_0, \nabla k_1$ as in Section 4; the number of quartets N .

Output: *FUTURE* or *random permutation*.

```

1  $c \leftarrow 0$ ;
2 for  $i \leftarrow 1$  to  $N$  do
3   choose  $P_1 \in \mathbb{F}_2^{64}$  uniformly at random;
4    $P_2 \leftarrow P_1 \oplus \Delta_{\text{in}}$ ;
5    $C_1 \leftarrow E^{K^A}(P_1)$ ,  $C_2 \leftarrow E^{K^B}(P_2)$ ;
6    $C_3 \leftarrow C_1 \oplus \nabla_{\text{out}}$ ,  $C_4 \leftarrow C_2 \oplus \nabla_{\text{out}}$ ;
7    $P_3 \leftarrow (E^{K^C})^{-1}(C_3)$ ,  $P_4 \leftarrow (E^{K^D})^{-1}(C_4)$ ;
8   if  $P_3 \oplus P_4 = \Delta_{\text{in}}$  then  $c \leftarrow c + 1$ ;
9 end
10 if  $c \geq 1$  then return FUTURE;
11 return random permutation;
```

One iteration of the loop is one quartet. It makes two encryption queries in line 5 and two decryption queries in line 7, and performs four 64-bit XOR operations, one in line 4, two in line 6 and one in line 8; it holds the counter and the current quartet and nothing else.

Corollary 1. *In the sandwich framework the full-round related-key boomerang distinguisher of Section 4 has probability*

$$P = \hat{p}^2 \bar{r} \hat{q}^2 = 2^{-25} \cdot \frac{11}{64} \cdot 2^{-8} = 11 \cdot 2^{-39} = 2^{-35.5406...},$$

where $\bar{r} = \frac{1}{8} \sum_{j=1}^8 r(\delta_j, \nabla_{\text{in}}) = 11/64 = 2^{-2.5406}$. Algorithm 1 distinguishes it from a random permutation, for which the probability is 2^{-64} , with $2^{35.54}$ quartets, that is $2^{37.54}$ chosen plaintexts and adaptively chosen ciphertexts under four related keys, $2^{37.54}$ XOR operations and negligible memory.

Proof. Proposition 1 gives $\hat{p}^2 = 2^{-25}$ and $\hat{q}^2 = 2^{-8}$, and makes ν a point measure, so $\bar{r} = \sum_j \mu(\delta_j) r(\delta_j, \nabla_{\text{in}}) = \frac{1}{8} \sum_j r(\delta_j, \nabla_{\text{in}})$. By Proposition 2 and the column of the BCT displayed above, $\sum_j r(\delta_j, \nabla_{\text{in}}) = 1 + 3 \cdot \frac{2}{16} = \frac{11}{8}$, whence $\bar{r} = 11/64$. Substitute into

Table 6: Three evaluations of the probability of the distinguisher

How the probability is computed	P	Quartets	Queries
Single characteristic, $p^2 r q^2$	$2^{-36.00}$	$2^{36.00}$	$2^{38.00}$
Upper bound $\hat{p}^2 \hat{q}^2$	$2^{-33.00}$	$2^{33.00}$	$2^{35.00}$
$\hat{p}^2 \bar{r} \hat{q}^2$ (this work)	$11 \cdot 2^{-39} = 2^{-35.54}$	$2^{35.54}$	$2^{37.54}$

Theorem 1. Each iteration of Algorithm 1 makes four queries and performs four 64-bit XOR operations, so $N = 2^{35.54}$ quartets amount to $2^{37.54}$ queries and to $2^{37.54}$ XOR operations, and the memory is that of one quartet and one counter. $\square$

With $N = 2^{35.54}$ the expected value of c is 1 for FUTURE and $2^{35.54-64} = 2^{-28.46}$ for a random permutation, so Algorithm 1 returns *FUTURE* with probability $1 - e^{-1} \approx 0.63$ and returns it for a random permutation with probability at most $2^{-28.46}$.

The three factors have distinct meanings. The factor $\hat{p}^2 = 2^{-25}$ exceeds the single-characteristic value $p^2 = 2^{-28}$ by 2^3 , which is the gain of the clustering effect; the factor $\bar{r} = 2^{-2.54}$ is below the value $r(\delta_7, \nabla_{\text{in}}) = 1$ at the characteristic, which is the cost of the same effect; and the product of the two, $2^{0.46}$, is the net gain. The value is an exact rational number: once the eight members of probability 2^{-14} and the member ∇_{in} of probability 2^{-4} are known, every remaining quantity is a lookup in the BCT.

Corollary 2. *For the same parameters,*

$$2^{-36} = p^2 r q^2 \leq P = 11 \cdot 2^{-39} \leq \hat{p}^2 \hat{q}^2 = 2^{-33},$$

the left-hand side being the single-characteristic probability of Section 4 and the right-hand side the product of the two differential effects. The value P exceeds the left-hand side by $2^{0.46}$ and is smaller than the right-hand side by $2^{2.54}$.

Proof. Theorem 1(a) and (b), the latter with $(\delta^*, \varepsilon^*) = (\delta_7, \nabla_{\text{in}})$, for which $r(\delta^*, \varepsilon^*) = 1$. $\square$

Corollary 2 is the numerical instance of the criterion of Theorem 1(c): here $r(\delta_7, \nabla_{\text{in}}) = 1$ while $\bar{r} = 11/64$, so $\hat{p}^2 \hat{q}^2 r(\delta_7, \nabla_{\text{in}}) = 2^{-33}$ exceeds P by the factor $64/11 = 2^{2.54}$.

Table 6 places the three quantities of Corollary 2 side by side. In each row the number of queries is four times the number of quartets, since a quartet consists of two encryption and two decryption queries. The reference point is the $2^{-45.8}$ of [JDC⁺26] and the $2^{47.8}$ queries it corresponds to. Against it the third row of Table 6 improves the probability and the data complexity alike by $2^{10.26}$. Of that factor, a factor $2^{9.8}$ comes from the exact search with the objective function (8) over both sub-ciphers and the middle layer, and a factor $2^{0.46}$ from the cluster weighting $\hat{p}^2 \bar{r} \hat{q}^2$. The measured counterpart of that comparison is given in Section 6.

5.4 Optimality over All Lower Middle-Layer Differences

With the upper sub-cipher fixed, the full-round probability depends on ∇_{in} through two quantities only: the sum $\sum_j r(\delta_j, \nabla_{\text{in}})$ of the connection probabilities over the cluster, and the minimum weight $w_1(\nabla_{\text{in}})$ of the lower sub-cipher.

Proposition 3. *Take the upper sub-cipher of Section 4, let ∇_{in} range over all 64-bit differences, and set*

$$P(\nabla_{\text{in}}) = 2^{-28} \cdot \sum_{j=1}^8 r(\delta_j, \nabla_{\text{in}}) \cdot 2^{-2w_1(\nabla_{\text{in}})}. \quad (11)$$

Then $\max_{\nabla_{\text{in}}} P(\nabla_{\text{in}}) = 11 \cdot 2^{-39}$, attained at $\nabla_{\text{in}} = 0c6000000000228c$.

Proof. By Proposition 2 the factors of (6) outside the cell set S are identically 16, so $\sum_j r(\delta_j, \nabla_{\text{in}})$ depends on ∇_{in} only through its six cells in S and is a table lookup on 16^6 values. Without any constraint on ∇_{in} the model of Section 4 gives minimum lower trail weight 4, with status OPTIMAL and lower bound 4, so $w_1(\nabla_{\text{in}}) \geq 4$ for every ∇_{in} and

$$P(\nabla_{\text{in}}) \leq 2^{-36} \cdot \sum_{j=1}^8 r(\delta_j, \nabla_{\text{in}}). \quad (12)$$

Hence every ∇_{in} with $\sum_j r(\delta_j, \nabla_{\text{in}}) \leq 11/8$ satisfies $P(\nabla_{\text{in}}) \leq 11 \cdot 2^{-39}$. Exhausting the 16^6 restrictions of ∇_{in} to S leaves exactly 50 values with $\sum_j r(\delta_j, \nabla_{\text{in}}) > 11/8$. Solving the lower sub-cipher for each of the 50 returns status OPTIMAL with lower bound equal to the optimum in every case; the smallest w_1 among them is 7 and the largest resulting $P(\nabla_{\text{in}})$ is 2^{-41} . So on that side $P(\nabla_{\text{in}}) \leq 2^{-41} < 11 \cdot 2^{-39}$, and the maximum is attained where $\sum_j r(\delta_j, \nabla_{\text{in}}) = 11/8$ and $w_1 = 4$, that is at the value of Section 4. $\square$

Corollary 3. *Over the differences ∇_{in} whose support avoids S , all eight members satisfy $r(\delta_j, \nabla_{\text{in}}) = 1$, so that $\sum_j r(\delta_j, \nabla_{\text{in}}) = 8$; the smallest weight of the lower sub-cipher is $w_1 = 8$, and the largest value of (11) is $P = 2^{-28} \cdot 8 \cdot 2^{-16} = 2^{-41}$. Raising the sum of the connection probabilities from 11/8 to 8 gains a factor $2^{2.54}$, while raising the weight of the lower trail from 4 to 8 costs a factor 2^{-8} ; the product is $2^{-5.46}$.*

Proof. A difference with $\text{supp } \nabla_{\text{in}} \cap S = \emptyset$ has all sixteen factors of (6) of the form $\text{BCT}_S(a, 0)$ or $\text{BCT}_S(0, b)$, since the supports of the eight members are contained in S by Proposition 2; hence $r(\delta_j, \nabla_{\text{in}}) = 1$ for every j . Minimizing the weight of the lower sub-cipher under the constraint $\text{supp } \nabla_{\text{in}} \cap S = \emptyset$ returns $w_1 = 8$ with status OPTIMAL and lower bound 8. Substituting $\sum_j r(\delta_j, \nabla_{\text{in}}) = 8$ and $w_1 = 8$ into (11) gives 2^{-41} , and $8/(11/8) = 64/11 = 2^{2.54}$ while $2^{-2.8}/2^{-2.4} = 2^{-8}$. $\square$

The value of Section 4 therefore dominates both directions in which ∇_{in} can be varied. Corollary 3 covers the sparse direction, in which the cluster correction is removed entirely at the price of a heavier lower trail. In the dense direction the objective function alone already grows. Fixing the number of active cells of ∇_{in} and re-minimizing (8) gives the values 44, 42, 44, 36, 37 and 37.42 for 3, 4, 5, 6, 7 and 8 active cells. Each of these is returned with status OPTIMAL and a matching lower bound, and one and two active cells are infeasible. Six active cells, the value of Section 4, is the minimum. The quantity $P(\nabla_{\text{in}})$ of (11) is a probability and not the objective function: it combines the cluster weighting of the upper sub-cipher with the best characteristic of the lower sub-cipher. Proposition 3 therefore compares probabilities. On the differences at hand the two comparisons agree: the difference of Corollary 3 has $P = 2^{-41}$ and objective function value $2 \cdot 14 + 2 \cdot 8 + 0 = 44$, so it is below the value of Section 4 under (11) and under (8) alike.

Remark 2. Proposition 3 is the optimality over ∇_{in} with the upper sub-cipher fixed, and Table 4 is the joint optimality of the two sub-ciphers and the middle layer over all switching rounds in the objective function (8). Together they place the parameters of Section 4 at the optimum of the objective function and at the optimum of (11) alike. That coincidence is a property of these parameters: in general a characteristic of slightly larger weight combined with a sparser ∇_{in} retains more of the cluster gain, so the two optima need not agree.

6 Experimental Verification

Algorithm 1 was executed on the full cipher over $2^{44.34}$ quartets, some $2^{8.80}$ times the data the distinguisher itself requires, so that its probability is measured from several hundred right quartets rather than computed.

Table 7: Measured and predicted values of the three parts

j	δ_j	$\delta_j[15]$	$2^{14}p(\delta_j)$	$\text{BCT}_S(\delta_j[15], c)$	$16 r(\delta_j, \nabla_{\text{in}})$
1	f0010d0006700009	9	0.9994	0	0.0000
2	d001090006e00006	6	0.9998	2	1.9984
3	8001030006200005	5	0.9999	2	1.9985
4	2001040006900003	3	0.9999	0	0.0000
5	300106000640000d	d	1.0000	2	2.0001
6	60010c000680000e	e	1.0003	0	0.0000
7	70010e0006500000	0	1.0004	16	16.0000
8	a001070006b0000a	a	0.9995	0	0.0000

6.1 Probability of the Full-Round Distinguisher

The cipher, its key schedule and Algorithm 1 were implemented in C with the parameters of Section 4. The implementation passes three checks: decryption inverts encryption on random pairs of a key and a plaintext; the routines for E_0 , E_m and E_1 compose into the full cipher at both boundaries of the middle layer; and the two characteristics of Appendix A occur at their predicted probabilities, as Table 7 below records.

Every quartet takes its randomness from a stream determined by its index, and a fresh quadruple of related keys is drawn every 2^{20} quartets. The whole experiment covers about $2^{24.3}$ such quadruples, so what it measures is the probability averaged over the key, which is the quantity that Corollary 1 predicts. The implementation forms and tests an average of 9.7×10^7 quartets per second on an ordinary personal computer, so the whole experiment amounts to 63.9 hours of run time. It uses no dedicated hardware and no server.

Over $N = 2.2235 \times 10^{13} = 2^{44.34}$ quartets, $c = 341$ quartets were right, that is

$$\hat{P} = \frac{341}{N} = 2^{-35.9243},$$

with a relative standard deviation $1/\sqrt{341} = 5.4\%$ and a 95% confidence interval $[2^{-36.086}, 2^{-35.779}]$. A random permutation yields $2^{44.34-64} = 2^{-19.66}$ right quartets in expectation over the same data, so the observed count separates the full cipher from a random permutation. The value $11 \cdot 2^{-39}$ of Corollary 1 predicts 444.9 right quartets, a factor of $2^{0.38}$ from the observed count. Section 6.2 resolves that factor into its two parts.

6.2 The Three Parts Measured Separately

Each factor of Corollary 1 admits its own measurement, at a cost far below that of the distinguisher itself. Over 2^{38} plaintexts, again with the four related keys redrawn every 2^{20} of them, the following three quantities were counted: the probability $p(\delta_j)$ with which E_0 maps Δ_{in} to the member δ_j ; the probability $r(\delta_j, \nabla_{\text{in}})$ with which the middle layer preserves δ_j when ∇_{in} is imposed on its output; and the probability $q(\nabla_{\text{in}})$ with which E_1 maps ∇_{in} to ∇_{out} .

Table 7 collects the outcome. The fourth column would be 1 if $p(\delta_j)$ were exactly 2^{-14} , and the sixth would equal the fifth if $r(\delta_j, \nabla_{\text{in}})$ were exactly the entry of the BCT. For each of the eight members the deviation is at most 2×10^{-3} in both columns. Conditioned on E_0 producing δ_j , the value taken by cell 15 of the state entering the middle SubCell layer is uniform over its sixteen values up to 0.2%, for every j . That uniformity is the hypothesis under which the BCT applies. For the lower sub-cipher the count gives $q(\nabla_{\text{in}}) = 2^{-3.9996}$, against the value 2^{-4} of the characteristic.

Assembling the measured factors,

$$\sum_{j=1}^8 p(\delta_j)^2 r(\delta_j, \nabla_{\text{in}}) = 2^{-27.5399}, \quad q(\nabla_{\text{in}})^2 = 2^{-7.9992},$$

against the closed forms $11 \cdot 2^{-31} = 2^{-27.5406}$ and 2^{-8} , and their product is $2^{-35.5391}$, which is the value of Corollary 1 to four digits. The three parts are thus each confirmed at their predicted values, and the factor $2^{0.38}$ between them and the count of Section 6.1 comes from the three-part independence of the sandwich framework.

A third measurement isolates where that factor arises. The quantity

$$\sum_{j=1}^8 p(\delta_j)^2 r(\delta_j, \nabla_{\text{in}})$$

can be measured with the middle layer intact but the lower sub-cipher replaced by the condition that the difference at its input is ∇_{in} : draw a plaintext P_1 uniformly, form $P_2 = P_1 \oplus \Delta_{\text{in}}$, compute $x_3 = S^{-1}(S(E_0^{K^A}(P_1)) \oplus \nabla_{\text{in}})$ and $x_4 = S^{-1}(S(E_0^{K^B}(P_2)) \oplus \nabla_{\text{in}})$, and test whether $(E_0^{K^C})^{-1}(x_3) \oplus (E_0^{K^D})^{-1}(x_4) = \Delta_{\text{in}}$, where $K^C = K^A \oplus \nabla k$ and $K^D = K^B \oplus \nabla k$. Over 3×10^{11} samples, again with the four related keys redrawn every 2^{20} of them, this count gives

$$\sum_{j=1}^8 p(\delta_j)^2 r(\delta_j, \nabla_{\text{in}}) = 2^{-28.0281},$$

with 1096 hits, a relative standard deviation 3.0%, and a 95% confidence interval $[2^{-28.1162}, 2^{-27.9452}]$. This measurement resolves the dependency that remains between the three parts of the sandwich framework once the middle layer accounts for the switching round. Between the upper sub-cipher and the middle layer it amounts to a factor of $2^{0.49}$, which the sample size separates from 1 by 11.2 standard deviations. Between the two sub-ciphers it amounts to a factor of $2^{0.10}$ in the opposite direction. The two together account for the $2^{0.38}$ of Section 6.1.

Remark 3. Read as a complexity, the experimental probability $2^{-35.92}$ puts Algorithm 1 at $2^{35.92}$ quartets, that is $2^{37.92}$ queries and $2^{37.92}$ XOR operations, and keeps the improvement over the $2^{-45.8}$ of [JDC⁺26] at a factor of $2^{9.88}$ in probability, in data and in time.

The same factor separates the two distinguishers in the cost of running them. A probability is measured from a count of right quartets, and a distinguisher of probability P returns 341 of them over $341/P$ quartets. At $2^{-35.54}$ that is the $2^{44.34}$ quartets counted above, and at $2^{-45.8}$ it is $341 \cdot 2^{45.8} = 2^{54.21}$ quartets. At the rate of Section 6.1, $2^{44.34}$ quartets take 63.9 hours of run time on an ordinary personal computer, and $2^{54.21}$ quartets take about 6.8 years on that same computer. The distinguisher of this paper is therefore a practical one: a single ordinary personal computer runs it on the full cipher and counts its right quartets within hours.

6.3 Key Recovery

The exact model also yields a key recovery attack on the full cipher. Reference [JDC⁺26] attacks all ten rounds by taking an 8-round boomerang distinguisher of probability 2^{-16} and extending it by two rounds of key guessing within the unified framework of [SZY⁺22, YSZ⁺24], obtaining 2^{18} data, 2^{70} time and 2^{64} memory. We show that the exact model and the structure of FUTURE's key schedule together reduce the time complexity to 2^{64} , the optimum of the framework at a memory that stays within the codebook. Table 2 lists the attacks on the full cipher.

The unified framework [YSZ⁺24, Appendix A.2] decomposes the cipher as $E = E_1 \circ E_m \circ E_0$, builds a boomerang distinguisher over $E_m \circ E_0$ and extends it by E_1 via key guessing. It computes the data, time and memory complexities as functions of the distinguisher probability, the number of rounds in E_1 , and the numbers of guessed key bits and bit conditions that E_1 supplies. For FUTURE, round 9 has no MixColumn, so the pattern of active cells is fixed before any key is guessed. The output difference ∇_{out} of an 8-round distinguisher determines which cells of the ciphertext are active, and those cells alone determine which key bits of $RK[9]$ and $RK[10]$ enter the guess. If ∇_{out} occupies exactly c columns with one active cell in each, the MixColumn of round 8 makes all four cells of such a column active and round 9 only permutes them. Rounds 8 and 9 therefore activate exactly $4c$ cells of the ciphertext, the number of bit conditions is $r_f = 16c$ and the number of guessed key bits is $m_f = 32c$. Both are determined by c , the column count of ∇_{out} .

Our exact model gives the optimal 8-round related-key boomerang distinguisher over rounds 0 to 7 for each value of c . For $c = 1$, the best distinguisher under the constraint that ∇_{out} occupies one column has probability 2^{-24} and yields $m_f = 32$ key bits, so at most $h = 32$ bits of advantage are available and the remaining $2^{128-32} = 2^{96}$ master key bits must be searched exhaustively. For $c = 3$, the counter table of the framework has size $2^{m_f-m'_f} = 2^{96}$ when no key bits are guessed ($m'_f = 0$), exceeding the codebook. For $c = 2$, we obtain an 8-round distinguisher of probability 2^{-20} with $\nabla_{\text{out}} = 000000b000006000$, whose active cells 6 and 12 occupy columns 1 and 3 with one active cell in each, giving $r_f = 32$ and $m_f = 64$. The 64 key bits fall in two disjoint halves of the master key: $RK[9] = K_1 \lll 20$ supplies 32 bits of K_1 and $RK[10] = K_0 \lll 25$ supplies 32 bits of K_0 , so the advantage h can reach its maximum 64 and the remaining master key bits number $2^{128-64} = 2^{64}$.

The framework formulas [YSZ⁺24, Appendix A.2] with E_b empty ($m_b = r_b = r_b^* = m'_b = t = 0$) and no key guessing ($m'_f = 0$) give

$$\begin{aligned} D_0 &= 2s/P^2, & D_1 &= \min\{D_0 \cdot 2^{r_f}, 2^{n+1}\}, & D &= D_0 + D_1, \\ M &= D_0 + D_1 + 2^{m_f}, & T &= D_0 + D_1 + 2^{m'_f} D_0 \cdot 2^{r_f^*} + 2^{m'_f} D_0 \cdot 2^{2r_f^*-n} \varepsilon + 2^{k-h}, \end{aligned}$$

where s is the expected number of right quartets, $n = 64$, $k = 128$, ε is the cost of processing one candidate quartet, and $r_f^* = r_f - r'_f$ is the number of bit conditions available after guessing m'_f key bits. With $c = 2$, $P^2 = 2^{-20}$, $s = 4$, $m'_f = 0$, $r_f = r_f^* = 32$ and $h = 64$, we obtain

$$D_0 = 2^{23}, \quad D = 2^{55}, \quad M = 2^{64}, \quad T = 2^{55} + 2^{55} + 2^{23}\varepsilon + 2^{64}.$$

A quartet that passes the $r_f = 32$ bit conditions suggests $2^{m_f-r_f} = 2^{32}$ values of the $m_f = 64$ guessed key bits, so $\varepsilon = 2^{32}$ and the third term of T equals 2^{55} . All three leading terms of T are thus 2^{55} , and the time is $T = 2^{64}$, dominated by the exhaustive search over the 64 unrecovered master key bits.

Direct measurement over 2^{25} quartets under eight uniformly drawn master keys gives 496 right quartets and probability $2^{-16.05}$, so the distinguisher probability exceeds the single-characteristic value 2^{-20} by a factor of $2^{3.95}$. With the experimental probability, the data complexity becomes $D = 2^{51.05}$ while the time and memory remain 2^{64} each.

The counter table in the unified framework has size $2^{t+m_b+m_f-m'_b-m'_f}$, and the advantage h is bounded by the same exponent [YSZ⁺24, Appendix A.2, step (iv)]. With $t = m_b = m'_b = m'_f = 0$ and $m_f = 64$ that bound is 64, which the choice $c = 2$ attains. It follows that $M \geq 2^h$ and $T \geq 2^{k-h}$, so

$$M \cdot T \geq 2^h \cdot 2^{k-h} = 2^k.$$

For FUTURE this gives $M \cdot T \geq 2^{128}$, so at any memory that stays within the 2^{64} codebook the optimal time within this framework is 2^{64} , and our attack attains it. Compared to the

attack of [JDC⁺26] at 2^{18} data, 2^{70} time and 2^{64} memory, the time complexity improves by a factor of 2^6 and reaches the optimum of the framework.

7 Conclusion

We have given a related-key boomerang distinguisher over all 10 rounds of FUTURE with probability $\hat{p}^2 \hat{r} \hat{q}^2 = 11 \cdot 2^{-39} \approx 2^{-35.54}$ and $2^{37.54}$ data, improving the best previously known such distinguisher by a factor of $2^{10.26}$ in probability and by the same factor in data and in time. Two ingredients produce the improvement: an exact bit-level model in which the objective function is the probability weight over both sub-ciphers and the middle layer jointly, and a cluster weighting available in closed form. The closed form is available because the whole correction is determined by a single column of the BCT. The objective function value 36 is optimal over all switching rounds, and the lower middle-layer difference maximizes the probability over all 64-bit differences once the upper sub-cipher is fixed. Running the distinguisher over $2^{44.34}$ quartets gives 341 right quartets and an experimental probability of $2^{-35.92}$, an improvement of $2^{9.88}$ over the same reference. That count of right quartets is reached in 63.9 hours of run time on an ordinary personal computer, where the previously available probability requires about 6.8 years on the same computer. The full-round related-key boomerang distinguisher of FUTURE thereby becomes a practical one, run end to end within hours on a single ordinary personal computer. The data also resolve the dependency that remains between the three parts of the sandwich framework: a factor of $2^{0.49}$ between the upper sub-cipher and the middle layer, separated from 1 by 11.2 standard deviations, and a factor of $2^{0.10}$ between the two sub-ciphers in the opposite direction. The two together account for the $2^{0.38}$ that separates the experimental value from the closed-form one.

The same model also drives a key recovery attack on the full cipher. Placing an 8-round distinguisher with a two-column output difference inside the unified framework of [SZY⁺22, YSZ⁺24] attains $2^{51.05}$ data, 2^{64} time and 2^{64} memory. This improves the time complexity of the corresponding attack of [JDC⁺26] by a factor of 2^6 and attains the optimum of the framework at any memory within the codebook.

The model and the closed-form evaluation open two further directions. First, a middle layer of two rounds becomes accessible once the entries of the extended BCT of [DDV20] are built into the constraint model as a second constraint table, which replaces the sparsity of the boundary differences by the connection probability of two rounds itself. Second, the 8-round boomerang distinguisher of [JDC⁺26] uses a single output difference, so replacing its statistic by a log-likelihood ratio over the full support of the output distribution should reduce its data complexity further; the collision counting of Lemma 1 already supplies that distribution.

A Round-by-Round States of the Two Characteristics

The two characteristics behind the parameters of Section 4 are given round by round. For the upper one, let $\Delta RK[r]$ be the difference of $RK[r]$ under Δk , let Δx_r be the difference at the input of the SubCell layer of round r and Δy_r the difference at its output, so that

$$\Delta x_0 = \Delta_{\text{in}} \oplus \Delta RK[0], \quad \Delta x_r = \text{ShiftRow}(\text{MixColumn}(\Delta y_{r-1})) \oplus \Delta RK[r]$$

for $1 \leq r \leq 5$, and $\Delta x_5 = \Delta_{\text{out}}$. The lower characteristic uses ∇ in place of Δ and starts from $\nabla x_6 = \text{ShiftRow}(\text{MixColumn}(\nabla_{\text{in}})) \oplus \nabla RK[6]$; round 9 has no MixColumn, so it ends with $\nabla_{\text{out}} = \text{ShiftRow}(\nabla y_9) \oplus \nabla RK[10]$. The weight of a round is the sum of $-\log_2(\text{DDT}_S(a, b)/16)$ over its cells.

Table 8: Round-by-round states of the upper characteristic

r	$\Delta RK[r]$	Δx_r	Δy_r	Weight
0	0034000000000000	000000000000a011	0000000000005029	6
1	0300004000018000	0000000000000000	0000000000000000	0
2	0680000000000000	0680000000000000	07c0000000000000	5
3	6000080000300000	0000000000000000	0000000000000000	0
4	d000000000000000	d000000000000000	b000000000000000	3

Table 9: Round-by-round states of the lower characteristic

r	$\nabla RK[r]$	∇x_r	∇y_r	Weight
6	0000088000200003	0000000000000000	0000000000000000	0
7	0000010000000000	0000010000000000	0000020000000000	2
8	0001100004000060	0000000000000000	0000000000000000	0
9	0000200000000000	0000200000000000	0000100000000000	2

In rounds 1 and 3 of Table 8 the difference reaching the round equals $\Delta RK[r]$, so the SubCell layer is crossed with a zero difference and the round costs nothing. The three active rounds carry 3, 2 and 1 active cells and weights 6, 5 and 3, so $w_0 = 14$ over 5 rounds. The middle-layer upper boundary is $\Delta_{\text{out}} = \text{ShiftRow}(\text{MixColumn}(\Delta y_4)) \oplus \Delta RK[5]$ with $\Delta RK[5] = 000100000600000c$.

The linear image $\text{ShiftRow}(\text{MixColumn}(\nabla_{\text{in}}))$ of the middle-layer lower boundary equals $\nabla RK[6]$, so round 6 of Table 9 is crossed with a zero difference, and so is round 8. The two active rounds carry one active cell each, so $w_1 = 4$ over 4 rounds, and $\nabla_{\text{out}} = \text{ShiftRow}(\nabla y_9) \oplus \nabla RK[10]$ with $\nabla RK[10] = 0022000080000c00$.

Table 10 lists the eight active S-boxes of the two characteristics. Their weights sum to 14 for the upper one and to 4 for the lower one, which gives $p = 2^{-14}$ and $q = 2^{-4}$. With $\text{supp } \Delta_{\text{out}} = \{0, 3, 5, 9, 10\}$ and $\text{supp } \nabla_{\text{in}} = \{1, 2, 12, 13, 14, 15\}$ disjoint, the middle layer has $r = 1$, and the objective function (8) takes the value $2 \cdot 14 + 2 \cdot 4 + 0 = 36$ of Table 4.

Table 10: Active S-box transitions of the two characteristics

Characteristic	r	Cell	$a \rightarrow b$	$\text{DDT}_S(a, b)$	Weight
upper	0	12	a $\rightarrow$ 5	4	2
upper	0	14	1 $\rightarrow$ 2	4	2
upper	0	15	1 $\rightarrow$ 9	4	2
upper	2	1	6 $\rightarrow$ 7	2	3
upper	2	2	8 $\rightarrow$ c	4	2
upper	4	0	d $\rightarrow$ b	2	3
lower	7	5	1 $\rightarrow$ 2	4	2
lower	9	4	2 $\rightarrow$ 1	4	2

References

- [BDG25] Christina Boura, Patrick Derbez, and Baptiste Germon. Extending the quasidifferential framework: From fixed-key to expected differential probability. *IACR Transactions on Symmetric Cryptology*, 2025(1):515–541, 2025. doi:10.46586/tosc.v2025.i1.515-541.
- [BDK05] Eli Biham, Orr Dunkelman, and Nathan Keller. Related-key boomerang and rectangle attacks. In *Advances in Cryptology – EUROCRYPT 2005*, pages 507–525. Springer, 2005. doi:10.1007/11426639_30.
- [BL23] Augustin Bariant and Gaëtan Leurent. Truncated boomerang attacks and application to AES-based ciphers. In *Advances in Cryptology – EUROCRYPT 2023*, pages 3–35. Springer, 2023. doi:10.1007/978-3-031-30634-1_1.
- [CHH⁺25] Chengcheng Chang, Hosein Hadipour, Kai Hu, Muzhou Li, and Meiqin Wang. Mix-basis geometric approach to boomerang distinguishers. *IACR Transactions on Symmetric Cryptology*, 2025(3):693–728, 2025. doi:10.46586/tosc.v2025.i3.693-728.
- [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. doi:10.1007/978-3-319-78375-8_22.
- [CMY⁺24] Ting Cui, Yiming Mao, Yang Yang, Yi Zhang, Jiyan Zhang, and Chenhui Jin. Congruent differential cluster for binary SPN ciphers. *IEEE Transactions on Information Forensics and Security*, 19:2385–2397, 2024. doi:10.1109/TIFS.2024.3350374.
- [dCK26] Gabriel de Carvalho and Luis Kowada. Improved biclique cryptanalysis of the lightweight cipher FUTURE. *Journal of the Brazilian Computer Society*, 32(1):539–554, 2026. doi:10.5753/jbcs.2026.5390.
- [DDG⁺24] François Delobel, Patrick Derbez, Arthur Gontier, Loïc Rouquette, and Christine Solnon. A CP-based automatic tool for instantiating truncated differential characteristics. In *Progress in Cryptology – INDOCRYPT 2023*, pages 247–268. Springer, 2024. doi:10.1007/978-3-031-56232-7_12.
- [DDV20] Stéphanie Delaune, Patrick Derbez, and Mathieu Vavrille. Catching the fastest boomerangs: Application to SKINNY. *IACR Transactions on Symmetric Cryptology*, 2020(4):104–129, 2020. doi:10.46586/tosc.v2020.i4.104-129.
- [DEFN22] Patrick Derbez, Marie Euler, Pierre-Alain Fouque, and Phuong Hoa Nguyen. Revisiting related-key boomerang attacks on AES using computer-aided tool. In *Advances in Cryptology – ASIACRYPT 2022*, pages 68–88. Springer, 2022. doi:10.1007/978-3-031-22969-5_3.
- [DG26] A. K. Das and A. Gautam. Optimizing SPN full key recovery via MILP-aided trail selection. In *Proceedings of the 23rd International Conference on Security and Cryptography – SECRYPT 2026*, pages 309–316. SCITEPRESS, 2026. doi:10.5220/0015075000004103.
- [DKS14] Orr Dunkelman, Nathan Keller, and Adi Shamir. A practical-time related-key attack on the KASUMI cryptosystem used in GSM and 3G telephony. *Journal of Cryptology*, 27(4):824–849, 2014. doi:10.1007/s00145-013-9154-9.

- [GPS22] Kishan Chand Gupta, Sumit Kumar Pandey, and Susanta Samanta. FUTURE: A lightweight block cipher using an optimal diffusion matrix. In *Progress in Cryptology – AFRICACRYPT 2022*, pages 28–52. Springer, 2022. doi:[10.1007/978-3-031-17433-9_2](https://doi.org/10.1007/978-3-031-17433-9_2).
- [HBS21] Hosein Hadipour, Nasour Bagheri, and Ling Song. Improved rectangle attacks on SKINNY and CRAFT. *IACR Transactions on Symmetric Cryptology*, 2021(2):140–198, 2021. doi:[10.46586/tosc.v2021.i2.140-198](https://doi.org/10.46586/tosc.v2021.i2.140-198).
- [İS23] Murat Burhan İlter and Ali Aydın Selçuk. MILP-aided cryptanalysis of the FUTURE block cipher. In *Innovative Security Solutions for Information Technology and Communications – SecITC 2022*, pages 153–167. Springer, 2023. doi:[10.1007/978-3-031-32636-3_9](https://doi.org/10.1007/978-3-031-32636-3_9).
- [JDC⁺26] Amit Jana, Smita Das, Ayantika Chatterjee, Debdeep Mukhopadhyay, and Yu Sasaki. Related-key cryptanalysis of FUTURE: The full round distinguishing attack. In *Applied Cryptography and Network Security – ACNS 2026*, pages 283–313. Springer, 2026. Cryptology ePrint Archive, Paper 2024/1614. doi:[10.1007/978-3-032-32578-5_11](https://doi.org/10.1007/978-3-032-32578-5_11).
- [LPS21] Gaëtan Leurent, Clara Pernot, and André Schrottenloher. Clustering effect in Simon and Simeck. In *Advances in Cryptology – ASIACRYPT 2021*, pages 272–302. Springer, 2021. doi:[10.1007/978-3-030-92062-3_10](https://doi.org/10.1007/978-3-030-92062-3_10).
- [LS19] Yunwen Liu and Yu Sasaki. Related-key boomerang attacks on GIFT with automated trail search including BCT effect. In *Information Security and Privacy – ACISP 2019*, pages 555–572. Springer, 2019. doi:[10.1007/978-3-030-21548-4_30](https://doi.org/10.1007/978-3-030-21548-4_30).
- [LZL23] H. Lin, J. Zou, and J. Li. The differential meet-in-the-middle attack on FUTURE and CRAFT. In *Proceedings of the 2023 13th International Conference on Communication and Network Security – CNS 2023*, pages 151–158. ACM, 2023. doi:[10.1145/3638782.3638805](https://doi.org/10.1145/3638782.3638805).
- [MRSA24] Sandip Kumar Mondal, Mostafizar Rahman, Santanu Sarkar, and Avishek Adhikari. Yoyo cryptanalysis on FUTURE. *International Journal of Applied Cryptography*, 4(3/4):238–249, 2024. doi:[10.1504/IJACT.2024.138453](https://doi.org/10.1504/IJACT.2024.138453).
- [PF25] Laurent Perron and Vincent Furnon. OR-Tools, version 9.15. Google, 2025. URL: <https://developers.google.com/optimization/>.
- [PT22] Thomas Peyrin and Quan Quan Tan. Mind your path: On (key) dependencies in differential characteristics. *IACR Transactions on Symmetric Cryptology*, 2022(4):179–207, 2022. doi:[10.46586/tosc.v2022.i4.179-207](https://doi.org/10.46586/tosc.v2022.i4.179-207).
- [RDMA24] Himadry Sekhar Roy, Prakash Dey, Sandip Kumar Mondal, and Avishek Adhikari. Cryptanalysis of full round FUTURE with multiple biclique structures. *Peer-to-Peer Networking and Applications*, 17(1):397–409, 2024. doi:[10.1007/s12083-023-01600-y](https://doi.org/10.1007/s12083-023-01600-y).
- [SQH19] Ling Song, Xianrui Qin, and Lei Hu. Boomerang connectivity table revisited: Application to SKINNY and AES. *IACR Transactions on Symmetric Cryptology*, 2019(1):118–141, 2019. doi:[10.46586/tosc.v2019.i1.118-141](https://doi.org/10.46586/tosc.v2019.i1.118-141).
- [SS23] André Schrottenloher and Marc Stevens. Simplified modeling of MITM attacks for block ciphers: New (quantum) attacks. *IACR Transactions on Symmetric Cryptology*, 2023(3):146–183, 2023. doi:[10.46586/tosc.v2023.i3.146-183](https://doi.org/10.46586/tosc.v2023.i3.146-183).

- [SWW18] Ling Sun, Wei Wang, and Meiqin Wang. More accurate differential properties of LED64 and Midori64. *IACR Transactions on Symmetric Cryptology*, 2018(3):93–123, 2018. doi:[10.13154/tosc.v2018.i3.93-123](https://doi.org/10.13154/tosc.v2018.i3.93-123).
- [SZY⁺22] Ling Song, Nana Zhang, Qianqian Yang, Danping Shi, Jiahao Zhao, Lei Hu, and Jian Weng. Optimizing rectangle attacks: A unified and generic framework for key recovery. In *Advances in Cryptology – ASIACRYPT 2022*, pages 410–440. Springer, 2022. doi:[10.1007/978-3-031-22963-3_14](https://doi.org/10.1007/978-3-031-22963-3_14).
- [Wag99] David Wagner. The boomerang attack. In *Fast Software Encryption – FSE 1999*, pages 156–170. Springer, 1999. doi:[10.1007/3-540-48519-8_12](https://doi.org/10.1007/3-540-48519-8_12).
- [WP19] Haoyang Wang and Thomas Peyrin. Boomerang switch in multiple rounds: Application to AES variants and Deoxys. *IACR Transactions on Symmetric Cryptology*, 2019(1):142–169, 2019. doi:[10.46586/tosc.v2019.i1.142-169](https://doi.org/10.46586/tosc.v2019.i1.142-169).
- [WZZ⁺26] Yunong Wu, Yanyan Zhou, Zongsheng Zhang, Tairong Shi, Bin Hu, Kai Zhang, and Senpeng Wang. Improved search models of boomerang distinguishers and application to LILLIPUT. *Cybersecurity*, 9(1):100, 2026. doi:[10.1186/s42400-025-00540-9](https://doi.org/10.1186/s42400-025-00540-9).
- [XCHW25] Zeyu Xu, Jiamin Cui, Kai Hu, and Meiqin Wang. Integral attack on the full FUTURE block cipher. *Tsinghua Science and Technology*, 30(1):161–170, 2025. doi:[10.26599/TST.2024.9010007](https://doi.org/10.26599/TST.2024.9010007).
- [YSZ⁺24] Qianqian Yang, Ling Song, Nana Zhang, Danping Shi, Libo Wang, Jiahao Zhao, Lei Hu, and Jian Weng. Optimizing rectangle and boomerang attacks: A unified and generic framework for key recovery. *Journal of Cryptology*, 37(2):19, 2024. doi:[10.1007/s00145-024-09499-1](https://doi.org/10.1007/s00145-024-09499-1).
- [ZTX25] Xinhao Zeng, Lin Tan, and Hong Xu. Improved search of boomerang distinguishers for generalized feistel and application to WARP. *IACR Transactions on Symmetric Cryptology*, 2025(1):444–470, 2025. doi:[10.46586/tosc.v2025.i1.444-470](https://doi.org/10.46586/tosc.v2025.i1.444-470).