Revisiting the Boomerang Attack from a Perspective of 3-differential

Libo Wang[1], Ling Song[2], Baofeng Wu[3,4], Mostafizar Rahman[1], Takanori Isobe[1,5]

[1] University of Hyogo, Kobe, Japan
[2] Jinan University, Guangzhou, China
[3] Institute of Information Engineering, CAS, Beijing, China
[4] School of Cyber Security, UCAS, Beijing, China
[5] NICT, Tokyo, Japan
wanglibo12b@gmail.com,songling.qs@gmail.com,wubaofeng@iie.ac.cn,
mrahman454@gmail.com,takanori.isobe@ai.u-hyogo.ac.jp

Abstract. In this paper, inspired by the work of Beyne and Rijmen at CRYPTO 2022, we explore the accurate probability of d-differential in the fixed-key model. The theoretical foundations of our method are based on a special matrix − quasi-d-differential transition matrix, which is a natural extension of the quasidifferential transition matrix. The role of quasi-d-differential transition matrices in polytopic cryptananlysis is analogous to that of correlation matrices in linear cryptanalysis. Therefore, the fixed-key probability of a d-differential can be exactly expressed as the sum of the correlations of its quasi-d-differential trails.
Then we revisit the boomerang attack from a perspective of 3-differential. Different from previous works, the probability of a boomerang distinguisher can be exactly expressed as the sum of the correlations of its quasi-3-differential trails without any assumptions in our work.
In order to illustrate our theory, we apply it to the lightweight block cipher GIFT. It is interesting to find the probability of every optimal 3-differential characteristic of an existing 2-round boomerang is zero, which can be seen as an evidence that the security of block ciphers adopting half-round key XOR might be overestimated previously to some extent in differential-like attacks.

Keywords: Boomerang attack, d-differential, Hypothesis of stochastic equivalence, Correlation matrix, Quasidifferential transition matrix, GIFT

1 Introduction

Differential attacks and their various variations have emerged as crucial techniques for assessing the security of block ciphers in contemporary times. Differential attack, proposed by Biham and Shamir at CRYPTO 1990 [10], analyzes a cipher based on the probabilistic transition from an input difference to an output difference. However, constructing differentials for iterated ciphers with a large number of rounds can be exceptionally challenging. In the case of ciphers that

can be expressed as compositions of simple operations, the conventional approach involves tracing sequences of intermediate differences or characteristics. The probability of a characteristic is then estimated heuristically by multiplying the probabilities of the intermediate differentials. Lai *et al.* [25] showed that the above procedure yields the correct value of the key-averaged probability for *Markov ciphers*.

One important consideration in a differential attack is that the key remains fixed throughout the process. As a result, the actual probability may differ significantly from the key-averaged probability. To address this challenge, Lai *et al.* [25] introduced an additional assumption called the *hypothesis of stochastic equivalence*. This assumption suggests that the probability associated with each key is close to the average probability.

The use of averages in assessing probabilities is known to obscure potential weak-key attacks, which can significantly undermine the security of certain ciphers. Until recently, there were limited tools available to compute probabilities beyond the average case, but in [8], Beyne and Rijmen introduced a new tool called the quasidifferential transition matrix, which enables a more comprehensive analysis of differential probabilities.

The role of quasidifferential transition matrix in differential cryptanalysis is similar to that of correlation matrix in linear cryptanalysis. Like the correlation of a linear approximation, which is precisely equal to the sum of the correlations of all its linear trails, the fixed-key probability of a differential can be expressed as the sum of the correlations of all its quasidifferential trails. For the first time, the problem of exactly computing the differential probability has been solved in the fixed-key model without any assumptions, more than three decades since the differential cryptanalysis was proposed.

Inspired by the idea of differential cryptanalysis, *i.e.*, exploiting non-random pairs of input and output differences of a cipher, many variations of it were proposed, including boomerang attack [32], polytopic attack [31] and so on. One of the typical representatives is the boomerang attack, which combines two short differential trails to get a long one with a high probability. It stands out for its ability to potentially penetrate a larger number of rounds in block ciphers, making it a compelling choice for evaluating security.

Despite the potential of boomerang attacks, it has been noted that certain proposed instances of these attacks are incorrect. To mitigate this concern, several tools have been developed to provide a more accurate estimation of the probability associated with a boomerang attack. These tools serve to improve the reliability and precision of evaluating the effectiveness of boomerang distinguishers. In EUROCRYPT 2018, Cid *et al.* introduced a novel tool called Boomerang Connectivity Table (BCT) for estimating the theoretical probability of a middle round [15]. This development has sparked further research in the field, leading to the proposal of various types of tables. These include the Upper `BCT` (`UBCT`) [17,33], the Lower `BCT` (`LBCT`) [17, 29], the Extended `BCT` (`EBCT`) [13, 17], and the Double `BCT` (`DBCT`) [21, 35]. These different tables enhance the analysis and evaluation of boomerang distinguishers. Under specific assumptions, the aforementioned tables

enable the estimation of the probability of boomerang distinguishers, even when the middle part consists of multiple rounds. To reduce computational complexities, all of these tables share a requirement: the two upper differentials and the two lower differentials must be the same simultaneously. This condition allows for more efficient calculations and analysis within the context of boomerang attacks. Recently, when Li *et al.* [26] mounted the rectangle attack to the block cipher GIFT [4] by introducing a new tool named Generalized BCT (GBCT), which was a generalization of BCT and did not require the two upper differentials or the two lower differentials to be equal. This improves the probability of the rectangle distinguisher.

Although many tools were introduced for getting more accurate probabilities of boomerang distinguishers, the resulting probabilities are still key-averaged ones, as these tools implicitly rely on the assumption of stochastic equivalence. Then a question rises naturally *whether it is possible to calculate the exact probability of a boomerang distinguisher without any assumptions.* This also remains an open question in the field of symmetric cryptography as stated in [31] in a different way[1]. Also in [18], Dunkelman *et al.* proposed the open question: *Create a "grand unified theory" of boomerang-like attacks which will explore their hidden relationships and treat them rigorously.*

Our contributions: In this paper, we first devote to partially answering the above questions theoretically, and then explain our theory and show how to use it in practice through analysing the probability of one boomerang distinguisher of GIFT. The main contributions of this paper are summarized below.

1. We generalize the framework of differential cryptanalysis proposed in [8] to polytopic cryptanalysis. Specifically, we generalize the notion of quasidifferential transition matrix (QDTM) to d-differential, and get a matrix named quasi-d-differential transition matrix (d-QDTM), which is obtained by performing a change-of-basis on the transition matrix describing the propagation of probability distributions of $(d+1)$-tuples through $d+1$ functions (may be equal), analogous to the construction of correlation matrix using Fourier transformation. The role of d-QDTM in polytopic cryptanalysis is similar to that of correlation matrix in linear cryptanalysis. For example, composition of functions corresponding to multiplication of d-QDTMs. Therefore, as in linear/differential cryptanalysis, we can prove that the sum of the correlations of all quasi-d-differential trails in a d-differential characteristic is equal to its exact probability.
2. Then, we revisit the boomerang attack from a perspective of 3-differential. Using the above theory, we can give the exact expression of the probability of

[1] In [31], it is stated as follows. *Another open problem is the exact determination of the success probability of boomerang attacks and their extensions. It has correctly been observed that the correlation between differentials must be taken into account to accurately determine the success probability [27]. The true probability can otherwise deviate arbitrarily from the estimated one.*

boomerang distinguisher without any assumptions for the first time. Moreover, under the assumption that intermediate differentials are independent as usually done in classical differential attack, we dive into the sandwich attack and find that there is a gap between the real probability and the value calculated in the sandwich attack framework. Actually, the probability of boomerang distinguisher estimated in previous work under sandwich attack framework can be seen as a kind of average.

3. In order to illustrate the theory we build, we apply it to lightweight block cipher GIFT. For one boomerang distinguisher provided in previous work, we find that the probability of every optimal 3-differential characteristic (in total 2^{13} optimal 3-differential characteristics) is zero, independently with the round keys, *i.e.*, all of them are impossible. The above interesting result can be seen as an evidence that the security of block ciphers adopting half round key XOR against the differential-like attacks is overestimated in previous work. However, in the case of full round key XOR, for each characteristic, there at least exists one round key such that its probability is nonzero under the assumption that the round sub-keys are independent (Note that the assumption does not hold for the round keys of GIFT).

Organization: In Section 2, we recall the linear and differential cryptanalysis, and some notions in polytopic cryptanalysis. In Section 3, we introduce the d-QDTM, an important tool in our work. In Section 4, we mainly focus on 3-differential and prove that the sum of the correlations of all quasi-3-differential trails in a 3-differential characteristic is equal to its exact probability. We revisit the boomerang attack in Section 5 under the framework of 3-differential. Section 6 presents an automated search tool for quasi-3-differential trails in GIFT. Finally, we conclude this paper in Section 7.

2 Preliminaries and Related Work

In this section, we will quickly review the linear attack and differential attack, and the well known tools correlation matrix (CM), differential distribution table (DDT), and quasidifferential transition matrix (QDTM). Then, we recall some notions in polytopic cryptanalysis. The notations used in this paper are consistent with that in [6, 7, 8] as much as possible, and will be introduced where necessary. Throughout this paper, we abuse the notation $\boldsymbol{x}$, which represents a random variable over $\mathbb{F}_2^n$ and also represents a vector with components belonging to $\mathbb{F}_2^n$, *i.e.*, $\boldsymbol{x} = (x_0, x_1, \ldots, x_d)$, $x_i \in \mathbb{F}_2^n, 0 \leq i \leq d$.

2.1 Linear Cryptanalysis

Although the present paper only focuses on differential-like cryptanalysis, it is useful to introduce the background of linear cryptanalysis, especially the novel tool – correlation matrices [16], which provide an important motivation for the quasidifferentil framework of Beyne and Rijmen's work [8].

Let $\mathbb{R}[\mathbb{F}_2^n]$ be the set of all functions from $\mathbb{F}_2^n$ to the real number field $\mathbb{R}$. Obviously, $\mathbb{R}[\mathbb{F}_2^n]$ is a vector space over $\mathbb{R}$ with dimension 2^n. It is well known that $\{\delta_a \mid a \in \mathbb{F}_2^n\}$ is a basis of $\mathbb{R}[\mathbb{F}_2^n]$, where the function δ_a is defined as $\delta_a(x) = 1$ if $x = a$ and zero elsewhere. This basis is an orthonormal basis of $\mathbb{R}[\mathbb{F}_2^n]$ with respect to the inner product $\langle f, g\rangle = \sum_{x\in\mathbb{F}_2^n} f(x)g(x)$. We call it *standard basis* in the following.

There exists another well known basis for $\mathbb{R}[\mathbb{F}_2^n]$ consisting of the additive characters of $\mathbb{F}_2^n$. These are homomorphisms from $\mathbb{R}[\mathbb{F}_2^n]$ to the multiplicative group $\mathbb{C}\setminus\{0\}$. Because of the special structure of $\mathbb{F}_2^n$, every such homomorphism is of the form $\chi_u(x) = (-1)^{u^\top x}$ with $u, x \in \mathbb{F}_2^n$ being column vectors. All characters χ_u form an orthogonal basis for $\mathbb{R}[\mathbb{F}_2^n]$ with respect to the above defined inner product, *i.e.*, $\langle \chi_u, \chi_v\rangle = \sum_{x\in\mathbb{F}_2^n} \chi_u(x)\chi_v(x) = 2^n\delta_u(v)$. The basis $\{\chi_u \mid u \in \mathbb{F}_2^n\}$ will be called the *character basis* throughout this paper. It is well known that from linear algebra every function $f \in \mathbb{R}[\mathbb{F}_2^n]$ can be uniquely expressed as a linear combination of a basis for $\mathbb{R}[\mathbb{F}_2^n]$ with coefficients in $\mathbb{R}$. Therefore, f can be written as

$$f = \sum_{u\in\mathbb{F}_2^n} f(u)\delta_u = \sum_{u\in\mathbb{F}_2^n} \frac{\langle f, \chi_u\rangle}{2^n}\chi_u. \tag{1}$$

Now we give the define of Fourier transformation.

Definition 1 (Fourier transformation). *Let $f \in \mathbb{R}[\mathbb{F}_2^n]$. The Fourier transformation of f, denoted by $\mathcal{F}_n f$, is also a function in $\mathbb{R}[\mathbb{F}_2^n]$, and defined by $(\mathcal{F}_n f)(u) = \langle \chi_u, f\rangle$ for each $u \in \mathbb{F}_2^n$. That is, $\mathcal{F}_n f = \sum_{u\in\mathbb{F}_2^n}\langle \chi_u, f\rangle\delta_u$.*

Remark 1. we remind that in some literatures, the Fourier transformation of f, *i.e.*, $\mathcal{F}_n f$, belongs to $\mathbb{R}[\widehat{\mathbb{F}}_2^n]$, where $\widehat{\mathbb{F}}_2^n$ is the group formed by all the characters of $\mathbb{F}_2^n$, and $\mathcal{F}_n f$ is defined by $(\mathcal{F}_n f)(\chi_u) = \langle \chi_u, f\rangle$. Because there is an isomorphism between $\widehat{\mathbb{F}}_2^n$ and $\mathbb{F}_2^n$, for convenience, throughout this paper we assume $\mathcal{F}_n f \in \mathbb{R}[\mathbb{F}_2^n]$.

From Definition 1 and Formula (1), we find that when f and its Fourier transformation $\mathcal{F}_n f$ are respectively expressed by the character basis and the standard basis, they have the same coefficients (up to a constant factor $\frac{1}{2^n}$). It is easy to see that Fourier transformation is a linear operator and changes the character basis $\{\chi_u \mid u \in \mathbb{F}_2^n\}$ to standard basis $\{\delta_u \mid u \in \mathbb{F}_2^n\}$, that is, $\mathcal{F}_n\chi_u = 2^n\delta_u$ for any $u \in \mathbb{F}_2^n$.

Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. We define the linear operator $T^{\mathsf{F}} : \mathbb{R}[\mathbb{F}_2^n] \to \mathbb{R}[\mathbb{F}_2^m]$ by $T^{\mathsf{F}}(\delta_x) = \delta_{\mathsf{F}(x)}$ for all $x \in \mathbb{F}_2^n$. The transition matrix of F is the coordinate representation of T^{F} with respect to the standard bases of $\mathbb{R}[\mathbb{F}_2^n]$ and $\mathbb{R}[\mathbb{F}_2^m]$. It is interesting that Beyne found the correlation matrix, proposed by Daemen *et al.* [16], can be seen as the transition matrix of T^{F} with respect to the character bases of $\mathbb{R}[\mathbb{F}_2^n]$ and $\mathbb{R}[\mathbb{F}_2^m]$ [7]. Therefore, by the change-of-basis operation of Fourier transformation, the correlation matrix can be defined as follows.

Definition 2 (Correlation matrix). *Let* $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. *Define* $C^{\mathsf{F}} : \mathbb{R}[\mathbb{F}_2^n] \to \mathbb{R}[\mathbb{F}_2^m]$ *as the Fourier transformation of* T^{F}. *That is,* $C^{\mathsf{F}} = \mathcal{F}_m T^{\mathsf{F}} \mathcal{F}_n^{-1}$, *as illustrated in Fig. 1.*

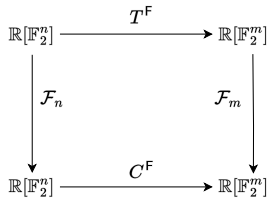

Fig. 1: The relationship between correlation matrix and transition matrix

The `CM` by Definition 2 is consistent with the original definition due to Daemen *et al.* [16], in which the (u, v)-entry is defined by $C^{\mathsf{F}}_{u,v} = 2\Pr[v^\top \mathsf{F}(\boldsymbol{x}) + u^\top \boldsymbol{x} = 0] - 1$ with $\boldsymbol{x}$ uniform random on $\mathbb{F}_2^n$. `CM`s satisfy several properties, one of which is that, for a function $\mathsf{F} = \mathsf{F}_r \circ \mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1$, it holds that

$$C^{\mathsf{F}} = C^{\mathsf{F}_r} C^{\mathsf{F}_{r-1}} \cdots C^{\mathsf{F}_1}.$$

Expanding the above equation in coordinates yields the following identity:

$$C^{\mathsf{F}}_{u_{r+1},u_1} = \sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} C^{\mathsf{F}_i}_{u_{i+1},u_i}. \tag{2}$$

That is, the correlation of a linear approximation is equal to the sum of the correlations of all linear trails defined by the intermediate masks $u_2, \ldots, u_r$.

2.2 Differential Cryptanalysis

The central problem of differential cryptanalysis is to find a high probability differential characteristic for a target cipher. To achieve this goal cryptanalyst usually analyzes the propagation of differences through components, and each component can be seen as a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Thus, the cryptanalyst attempts to find a differential $(a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$ such that the number of solutions of the difference equation

$$\mathsf{F}(x \oplus a) \oplus \mathsf{F}(x) = b$$

is as large as possible. The difference distribution table $\mathtt{DDT}^{\mathsf{F}}$ is a $2^n \times 2^m$ table with rows and columns indexed by input and output differences, respectively. The corresponding entries are equal to the number of solutions for a particular differential:

$$\mathtt{DDT}^{\mathsf{F}}_{a,b} = |\{x \in \mathbb{F}_2^n \mid \mathsf{F}(x \oplus a) \oplus \mathsf{F}(x) = b\}| = 2^n \Pr[\mathsf{F}(\boldsymbol{x} \oplus a) \oplus \mathsf{F}(\boldsymbol{x}) = b]$$

with $\boldsymbol{x}$ uniform random on $\mathbb{F}_2^n$.

Directly computing or estimating the probability of a differential for a cipher of large input size (such as 128 bits) is computationally difficult. However, many ciphers are of the form $\mathsf{F} = \mathsf{F}_r \circ \mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1$, where the functions F_i admit differentials with relatively high probabilities and are usually easier to analyze. In this case, the probability of a differential (a_1, a_{r+1}) can be estimated based on *characteristics*. A characteristic is a sequence $(a_1, a_2, \ldots, a_{r+1})$ of compatible intermediate input and output differences for each of the functions F_i. For the sake of simplicity, we assume that the functions F_i are all n-bit functions, *i.e.*, $m = n$. It holds that

$$\Pr[\mathsf{F}(\boldsymbol{x} \oplus a_1) \oplus \mathsf{F}(\boldsymbol{x}) = a_{r+1}] = \sum_{a_2,\ldots,a_r} \Pr[\wedge_{i=1}^{r} \mathsf{F}_i(\boldsymbol{x}_i \oplus a_i) \oplus \mathsf{F}_i(\boldsymbol{x}_i) = a_{i+1}]$$

with $\boldsymbol{x}_1$ uniform random on $\mathbb{F}_2^n$ and $\boldsymbol{x}_i = \mathsf{F}_{i-1}(\boldsymbol{x}_{i-1})$ for $2 \le i \le r$. The probability of a characteristic is often estimated using the assumption that intermediate differentials are independent:

$$\Pr[\wedge_{i=1}^{r} \mathsf{F}_i(\boldsymbol{x}_i \oplus a_i) \oplus \mathsf{F}_i(\boldsymbol{x}_i) = a_{i+1}] \approx \prod_{i=1}^{r} \Pr[\mathsf{F}_i(\boldsymbol{x}_i \oplus a_i) \oplus \mathsf{F}_i(\boldsymbol{x}_i) = a_{i+1}].$$

Combining the above three equations, we get

$$\mathtt{DDT}^{\mathsf{F}}_{a_1,a_{r+1}}/2^n \approx \sum_{a_2,\ldots,a_r} \prod_{i=1}^{r} \mathtt{DDT}^{\mathsf{F}_i}_{a_i,a_{i+1}}/2^n. \tag{3}$$

Equation (3) for differential probability should be compared with Equation (2) for liner approximation. However, there is a fundamental difference: whereas Equation (3) is heuristic and at best true on average with respect to independent uniform random round keys, Equation (2) holds exactly without any assumptions. Closing the gap between Equation (3) and Equation (2) is essential to achieve a more complete understanding of differential cryptanalysis. Recently, Beyne and Rijmen solved the problem through introducing the following new tool.

Definition 3 (Quasidifferential transition matrix). *Let n and m be two positive integers, and $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. The quasidifferential transition matrix D^{F} is a matrix with size $2^{2m} \times 2^{2n}$, and its $((v,b),(u,a))$-entry is defined by*

$$D^{\mathsf{F}}_{(v,b),(u,a)} = \frac{1}{2^n} \sum_{\substack{x \in \mathbb{F}_2^n \\ \mathsf{F}(x) \oplus \mathsf{F}(x \oplus a) = b}} (-1)^{v^\top \mathsf{F}(x) + u^\top x},$$

where $v, b \in \mathbb{F}_2^m$, $u, a \in \mathbb{F}_2^n$.

The role of `QDTM` in differential cryptanalysis is similar as that of `CM` in linear cryptanalysis, so they have the similar properties. For example, for a function $\mathsf{F} = \mathsf{F}_r \circ \mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1$, it also holds that

$$D^{\mathsf{F}} = D^{\mathsf{F}_r} D^{\mathsf{F}_{r-1}} \cdots D^{\mathsf{F}_1}.$$

We denote $\varpi_i = (u_i, a_i)$ for $1 \leq i \leq r+1$. Expanding the above equation in coordinates yields the following identity:

$$D^{\mathsf{F}}_{\varpi_{r+1},\varpi_1} = \sum_{\varpi_2,\ldots,\varpi_r} \prod_{i=1}^{r} D^{\mathsf{F}_i}_{\varpi_{i+1},\varpi_i}. \tag{4}$$

When we restrict $u_{r+1} = u_1 = 0$, from Equation (4), we can get that the exact probability of differential (a_1, a_{r+1}) is equal to the sum of the correlations of all quasidifferential trails with input and output mask-difference pairs $\varpi_1 = (0, a_1)$ and $\varpi_{r+1} = (0, a_{r+1})$, respectively. Actually, Equation (4) is more like Equation (2), which puts the theory of differential and linear cryptanalysis on an equal footing.

2.3 *d*-differences and Their Transitions

In this subsection, we recall the definitions of d-differences and their transitions. The readers can refer to [31] for more details.

Different from original differential cryptanalysis that analyses the propagation of difference between only two plaintexts, differential-like cryptanalysis usually have to track the differences among multiple plaintexts. We are usually not interested in the absolute position of texts in the state space but only in their relative positions, *i.e.*, their differences. The relative positions of a tuple of $d+1$ texts can be defined by the differences of the last d texts with respect to the first one, which leads to the following definition of d-difference.

Definition 4 (d-difference). *For a $(d+1)$-tuple $(m_0, m_1, \ldots, m_d)$, its d-difference is defined as the following d-tuple*

$$(m_1 \oplus m_0, m_2 \oplus m_0, \ldots, m_d \oplus m_0).$$

We refer to the first text m_0 of the $(d+1)$-tuple of messages as the anchor of the d-difference.

It is easy to see that a $(d+1)$-tuple is uniquely determined by its d-difference together with its anchor. Similar to the propagation of difference, we can define the transition of d-differences.

Definition 5 (d-difference transition). *Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^n$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_d)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)$ be two d-differences over $\mathbb{F}_2^n$. We use the notation $\boldsymbol{\alpha} \xrightarrow[x]{\mathsf{F}} \boldsymbol{\beta}$ to denote the event that F maps the $(d+1)$-tuple of messages corresponding to the d-difference $\boldsymbol{\alpha}$ with anchor x to a $(d+1)$-tuple of messages corresponding to d-difference $\boldsymbol{\beta}$ with anchor $\mathsf{F}(x)$. More precisely, the notation $\boldsymbol{\alpha} \xrightarrow[x]{\mathsf{F}} \boldsymbol{\beta}$ implies that*

$$\begin{aligned} \mathsf{F}(x \oplus \alpha_1) \oplus \mathsf{F}(x) &= \beta_1, \\ \mathsf{F}(x \oplus \alpha_2) \oplus \mathsf{F}(x) &= \beta_2, \\ &\cdots \\ \mathsf{F}(x \oplus \alpha_d) \oplus \mathsf{F}(x) &= \beta_d. \end{aligned}$$

Example 1. Let (m_0, m_1, m_2, m_3) be a tuple of four plaintexts, and the corresponding 3-difference be $\boldsymbol{\alpha} = (m_1 \oplus m_0, m_2 \oplus m_0, m_3 \oplus m_0)$. For some 3-difference $\boldsymbol{\beta}$, if $\boldsymbol{\alpha} \xrightarrow[m_0]{\mathsf{F}} \boldsymbol{\beta}$, then $\boldsymbol{\beta} = (\mathsf{F}(m_1) \oplus \mathsf{F}(m_0), \mathsf{F}(m_2) \oplus \mathsf{F}(m_0), \mathsf{F}(m_3) \oplus \mathsf{F}(m_0))$.

Note that when $d = 1$, the d-difference is reduced to the classical difference. Like in standard differential cryptanalysis, we are also interested in the probability of the d-difference transition in differential-like cryptanalysis.

Definition 6 (Probability of transition). *Let F, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ be same as in Definition 5. The probability of the transition $\boldsymbol{\alpha} \xrightarrow{\mathsf{F}} \boldsymbol{\beta}$ is defined as*

$$\Pr(\boldsymbol{\alpha} \xrightarrow{\mathsf{F}} \boldsymbol{\beta}) = \Pr_{\boldsymbol{x}}(\boldsymbol{\alpha} \xrightarrow[\boldsymbol{x}]{\mathsf{F}} \boldsymbol{\beta}) = \frac{\#\{x \in \mathbb{F}_2^n \mid \mathsf{F}(x \oplus \alpha_i) \oplus \mathsf{F}(x) = \beta_i, 1 \le i \le d\}}{2^n}$$

where $\boldsymbol{x}$ is uniform random on $\mathbb{F}_2^n$.

In order to estimate the probability of d-differential for a cipher, we first need to introduce the notion of d-differential characteristic. In fact, it is the generalization of differential characteristic.

Definition 7 (A d-differential characteristic). *Let a cipher $\mathsf{F} = \mathsf{F}_r \circ \mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1$ be a composition of r n-bits functions, and $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_{r+1}} \in \mathbb{F}_2^{dn}$ be a sequence of d-differences. We refer to the sequence $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_{r+1}}$ as a characteristic over F if it satisfies $\boldsymbol{\alpha_1} \xrightarrow[x]{\mathsf{F}_1} \boldsymbol{\alpha_2} \xrightarrow[\mathsf{F}_1(x)]{\mathsf{F}_2} \boldsymbol{\alpha_3} \to \cdots \to \boldsymbol{\alpha_r} \xrightarrow[\mathsf{F}_{r-1} \circ \cdots \circ \mathsf{F}_1(x)]{\mathsf{F}_r} \boldsymbol{\alpha_{r+1}}$. That is, the $(d+1)$-tuple corresponding to $\boldsymbol{\alpha}_1$ with the anchor x follows the trail $(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_{r+1}})$.*

Then we can estimate the probability of a d-differential characteristic as we usually do for differential characteristic, namely we make the same assumption as in differential cryptanalysis. Therefore, we can approximate the probability of a d-differential characteristic by considering the individual transitions as independent, and we can further estimate the probability of a d-differential. Similar as Equation (3), we have the following formula

$$\Pr[\wedge_{i=1}^{r}(\boldsymbol{\alpha_i} \xrightarrow[\boldsymbol{x}_i]{\mathsf{F}_i} \boldsymbol{\alpha_{i+1}})] \approx \prod_{i=1}^{r} \Pr[(\boldsymbol{\alpha_i} \xrightarrow[\boldsymbol{x}_i]{\mathsf{F}_i} \boldsymbol{\alpha_{i+1}})].$$

Furthermore, the probability of a d-differential $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_{r+1})$ can be estimated by

$$\Pr[(\boldsymbol{\alpha_1} \xrightarrow[\boldsymbol{x}]{\mathsf{F}} \boldsymbol{\alpha_{r+1}})] \approx \sum_{\boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r} \prod_{i=1}^{r} \Pr[(\boldsymbol{\alpha_i} \xrightarrow[\boldsymbol{x}_i]{\mathsf{F}_i} \boldsymbol{\alpha_{i+1}})]. \tag{5}$$

We emphasize that Equation (5) is heuristic and at best true on average with respect to independent uniform random round keys. In Section 3, we will generalize the notion of `QDTM` to d-differential, and further give a formula for the exact probability of d-differential in theoretical sound way.

Before going to next section, we give the definition of truncated d-difference, which is useful when we describe the probability of boomerang distinguisher.

Definition 8 (Transitions of truncated d-differences). *A truncated d-difference is an affine subspace in the linear space of d-differences. Let A, B be two truncated d-differences, the probability of d-difference transition $A \xrightarrow{\mathsf{F}} B$ is defined as the probability that an input d-difference, chosen uniformly at random from A, maps to a d-difference in B:*

$$\Pr[A \xrightarrow{\mathsf{F}} B] = |A|^{-1} \sum_{\alpha \in A, \beta \in B} \Pr[\alpha \xrightarrow{\mathsf{F}} \beta]. \tag{6}$$

Note that when F is bijective on $\mathbb{F}_2^n$, usually $\Pr[B \xrightarrow{\mathsf{F}^{-1}} A] \neq \Pr[A \xrightarrow{\mathsf{F}} B]$, but they have the following relationship:

$$|A| \Pr[A \xrightarrow{\mathsf{F}} B] = |B| \Pr[B \xrightarrow{\mathsf{F}^{-1}} A].$$

3 Quasi-d-differential Transition Matrices

In this section we prove that the notion of `QDTM` can be generalized to d-difference and get the d-`QDTM`.

3.1 Quasi-d-differential Basis

Let $\mathbb{P} = \underbrace{\mathbb{F}_2^n \times \cdots \times \mathbb{F}_2^n}_{d+1 \text{ times}}$ and $\mathbb{R}[\mathbb{P}]$ be the set of all functions from $\mathbb{P}$ to the real number field $\mathbb{R}$. Obviously, $\mathbb{R}[\mathbb{P}]$ is a vector space over $\mathbb{R}$ with dimension $2^{(d+1)n}$.

Definition 9 (Quasi-d-differential basis). *For any $(u, \boldsymbol{\alpha}) = (u, \alpha_1, \ldots, \alpha_d) \in \mathbb{P}$, the function $\mathcal{B}_{u,\boldsymbol{\alpha}} : \mathbb{P} \to \mathbb{R}$ is defined by*

$$\mathcal{B}_{u,\boldsymbol{\alpha}}(\boldsymbol{x}) = \chi_u(x_0) \prod_{i=1}^{d} \delta_{\alpha_i}(x_0 \oplus x_i)$$

for every $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{P}$. Then all elements in the set $\{\mathcal{B}_{u,\boldsymbol{\alpha}} \mid (u, \boldsymbol{\alpha}) \in \mathbb{P}\}$ will be called the quasi-d-differential basis for $\mathbb{R}[\mathbb{P}]$ in this paper.

The basis in Defintion 9 is orthogonal, which is shown in the following theorem.

Theorem 1. *The quasi-d-differential basis defined in Definition 9 is orthogonal.*

Proof. For arbitrary $(u, \boldsymbol{\alpha}), (v, \boldsymbol{\beta}) \in \mathbb{P}$, we have

$$\begin{aligned}
\langle \mathcal{B}_{u,\boldsymbol{\alpha}}, \mathcal{B}_{v,\boldsymbol{\beta}} \rangle &= \sum_{\boldsymbol{x} \in \mathbb{P}} \chi_u(x_0)\chi_v(x_0) \prod_{i=1}^{d} \delta_{\alpha_i}(x_0 \oplus x_i)\delta_{\beta_i}(x_0 \oplus x_i) \\
&= \prod_{i=1}^{d} \delta_{\alpha_i}(\beta_i) \sum_{x_0 \in \mathbb{F}_2^n} \chi_u(x_0)\chi_v(x_0)
\end{aligned}$$

$$= 2^n \delta_u(v) \prod_{i=1}^{d} \delta_{\alpha_i}(\beta_i),$$

which means that the quasi-d-differential basis is orthogonal.

Similar to the Fourier transformation, we can define the *change-of-basis* operator $\mathfrak{D}_n : \mathbb{R}[\mathbb{P}] \to \mathbb{R}[\mathbb{P}]$ by $(\mathfrak{D}_n f)(u, \boldsymbol{\alpha}) = \langle \mathcal{B}_{u,\boldsymbol{\alpha}}, f \rangle$, where $f \in \mathbb{R}[\mathbb{P}]$. It is easy to see that $\mathfrak{D}_n$ is a linear operator, and we also use it to represent the corresponding matrix in the following. In fact, $(\mathfrak{D}_n f)(u, \boldsymbol{\alpha})/2^n$ is the coordinate corresponding to basis function $\mathcal{B}_{u,\boldsymbol{\alpha}}$ when f is expressed in the quasi-d-differential basis, *i.e.*,

$$f = \sum_{(u,\boldsymbol{\alpha}) \in \mathbb{P}} \frac{(\mathfrak{D}_n f)(u, \boldsymbol{\alpha})}{2^n} \mathcal{B}_{u,\boldsymbol{\alpha}}.$$

3.2 Quasi-d-differential Transition Matrices

Let n and m be two positive integers and $\mathsf{F}_i : \mathbb{F}_2^n \to \mathbb{F}_2^m$, for $1 \le i \le d+1$. The transition matrix of $(d+1)$-tuples through $d+1$ functions can be written as the Kronecker product $\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i}$, which is defined as a $2^{(d+1)m} \times 2^{(d+1)n}$ matrix with coordinates

$$\left(\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i} \right)_{(y_1,y_2,\ldots,y_{d+1}),(x_1,x_2,\ldots,x_{d+1})} = \prod_{i=1}^{d+1} T^{\mathsf{F_i}}_{y_i,x_i} = \prod_{i=1}^{d+1} \delta_{y_i}(\mathsf{F}_i(x_i)).$$

For convenience, we index the coordinates of $\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i}$ directly by pairs of bitvectors.

Definition 10 (Quasi-d-differential transition matrix). *The quasi-d-differential transition matrix $D^{\mathsf{F}_1,\mathsf{F}_2,\ldots,\mathsf{F}_{d+1}}$ is defined as the matrix representation of the linear transformation induced by $\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i}$ with respect to the quasi-d-differential basis, as illustrated in Fig. 2. That is, $D^{\mathsf{F}_1,\mathsf{F}_2,\ldots,\mathsf{F}_{d+1}} = \mathfrak{D}_m \left(\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i} \right) \mathfrak{D}_n^{-1}$.*

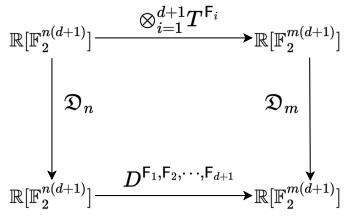

Fig. 2: The quasi-d-differential transition matrix

When $\mathsf{F}_1 = \mathsf{F}_2 = \cdots = \mathsf{F}_{d+1} = \mathsf{F}$, for convenience, we denote $D^{\mathsf{F}_1,\mathsf{F}_2,\ldots,\mathsf{F}_{d+1}}$ by D^{F} in the following. In particular, when $d = 1$ and $\mathsf{F}_1 = \mathsf{F}_2 = \mathsf{F}$, the above d-QDTM becomes $D^{\mathsf{F}} = \mathfrak{D}_m (T^{\mathsf{F}} \otimes T^{\mathsf{F}}) \mathfrak{D}_n^{-1}$, which is the QDTM proposed in [8].

Note that for arbitrary $f, g \in \mathbb{R}[\mathbb{P}]$, we have

$$\begin{aligned}
\langle \mathfrak{D}_n f, \mathfrak{D}_n g \rangle &= \sum_{(u,\boldsymbol{\alpha})\in\mathbb{P}} (\mathfrak{D}_n f)(u,\boldsymbol{\alpha}) \cdot (\mathfrak{D}_n g)(u,\boldsymbol{\alpha}) \\
&= \sum_{(u,\boldsymbol{\alpha})\in\mathbb{P}} \langle \mathcal{B}_{u,\boldsymbol{\alpha}}, f\rangle \langle \mathcal{B}_{u,\boldsymbol{\alpha}}, g\rangle \\
&= \sum_{(u,\boldsymbol{\alpha})\in\mathbb{P}} \Big(\sum_{\boldsymbol{x}\in\mathbb{P}} f(\boldsymbol{x})\chi_u(x_0) \prod_{i=1}^{d} \delta_{\alpha_i}(x_0 \oplus x_i) \Big) \Big(\sum_{\boldsymbol{y}\in\mathbb{P}} g(\boldsymbol{y})\chi_u(y_0) \prod_{i=1}^{d} \delta_{\alpha_i}(y_0 \oplus y_i) \Big) \\
&= 2^n \sum_{\boldsymbol{x}\in\mathbb{P}} f(\boldsymbol{x}) g(\boldsymbol{x}) \\
&= 2^n \langle f, g\rangle,
\end{aligned}$$

which implies that $\mathfrak{D}_n^{-1} = \mathfrak{D}_n^{\top}/2^n$.

To make Definition 10 more concrete, we compute the entries of $D^{\mathsf{F}_1,\mathsf{F}_2,\ldots,\mathsf{F}_{d+1}}$ as follows, where $(u,\boldsymbol{\alpha}) = (u,\alpha_1,\ldots,\alpha_d) \in \mathbb{F}_2^{n(d+1)}$ and $(v,\boldsymbol{\beta}) = (v,\beta_1,\ldots,\beta_d) \in \mathbb{F}_2^{m(d+1)}$, which are clear in Fig. 2.

$$\begin{aligned}
D^{\mathsf{F}_1,\mathsf{F}_2,\ldots,\mathsf{F}_{d+1}}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})} &= \left\langle \delta_{(v,\boldsymbol{\beta})}, \mathfrak{D}_m \left(\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i} \right) \mathfrak{D}_n^{-1} \delta_{(u,\boldsymbol{\alpha})} \right\rangle \\
&= \frac{1}{2^n} \left\langle \delta_{(v,\boldsymbol{\beta})}, \mathfrak{D}_m \left(\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i} \right) \mathfrak{D}_n^{\top} \delta_{(u,\boldsymbol{\alpha})} \right\rangle \\
&= \frac{1}{2^n} \left\langle \mathcal{B}_{(v,\boldsymbol{\beta})}, \left(\bigotimes_{i=1}^{d+1} T^{\mathsf{F}_i} \right) \mathcal{B}_{(u,\boldsymbol{\alpha})} \right\rangle \\
&= \frac{1}{2^n} \sum_{\boldsymbol{x}\in\mathbb{P}} \chi_u(x_0)\chi_v(\mathsf{F}_1(x_0)) \prod_{i=1}^{d} \delta_{\alpha_i}(x_0 \oplus x_i) \delta_{\beta_i}(\mathsf{F}_1(x_0) \oplus \mathsf{F}_{i+1}(x_i)) \\
&= \frac{1}{2^n} \sum_{\substack{\boldsymbol{x}\in\mathbb{P} \\ x_0\oplus x_i = \alpha_i \\ \mathsf{F}_1(x_0)\oplus\mathsf{F}_{i+1}(x_i)=\beta_i, 1\le i\le d}} (-1)^{u^{\top}x_0 + v^{\top}\mathsf{F}_1(x_0)} \\
&= \frac{1}{2^n} \sum_{\substack{x_0\in\mathbb{F}_2^n \\ \mathsf{F}_1(x_0)\oplus\mathsf{F}_{i+1}(x_0\oplus\alpha_i)=\beta_i, 1\le i\le d}} (-1)^{u^{\top}x_0 + v^{\top}\mathsf{F}_1(x_0)}. \qquad (7)
\end{aligned}$$

In particular, when $\mathsf{F}_1 = \mathsf{F}_2 = \cdots = \mathsf{F}_{d+1} = \mathsf{F}$, we have

$$D^{\mathsf{F}}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})} = \frac{1}{2^n} \sum_{\substack{x\in\mathbb{F}_2^n \\ \boldsymbol{\alpha} \xrightarrow[x]{\mathsf{F}} \boldsymbol{\beta}}} (-1)^{u^{\top}x + v^{\top}\mathsf{F}(x)}. \qquad (8)$$

In this case, for $u = v = 0$, the above Equation (8) reduces to the probability of the d-differential with input d-difference $\boldsymbol{\alpha}$ and output d-difference $\boldsymbol{\beta}$, respectively.

For $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, one obtains the coordinates of the CM of the function F. More generally, the right hand side of Equation (8) can be interpreted as a kind of correlation matrix for the function F when restricted to the values satisfying the d-differential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

The d-QDTM has some basic and important properties, which are useful to analyse the probability of d-differential. In next section we will present some properties for 3-differential, actually all of which can be generalized to d-differential naturally.

4 Quasi-3-differential Trails

In this section, we restrict $d = 3$, because the boomerang attack can be described by using 3-differentials, which will be clear in the next section.

4.1 Basic Properties for Quasi-3-differential Transition Matrices

First of all, we list some useful properties for 3-QDTMs in the following theorem. The Kronecker product of two 3-QDTMs, which will be used in the following theorem, is defined by

$$\left(D^{\mathsf{F}_1^1,\mathsf{F}_2^1,\mathsf{F}_3^1,\mathsf{F}_4^1} \bigotimes D^{\mathsf{F}_1^2,\mathsf{F}_2^2,\mathsf{F}_3^3,\mathsf{F}_4^4}\right)_{(v_1||v_2,\boldsymbol{\alpha}_1||\boldsymbol{\alpha}_2),(u_1||u_2,\boldsymbol{\beta}_1||\boldsymbol{\beta}_2)} = D^{\mathsf{F}_1^1,\mathsf{F}_2^1,\mathsf{F}_3^1,\mathsf{F}_4^1}_{(v_1,\boldsymbol{\alpha}_1),(u_1,\boldsymbol{\beta}_1)} D^{\mathsf{F}_1^2,\mathsf{F}_2^2,\mathsf{F}_3^2,\mathsf{F}_4^2}_{(v_2,\boldsymbol{\alpha}_2),(u_2,\boldsymbol{\beta}_2)},$$

where $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$ are all 3-differences.

Theorem 2. *Let n and m be two positive integers, and $\mathsf{F}_i : \mathbb{F}_2^n \to \mathbb{F}_2^m$, $1 \le i \le 4$. The quasi-3-differential transition matrix $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}$ has the following properties:*
(1) If each F_i is a bijective, then $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}$ is an orthogonal matrix.
(2) If $\mathsf{F}_i = (\mathsf{F}_i^1, \mathsf{F}_i^2, \ldots, \mathsf{F}_i^k)$, where $\mathsf{F}_i^j : \mathbb{F}_2^{n_j} \to \mathbb{F}_2^{m_j}$, $\sum_{j=1}^k n_j = n$ and $\sum_{j=1}^k m_j = m$, then $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4} = \bigotimes_{j=1}^k D^{\mathsf{F}_1^j,\mathsf{F}_2^j,\mathsf{F}_3^j,\mathsf{F}_4^j}$.
(3) If $\mathsf{F}_i = \mathsf{F}_i^2 \circ \mathsf{F}_i^1$, where $\mathsf{F}_i^1 : \mathbb{F}_2^n \to \mathbb{F}_2^\ell$ and $\mathsf{F}_i^2 : \mathbb{F}_2^\ell \to \mathbb{F}_2^m$, then $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4} = D^{\mathsf{F}_1^2,\mathsf{F}_2^2,\mathsf{F}_3^2,\mathsf{F}_4^2} D^{\mathsf{F}_1^1,\mathsf{F}_2^1,\mathsf{F}_3^1,\mathsf{F}_4^1}$.
(4) If $\mathsf{F}_i(x) = Ax \oplus c_i$, where A is an $m \times n$ matrix and $c \in \mathbb{F}_2^m$, and $(u, \boldsymbol{\alpha}) = (u, \alpha_1, \alpha_2, \alpha_3) \in \mathbb{F}_2^{4n}$, $(v, \boldsymbol{\beta}) = (v, \beta_1, \beta_2, \beta_3) \in \mathbb{F}_2^{4m}$, then

$$D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})} = \chi_v(c_1)\delta_u(A^\top v)\delta_{\beta_1}(A\alpha_1 \oplus c_2 \oplus c_1)\delta_{\beta_2}(A\alpha_2 \oplus c_3 \oplus c_1)\delta_{\beta_3}(A\alpha_3 \oplus c_4 \oplus c_1).$$

In particular, when A is an identity matrix and $c_1 = c_2 = c_3 = c_4 = c$, i.e., $\mathsf{F}_i(x) = x \oplus c$, then $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})} = \chi_v(c)\delta_u(v)\delta_{\beta_1}(\alpha_1)\delta_{\beta_2}(\alpha_2)\delta_{\beta_3}(\alpha_3)$. Thus, $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})}$ is nonzero if and only if $(u, \boldsymbol{\alpha}) = (v, \boldsymbol{\beta})$, i.e., $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4}$ is a diagonal matrix in this case.

Proof. Note that $D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4} = \mathfrak{D}_n(T^{\mathsf{F}_1} \otimes T^{\mathsf{F}_2} \otimes T^{\mathsf{F}_3} \otimes T^{\mathsf{F}_4})\mathfrak{D}_n^{-1} = \mathfrak{D}_n(T^{\mathsf{F}_1} \otimes T^{\mathsf{F}_2} \otimes T^{\mathsf{F}_3} \otimes T^{\mathsf{F}_4})\mathfrak{D}_n^\top/2^n$. Property (1) follows from the fact that $T^{\mathsf{F}_1} \otimes T^{\mathsf{F}_2} \otimes T^{\mathsf{F}_3} \otimes T^{\mathsf{F}_4}$ is a permutation matrix when F_i are bijective, $1 \le i \le 4$, and the fact that $\mathfrak{D}_n/\sqrt{2^n}$

is an orthogonal matrix. Property (2) can be easily obtained form Equation (8). Property (3) comes from

$$\begin{aligned} D^{\mathsf{F}_1,\mathsf{F}_2,\mathsf{F}_3,\mathsf{F}_4} &= \mathfrak{D}_m(T^{\mathsf{F}_1} \otimes T^{\mathsf{F}_2} \otimes T^{\mathsf{F}_3} \otimes T^{\mathsf{F}_4})\mathfrak{D}_n^{-1} \\ &= \mathfrak{D}_m(T^{\mathsf{F}_1^2} \otimes T^{\mathsf{F}_2^2} \otimes T^{\mathsf{F}_3^2} \otimes T^{\mathsf{F}_4^2})(T^{\mathsf{F}_1^1} \otimes T^{\mathsf{F}_2^1} \otimes T^{\mathsf{F}_3^1} \otimes T^{\mathsf{F}_4^1})\mathfrak{D}_n^{-1} \\ &= \mathfrak{D}_m(T^{\mathsf{F}_1^2} \otimes T^{\mathsf{F}_2^2} \otimes T^{\mathsf{F}_3^2} \otimes T^{\mathsf{F}_4^2})\mathfrak{D}_\ell^{-1}\mathfrak{D}_\ell(T^{\mathsf{F}_1^1} \otimes T^{\mathsf{F}_2^1} \otimes T^{\mathsf{F}_3^1} \otimes T^{\mathsf{F}_4^1})\mathfrak{D}_n^{-1} \\ &= D^{\mathsf{F}_1^2,\mathsf{F}_2^2,\mathsf{F}_3^2,\mathsf{F}_4^2} D^{\mathsf{F}_1^1,\mathsf{F}_2^1,\mathsf{F}_3^1,\mathsf{F}_4^1}. \end{aligned}$$

Property (4) can also be obtained from Equation (8).

Property (3) is interesting and important, which makes 3-QDTMs behave like CMs, and further it allows to give the exact probability of 3-differential.

4.2 Exact Probabilities from Quasi-3-differential Trails

Motivated by the notion of quasidifferential trail, we first define the quasi-3-differential trail, and then show that the exact expressions for the probabilities of 3-differentials can be given in terms of the correlations of quasi-3-differential trails. Form now on, we assume that $\mathsf{F}_1 = \mathsf{F}_2 = \mathsf{F}_3 = \mathsf{F}_4$, whose meaning is that we use the same encrypt oracle to encrypt multiple plaintexts. However, the results in this subsection can be generalized to the related-key setting easily, where these functions differ only in additive-key operation which is easy to deal with in practice.

Definition 11. *A quasi-3-differential trail for a function* $\mathsf{E} = \mathsf{E}_r \circ \cdots \circ \mathsf{E}_2 \circ \mathsf{E}_1$ *is a sequence* $\varpi_1, \varpi_2, \ldots, \varpi_{r+1}$ *of mask-3-differential quartets, where* $\varpi_i = (u, \alpha_i, \beta_i, \gamma_i), 1 \le i \le r+1$. *The correlation of this quasi-3-differential trail is defined as* $\prod_{i=1}^{r} D^{\mathsf{E}_i}_{\varpi_{i+1},\varpi_i}$.

Quasi-3-differential trails with $u_1 = u_2 = \cdots = u_{r+1} = 0$ correspond to 3-differential characteristics. Their correlation is equal to the product of one-round probabilities of the 3-differential characteristic with 3-differences sequence $(\alpha_1, \beta_1, \gamma_1), \ldots, (\alpha_{r+1}, \beta_{r+1}, \gamma_{r+1})$:

$$\prod_{i=1}^{r} D^{\mathsf{E}}_{(0,\alpha_{i+1},\beta_{i+1},\gamma_{i+1}),(0,\alpha_i,\beta_i,\gamma_i)} = \prod_{i=1}^{r} \Pr[(\alpha_i, \beta_i, \gamma_i) \xrightarrow[\boldsymbol{x}]{\mathsf{E}_i} (\alpha_{i+1}, \beta_{i+1}, \gamma_{i+1})]$$

with $\boldsymbol{x}$ uniform random on $\mathbb{F}_2^n$. This follows from Definition 11 and Equation (8).

Theorem 2 (3) indicates that the sum of the correlations of all quasi-3-differential trails with input and output mask-3-differential quartets $\varpi_1 = (0, \alpha_1, \beta_1, \gamma_1)$ and $\varpi_{r+1} = (0, \alpha_{r+1}, \beta_{r+1}, \gamma_{r+1})$ respectively, is equal to the exact probability of the 3-differential with input 3-difference $(\alpha_1, \beta_1, \gamma_1)$ and output 3-difference $(\alpha_{r+1}, \beta_{r+1}, \gamma_{r+1})$. Explicitly, expanding the coordinates of $D^{\mathsf{E}} = \prod_{i=1}^{r} D^{\mathsf{E}_i}$ corresponding to this 3-differential yields

$$D^{\mathsf{E}}_{\varpi_{r+1},\varpi_1} = \sum_{\varpi_2,\ldots,\varpi_r} \prod_{i=1}^{r} D^{\mathsf{E}_i}_{\varpi_{i+1},\varpi_i}. \tag{9}$$

This result can be trivially generalized to d-differential, *i.e.*, we can give the exact probabilitiy of polytopic cryptanalysis. Actually, quasi-3-differential trails also allow computing the probability of a 3-differential characteristic as shown in the following theorem.

Theorem 3. *Let* $\mathsf{E} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ *be a function and* $\mathsf{E} = \mathsf{E}_r \circ \cdots \circ \mathsf{E}_1$. *The probability of a 3-differential characteristic with 3-differences* $(\alpha_1, \beta_1, \gamma_1), \ldots, (\alpha_{r+1}, \beta_{r+1}, \gamma_{r+1})$ *is equal to the sum of the correlations of all quasi-3-differential trails with the same intermediate 3-differences:*

$$\Pr[\wedge_{i=1}^{r}(\alpha_i, \beta_i, \gamma_i) \xrightarrow[\boldsymbol{x}_i]{\mathsf{E}_i} (\alpha_{i+1}, \beta_{i+1}, \gamma_{i+1})] = \sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} D^{\mathsf{E}_i}_{(u_{i+1},\alpha_{i+1},\beta_{i+1},\gamma_{i+1}),(u_i,\alpha_i,\beta_i,\gamma_i)}$$

with $u_1 = u_{r+1} = 0$, $\boldsymbol{x}_i = \mathsf{E}_{i-1}(\boldsymbol{x}_{i-1})$ *for* $2 \le i \le r$, *and* $\boldsymbol{x}_1$ *uniform random on* $\mathbb{F}_2^n$.

Proof. Note that the expression of $D^{\mathsf{E}_i}_{(u_{i+1},\alpha_{i+1},\beta_{i+1},\gamma_{i+1}),(u_i,\alpha_i,\beta_i,\gamma_i)}$ is given by Equation (8), thus we have

$$\begin{aligned}
&\prod_{i=1}^{r} D^{\mathsf{E}_i}_{(u_{i+1},\alpha_{i+1},\beta_{i+1},\gamma_{i+1}),(u_i,\alpha_i,\beta_i,\gamma_i)} \\
&= \frac{1}{2^{nr}} \sum_{\substack{x_1,\ldots,x_r \\ (\alpha_i,\beta_i,\gamma_i) \xrightarrow[x_i]{\mathsf{E}_i} (\alpha_{i+1},\beta_{i+1},\gamma_{i+1})}} \prod_{i=1}^{r} (-1)^{u_i^\top x_i + u_{i+1}^\top \mathsf{E}_i(x_i)} \\
&= \frac{1}{2^{nr}} \sum_{\substack{x_1,\ldots,x_r \\ (\alpha_i,\beta_i,\gamma_i) \xrightarrow[x_i]{\mathsf{E}_i} (\alpha_{i+1},\beta_{i+1},\gamma_{i+1})}} \prod_{i=1}^{r-1} (-1)^{u_{i+1}^\top x_{i+1} + u_{i+1}^\top \mathsf{E}_i(x_i)}.
\end{aligned}$$

Summing over $u_2, \ldots, u_r$, then

$$\begin{aligned}
&\sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} D^{\mathsf{E}_i}_{(u_{i+1},\alpha_{i+1},\beta_{i+1},\gamma_{i+1}),(u_i,\alpha_i,\beta_i,\gamma_i)} \\
&= \frac{1}{2^{nr}} \sum_{\substack{x_1,\ldots,x_r \\ (\alpha_i,\beta_i,\gamma_i) \xrightarrow[x_i]{\mathsf{E}_i} (\alpha_{i+1},\beta_{i+1},\gamma_{i+1})}} \prod_{i=1}^{r-1} \Big(\sum_{u_{i+1}} (-1)^{u_{i+1}^\top x_{i+1} + u_{i+1}^\top \mathsf{E}_i(x_i)} \Big) \\
&= \frac{1}{2^{n}} \sum_{\substack{x_1,\ldots,x_r \\ (\alpha_i,\beta_i,\gamma_i) \xrightarrow[x_i]{\mathsf{E}_i} (\alpha_{i+1},\beta_{i+1},\gamma_{i+1})}} \prod_{i=1}^{r-1} \delta_{x_{i+1}}(\mathsf{E}_i(x_i)).
\end{aligned}$$

The right hand side is indeed equal to $\Pr[\wedge_{i=1}^{r}(\alpha_i, \beta_i, \gamma_i) \xrightarrow[\boldsymbol{x}_i]{\mathsf{E}_i} (\alpha_{i+1}, \beta_{i+1}, \gamma_{i+1})]$ when writing it in terms of probabilities.

Note that the complexity of calculating the probability of 3-differential is very high using Equation (9), because every $\varpi_i \in \mathbb{F}_2^{4n}$. However, Theorem 3 implies that the correlations of some quasi-3-differential trails (corresponding to the same characteristic) may be computed together, which can reduce the computing complexity. This is what we will do in the following subsection.

4.3 Interpretation of Quasi-3-differential Trails

The coordinates of D^{F} can be interpreted as the correlations of linear approximations between the input and output values satisfying a certain difference equation. From Equation (8), we know that $|D^{\mathsf{F}}_{(v,\boldsymbol{\beta}),(u,\boldsymbol{\alpha})}|$ never exceeds $D^{\mathsf{F}}_{(0,\boldsymbol{\beta}),(0,\boldsymbol{\alpha})}$, the probability of a 3-differential $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. While other quasi-3-differential trails with nontrivial masks may also have the absolute highest correlation. Quasi-3-differential trails with absolute correlation equaling the correlation of the corresponding 3-differential characteristic are of particular interest. They correspond to deterministic linear relations on the intermediate values which satisfy a certain difference equation, from which we have the following useful results. Actually, the following results are generalizations of [8, Theorem 4.2], and the proofs will be omitted; interested readers can refer to [8].

Theorem 4. *For a function $\mathsf{E} = \mathsf{E}_r \circ \cdots \circ \mathsf{E}_1$ and a 3-differential characteristic $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1}$ with correlation p (as quasi-3-differential trail with mask 0), it holds that:*
(1) If $(u_1, \boldsymbol{\alpha}_1), (u_2, \boldsymbol{\alpha}_2), \ldots, (u_{r+1}, \boldsymbol{\alpha}_{r+1})$ is a quasi-3-differential trail with correlation $(-1)^b p$ where $b \in \{0, 1\}$, then for any quasi-3-differential trail $(v_1, \boldsymbol{\alpha}_1), (v_2, \boldsymbol{\alpha}_2), \ldots, (v_{r+1}, \boldsymbol{\alpha}_{r+1})$ with correlation c, the correlation of the quasi-3-differential trail $(u_1 + v_1, \boldsymbol{\alpha}_1), (u_2 + v_2, \boldsymbol{\alpha}_2), \ldots, (u_{r+1} + v_{r+1}, \boldsymbol{\alpha}_{r+1})$ is $(-1)^b c$.
(2) If the correlations of any number of quasi-3-differential trails with 3-differences $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1}$ and correlation $\pm p$ sum to zero, then the probability of the 3-differential characteristic $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1}$ is zero.

From the above theorem, we find that if $(u_1, \boldsymbol{\alpha}_1), (u_2, \boldsymbol{\alpha}_2), \ldots, (u_{r+1}, \boldsymbol{\alpha}_{r+1})$ and $(v_1, \boldsymbol{\alpha}_1), (v_2, \boldsymbol{\alpha}_2), \ldots, (v_{r+1}, \boldsymbol{\alpha}_{r+1})$ are two quasi-3-differential trails with absolute correlation p, then $(u_1 + v_1, \boldsymbol{\alpha}_1), (u_2 + v_2, \boldsymbol{\alpha}_2), \ldots, (u_{r+1} + v_{r+1}, \boldsymbol{\alpha}_{r+1})$ is also a quasi-3-differential trail with absolute correlation p. Therefore, the masks (viewed as vectors, *e.g.*, $u_1||u_2||\ldots||u_{r+1}$) of all quasi-3-differential trails with absolute correlation p form a linear subspace. That is to say, the number of all quasi-3-differential trails with absolute correlation p is a power of two. Furthermore, for a 3-differential characteristic $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1}$, if there exists a quasi-3-differential trail with correlation $-p$, then the probability of this characteristic $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1}$ must be zero, this is because in this case the number of quasi-3-differential trails respectively with correlation p and $-p$ must be equal.

From the above analysis, quasi-3-differential trails with the highest absolute correlation are important to determine the probability of the corresponding 3-differential characteristic. Similar to the discussions in [8], quasi-3-differential trails with high correlations should not activate many differentially inactive S-boxes, *i.e.*, the masks of the quasi-3-differential trail should follow the 3-differences

as closely as possible, which is more likely if the linear layer L satisfies $L^\top = L^{-1}$ by Theorem 2 (4). Liner layer with this property is commonly used in lightweight block ciphers, including all bit-permutations.

5 Revisiting the Boomerang Attack

The boomerang attack proposed by Wagner is a variation of differential cryptanalysis. Its main idea is to combine two short differential trails to get a long one with a high probability.

5.1 The Boomerang Attack

The framework: In a boomerang attack, a cipher E is regarded as the composition of two sub-ciphers E_0 and E_1, *i.e.*, $E = E_1 \circ E_0$. Suppose there exists a differential $\alpha \to \beta$ of E_0 with probability p and a differential $\gamma \to \delta$ of E_1 with probability q. The two differentials are then combined in an adaptive chosen plaintext/ciphertext attack setting to construct a long boomerang distinguisher, as shown in Fig. 3 (left). Initially, the boomerang attack was proposed in the single-key setting. Later, the basic boomerang attack was extended to the related-key setting [9].

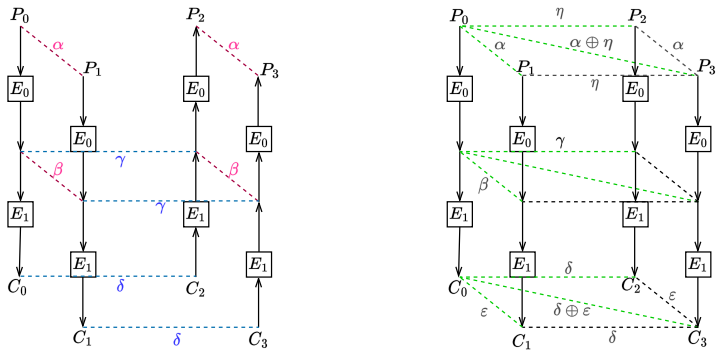

Fig. 3: Basic boomerang attack (left), and boomerang attack in the 3-difference view (right), the differences η and ε are allowed to take any value.

Let $E_K(P)$ and $E_K^{-1}(C)$ denote the encryption of P and the decryption of C under a key K, respectively. Suppose ΔK and ∇K are the master key differences of the mentioned two differentials. Then the boomerang attack in the related-key setting works as follows.

1. $K_0 \leftarrow K$, $K_1 \leftarrow K_0 \oplus \Delta K$, $K_2 \leftarrow K_0 \oplus \nabla K$ and $K_3 \leftarrow K_0 \oplus \Delta K \oplus \nabla K$.
2. Repeat the following steps many times.
(1) $P_0 \leftarrow random()$ and $P_1 \leftarrow P_0 \oplus \alpha$
(2) $C_0 \leftarrow E_{K_0}(P_0)$ and $C_1 \leftarrow E_{K_1}(P_1)$
(3) $C_2 = C_0 \oplus \delta$ and $C_3 = C_1 \oplus \delta$
(4) $P_2 = E_{K_2}^{-1}(C_2)$ and $P_3 = E_{K_3}^{-1}(C_3)$
(5) Check if $P_2 \oplus P_3 = \alpha$.

If $P_2 \oplus P_3 = \alpha$ holds, then a right quartet (P_0, P_1, P_2, P_3) is found such that $P_0 \oplus P_1 = P_2 \oplus P_3 = \alpha$ and $C_0 \oplus C_2 = C_1 \oplus C_3 = \delta$. This event happens with probability p^2q^2 under the assumption that the two differentials are independent and the probability is obtained as

$$\Pr[E^{-1}(E(P_0) + \delta) + E^{-1}(E(P_0 + \alpha) + \delta) = \alpha] = p^2q^2.$$

However, there maybe exists dependency between the two differential trails, which highly affects the probability of the boomerang distinguisher. As pointed out by Murphy in [27], there exist cases where the probabilities formulated by p^2q^2 are highly inaccurate. He showed that in some cases of S-box based ciphers, two independently chosen differential trails are incompatible, making the boomerang never return, and in other cases, the dependency leads to a higher probability than p^2q^2. Further, Biryukov *et al.* made an improvement on exploiting the positive dependency of boomerang distinguishers, which was named boomerang switch [11].

The above phenomenons show that the foundations of the boomerang attacks need to be revisited. Kidmose and Tiessen dived into the boomerang attacks and analysed the probabilities of boomerang distinguishers theoretically in [24], and got the following counter-intuitive result.

Theorem 5 (Theorem 1 in [24]). *Assume that we have a boomerang as described above of probability p^2q^2 and assume that the assumption of the independence of differentials holds (the upper and lower differentials). Then there exist differentials $\alpha \xrightarrow{E} \varepsilon$ and $\eta \xrightarrow{E} \delta$ over the whole cipher (see Fig. 3 (right)) with probabilities at least pq^2 and p^2q, respectively.*

Theorem 5 implies that there always exist differentials that beat boomerang distinguishers. However, in the view of 3-differential, as illustrated in Fig. 3 (right), Kidmose and Tiessen also gave the following result.

Theorem 6 (Theorem 2 in [24]). *Let A be the affine subspace of all 3-differences which correspond to an input quartet:*

$$A = \{(\alpha, \eta, \alpha \oplus \eta) \in \mathbb{F}_2^{3n} \mid \eta \in \mathbb{F}_2^n\}.$$

Let B be the set of all 3-differences which correspond to a right ciphertext quartet:

$$B = \{(\varepsilon, \delta, \varepsilon \oplus \delta) \in \mathbb{F}_2^{3n} \mid \varepsilon \in \mathbb{F}_2^n\}.$$

The probability of the return of the boomerang is then equal to the probability of the truncated 3-difference transition $A \xrightarrow{E} B$ multiplied by 2^n:

$$\Pr(\text{Boomerang returns}) = 2^n \cdot \Pr(A \xrightarrow{E} B).$$

From Equation (6), we have

$$\begin{aligned}\Pr(\text{Boomerang returns}) &= \sum_{\eta,\varepsilon\in\mathbb{F}_2^n} \Pr\Big((\alpha,\eta,\alpha\oplus\eta) \xrightarrow{E} (\varepsilon,\delta,\varepsilon\oplus\delta)\Big) \\ &\geq \Pr\Big((\alpha,\alpha,0) \xrightarrow{E} (\delta,\delta,0)\Big) \\ &= \Pr(\alpha \xrightarrow{E} \delta)\end{aligned}$$

which implies that for every differential, there always exists a boomerang distinguisher better than it. This result contradicts to the conclusion that there always exist differentials that beat boomerang distinguishers coming from Theorem 5. The root of the contradiction is the assumption of the independence of differentials (the upper and lower differentials). In fact, some researchers also noticed that there exists dependency between the two differential trails in previous work, which led to the sandwich attack [19, 20], an improvement of the boomerang attack. In a sandwich attack, E is divided into 3 parts, *i.e.*,

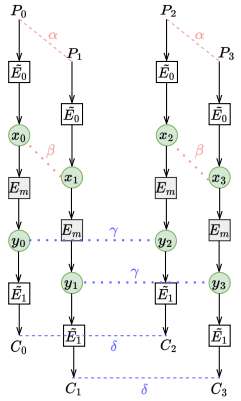

Fig. 4: The sandwich attack

$E = \tilde{E}_1 \circ E_m \circ \tilde{E}_0$ as illustrated in Fig. 4, where the middle part E_m specifically handles the dependency and contains a relatively small number of rounds. If the probability of generating a right quartet for E_m is r, then the probability of the whole boomerang distinguisher is

$$\Pr[E^{-1}(E(P_0)+\delta) + E^{-1}(E(P_0+\alpha)+\delta) = \alpha] = \tilde{p}^2\tilde{q}^2 r,$$

where $\tilde{p}$ (resp. $\tilde{q}$) is the probability of the differential of $\tilde{E_0}$ (resp. $\tilde{E_1}$). Let (x_0, x_1, x_2, x_3) and (y_0, y_1, y_2, y_3) be input and output quartets for E_m, where $y_i = E_m(x_i)$, $0 \leq i \leq 3$. Suppose $x_0 \oplus x_1 = x_2 \oplus x_3 = \beta$ and $y_0 \oplus y_2 = y_1 \oplus y_3 = \gamma$. Then, r was formally defined as

$$r = \Pr[x_2 \oplus x_3 = \alpha \mid (x_0 \oplus x_1 = \beta) \wedge (y_0 \oplus y_2 = \gamma) \wedge (y_1 \oplus y_3 = \gamma)].$$

In [19, 20], the probability r of E_m was evaluated by experiments. In [15], a new tool named `BCT` was proposed, which can calculate r theoretically when E_m is composed of a single S-box layer. When E_m contains multiple rounds, inspired by `BCT`, many tables were proposed to estimate the probability of it [13, 17, 21, 29, 33, 35]. Almost all the previous work calculating the probability of middle part E_m are under the assumption that the rounds of a cipher are independent and the round keys are random. However, for most lightweight block ciphers, the key schedules are simple and the probability of middle part E_m obtained under the standard assumption may deviate from the real value significantly. As far as we know, only the `DBCT`, proposed in [21] and further studied in [35], can handle dependency in two rounds to some extent.

5.2 The Probabilities of Boomerang Distinguishers

In boomerang attacks, calculating the probabilities of boomerang distinguishers is an essential step, because it directly determines the attack complexity. Although there are many works on this topic, the exact formula of the probability of boomerang distinguisher is unknown till now, let alone how to accurately compute it. Most previous works under the assumption that the differential propagation throughout the rounds are independent, and use all kinds of heuristic methods to deal with the dependency in upper and lower trails. Recently, from the perspective of 3-differential, Kidmose and Tiessen [24] provided a framework that allows formulating precisely the probability of a boomerang distinguisher without relying on independence assumptions of the trails, but only depending on the assumption that intermediate differentials are independent as commonly made in differential cryptanalysis. Note that for a single 3-differential characteristic, its exact probability is given in Theorem 3 in terms of quasi-3-differential trails. Therefore, we have the following result.

Theorem 7. *For a cipher* E*, it can be regarded as* $\mathsf{E} = \mathsf{E}_r \circ \cdots \circ \mathsf{E}_1$*. Assume that the input difference of the upper differential is* α *and the output difference of the lower differential is* δ *in the boomerang attack. Then the probability of the return of the boomerang is equal to*

$$\Pr(\text{Boomerang returns}) = \sum_{\varpi_1, \varpi_2, \ldots, \varpi_r, \varpi_{r+1}} \prod_{i=1}^{r} D^{\mathsf{E}_i}_{\varpi_{i+1}, \varpi_i} \tag{10}$$

where $\varpi_1 = (0, \alpha, \eta, \alpha \oplus \eta)$ *and* $\varpi_{r+1} = (0, \varepsilon, \delta, \varepsilon \oplus \delta)$*,* $\eta, \varepsilon \in \mathbb{F}_2^n$*, and mask-3-differential quartets* $\varpi_i = (u_i, \alpha_i, \beta_i, \gamma_i) \in \mathbb{F}_2^{4n}$ *for* $2 \leq i \leq r$*.*

In Theorem 7, we give the exact formula of the probability of the boomerang distinguisher theoretically without any assumptions. This result may be helpful to deepen the understanding of the boomerang attacks for researchers in this area. Naturally, one may ask how accurate is the probability obtained under the sandwich attack framework.

We find that under the assumption that intermediate differentials are independent the probability obtained under the sandwich attack framework can be seen as a kind of *average*. Specifically, if we assume that the three parts $\tilde{E}_0, E_m$, and $\tilde{E}_1$ are independent, then

$$
\begin{aligned}
&\Pr(\text{Boomerang returns}) = \sum_{\eta,\varepsilon\in\mathbb{F}_2^n} \Pr[(\alpha,\eta,\alpha\oplus\eta) \xrightarrow{E} (\varepsilon,\delta,\varepsilon\oplus\delta)] \\
&= \sum_{\eta,\varepsilon,x,y\in\mathbb{F}_2^n} \Pr[(\alpha,\eta,\alpha\oplus\eta) \xrightarrow{\tilde{E}_0} (\beta,x,\beta\oplus x)] \times \\
&\quad \Pr[(\beta,x,\beta\oplus x) \xrightarrow{E_m} (y,\gamma,y\oplus\gamma)] \cdot \Pr[(y,\gamma,y\oplus\gamma) \xrightarrow{\tilde{E}_1} (\varepsilon,\delta,\varepsilon\oplus\delta)] \\
&\overset{?}{=} \frac{\sum_{\eta,x\in\mathbb{F}_2^n} \Pr[(\alpha,\eta,\alpha\oplus\eta) \xrightarrow{\tilde{E}_0} (\beta,x,\beta\oplus x)]}{2^n} \times \Big(\sum_{x,y\in\mathbb{F}_2^n} \Pr[(\beta,x,\beta\oplus x) \xrightarrow{E_m} (y,\gamma,y\oplus\gamma)] \Big) \\
&\quad \times \frac{\sum_{y,\varepsilon\in\mathbb{F}_2^n} \Pr[(y,\gamma,y\oplus\gamma) \xrightarrow{\tilde{E}_1} (\varepsilon,\delta,\varepsilon\oplus\delta)]}{2^n} \\
&\overset{\star}{=} \Big(\Pr[\alpha \xrightarrow{\tilde{E}_0} \beta] \Big)^2 \Big(\Pr[\gamma \xrightarrow{\tilde{E}_1} \delta] \Big)^2 \Big(\sum_{x,y\in\mathbb{F}_2^n} \Pr[(\beta,x,\beta\oplus x) \xrightarrow{E_m} (y,\gamma,y\oplus\gamma)] \Big) \\
&= \tilde{p}^2\tilde{q}^2 r,
\end{aligned}
$$

where $r = \sum_{x,y\in\mathbb{F}_2^n} \Pr[(\beta,x,\beta\oplus x) \xrightarrow{E_m} (y,\gamma,y\oplus\gamma)]$ is equivalent to the definition in the sandwich attack, and "$\star$" comes from the following lemma.

Lemma 1 (Lemma 1 in [24]). *The average of the probability for a 3-difference $(\alpha,\eta,\alpha\oplus\eta)$ to be mapped by a function F to a 3-difference of type $(\beta,\gamma,\beta\oplus\gamma)$ for some $\gamma\in\mathbb{F}_2^n$ overall $\eta\in\mathbb{F}_2^n$ is equal to the square of the probability of the differential $\alpha \xrightarrow{\mathsf{F}} \beta$:*

$$
2^{-n} \sum_{\gamma,\eta\in\mathbb{F}_2^n} \Pr[(\alpha,\eta,\alpha\oplus\eta) \xrightarrow{\mathsf{F}} (\beta,\gamma,\beta\oplus\gamma)] = \Big(\Pr[\alpha \xrightarrow{\mathsf{F}} \beta] \Big)^2 .
$$

Actually, similar to the proof of the above lemma, we also have

$$
2^{-n} \sum_{\alpha,\beta\in\mathbb{F}_2^n} \Pr[(\alpha,\eta,\alpha\oplus\eta) \xrightarrow{\mathsf{F}} (\beta,\gamma,\beta\oplus\gamma)] = \Big(\Pr[\eta \xrightarrow{\mathsf{F}} \gamma] \Big)^2 .
$$

The "?" step can be seen as a kind of average, but it maybe deviates from the real probability significantly for some extreme cases. Therefore, estimating the probability under the sandwich attack framework must be careful.

If we do not care about the gap mentioned before, the remaining important task is calculating the probability of the middle part E_m in the sandwich attack. In recent years, there are really many works on it. In the following subsection, we will give a reasonable explanation for previous works on estimating the probability of the middle part E_m under the assumption that intermediate differentials are independent.

5.3 The Relationships Between 3-differential and all Kinds of Tables

Since Cid *et al.* introduced the BCT, many tables were proposed to calculate the probabilities of the middle part E_m. In fact, these tables can be described using 3-differentials, which may be an essential language to describe the quartet-based attacks. First, let's recall the definitions of some tables of an S-box.

Definition 12. *Let S be an n-bits S-box. The* BCT, UBCT, LBCT *and* EBCT *are defined as (refer to Fig. 5 (left)):*

$$\begin{aligned}
\mathtt{BCT}(\alpha_1,\beta_2) &= \#\{x\in\mathbb{F}_2^n \mid S^{-1}(S(x)\oplus\beta_2)\oplus S^{-1}(S(x\oplus\alpha_1)\oplus\beta_2)=\alpha_1\},\\
\mathtt{UBCT}(\alpha_1,\alpha_2,\beta_2) &= \#\left\{x\in\mathbb{F}_2^n \middle| \begin{array}{c} S(x)\oplus S(x\oplus\alpha_1)=\alpha_2 \\ S^{-1}(S(x)\oplus\beta_2)\oplus S^{-1}(S(x\oplus\alpha_1)\oplus\beta_2)=\alpha_1 \end{array}\right\},\\
\mathtt{LBCT}(\alpha_1,\beta_1,\beta_2) &= \#\left\{x\in\mathbb{F}_2^n \middle| \begin{array}{c} S(x)\oplus S(x\oplus\beta_1)=\beta_2 \\ S^{-1}(S(x)\oplus\beta_2)\oplus S^{-1}(S(x\oplus\alpha_1)\oplus\beta_2)=\alpha_1 \end{array}\right\},\\
\mathtt{EBCT}(\alpha_1,\alpha_2,\beta_1,\beta_2) &= \#\left\{x\in\mathbb{F}_2^n \middle| \begin{array}{c} S(x)\oplus S(x\oplus\alpha_1)=\alpha_2 \\ S(x)\oplus S(x\oplus\beta_1)=\beta_2 \\ S^{-1}(S(x)\oplus\beta_2)\oplus S^{-1}(S(x\oplus\alpha_1)\oplus\beta_2)=\alpha_1 \end{array}\right\}.
\end{aligned}$$

In [24, Theorem 6], the entries of BCT are expressed as

$$\mathtt{BCT}(\alpha_1,\beta_2)=\sum_{\eta,\varepsilon\in\mathbb{F}_2^n}\#\{x\in\mathbb{F}_2^n \mid (\alpha_1,\eta,\alpha_1\oplus\eta)\xrightarrow[x]{S}(\varepsilon,\beta_2,\varepsilon\oplus\beta_2)\}.$$

We can also rewrite the entries of other tables in terms of 3-differentials without proofs as follows:

$$\begin{aligned}
\mathtt{UBCT}(\alpha_1,\alpha_2,\beta_2) &= \sum_{\eta\in\mathbb{F}_2^n}\#\{x\in\mathbb{F}_2^n \mid (\alpha_1,\eta,\alpha_1\oplus\eta)\xrightarrow[x]{S}(\alpha_2,\beta_2,\alpha_2\oplus\beta_2)\},\\
\mathtt{LBCT}(\alpha_1,\beta_1,\beta_2) &= \sum_{\varepsilon\in\mathbb{F}_2^n}\#\{x\in\mathbb{F}_2^n \mid (\alpha_1,\beta_1,\alpha_1\oplus\beta_1)\xrightarrow[x]{S}(\varepsilon,\beta_2,\varepsilon\oplus\beta_2)\},\\
\mathtt{EBCT}(\alpha_1,\alpha_2,\beta_1,\beta_2) &= \#\{x\in\mathbb{F}_2^n \mid (\alpha_1,\beta_1,\alpha_1\oplus\beta_1)\xrightarrow[x]{S}(\alpha_2,\beta_2,\alpha_2\oplus\beta_2)\}.
\end{aligned}$$

Recently, when Li *et al.* [26] mounted the rectangle attack to GIFT [4], they introduced a new tool named GBCT, which is the generalization of BCT.

Definition 13. *Let S be an n-bits S-box. The* GBCT *is defined as (refer to Fig. 5 (right)):*

$$\mathtt{GBCT}(\alpha_1,\alpha_2,\beta_1,\beta_2)=\#\left\{x\in\mathbb{F}_2^n \mid S^{-1}(S(x)\oplus\beta_1)\oplus S^{-1}(S(x\oplus\alpha_1)\oplus\beta_2)=\alpha_2\right\}.$$

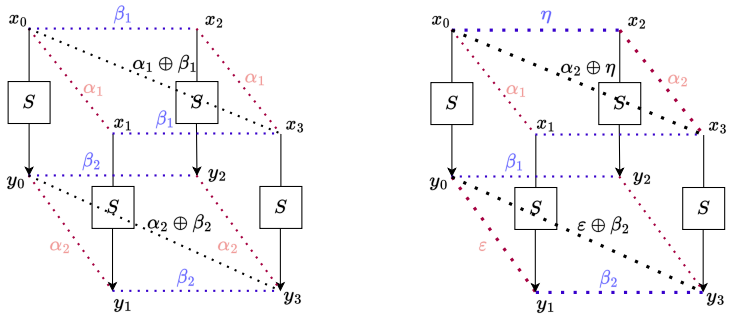

Fig. 5: Two upper (or lower) differentials are equal (left) or not equal (right).

It can be rewritten as

$$\texttt{GBCT}(\alpha_1, \alpha_2, \beta_1, \beta_2) = \sum_{\eta, \varepsilon \in \mathbb{F}_2^n} \#\{x \in \mathbb{F}_2^n \mid (\alpha_1, \eta, \alpha_2 \oplus \eta) \xrightarrow[x]{S} (\varepsilon, \beta_1, \varepsilon \oplus \beta_2)\}.$$

Actually, the definitions of other tables can also be generalized, as done for GBCT, as they also need not require that the two upper differentials or two lower differentials are equal.

All the above tables are defined for one S-box, while in order to more efficiently calculate the probabilities of boomerang distinguishers, Hadipour *et al.* introduced the DBCT, which is defined for two S-boxes in a row.

Definition 14. *Let S be an n-bits S-box. The* DBCT *is defined as (refer to Fig. 6 (left)):*

$$\begin{aligned}\texttt{DBCT}(\alpha_1, \beta_3) &= \sum_{\alpha_2, \beta_2 \in \mathbb{F}_2^n} \texttt{UBCT}(\alpha_1, \alpha_2, \beta_2)\texttt{LBCT}(\alpha_2, \beta_2, \beta_3) \\ &= \sum_{\alpha_2, \beta_2 \in \mathbb{F}_2^n} \Big(\sum_{\eta \in \mathbb{F}_2^n} \#\{x \in \mathbb{F}_2^n \mid (\alpha_1, \eta, \alpha_1 \oplus \eta) \xrightarrow[x]{S} (\alpha_2, \beta_2, \alpha_2 \oplus \beta_2)\} \Big) \times \\ &\quad \Big(\sum_{\varepsilon \in \mathbb{F}_2^n} \#\{x \in \mathbb{F}_2^n \mid (\alpha_2, \beta_2, \alpha_2 \oplus \beta_2) \xrightarrow[x]{S} (\varepsilon, \beta_3, \varepsilon \oplus \beta_3)\} \Big).\end{aligned}$$

In fact, using DBCT to compute the probability of E_m cannot bring any advantage on accuracy compared with previous techniques, but can reduce the time complexity. In order to improve the accuracy, we generalize the definition of DBCT as (see Fig. 6 (right)):

$$\begin{aligned}\texttt{DBCT}^*(\alpha_1, \beta_3) = \sum_{\alpha_2, \beta_2, \gamma_2 \in \mathbb{F}_2^n} &\Big(\sum_{\eta \in \mathbb{F}_2^n} \#\{x \in \mathbb{F}_2^n \mid (\alpha_1, \eta, \alpha_1 \oplus \eta) \xrightarrow[x]{S} (\alpha_2, \beta_2, \gamma_2)\} \Big) \times \\ &\Big(\sum_{\varepsilon \in \mathbb{F}_2^n} \#\{x \in \mathbb{F}_2^n \mid (\alpha_2, \beta_2, \gamma_2) \xrightarrow[x]{S} (\varepsilon, \beta_3, \varepsilon \oplus \beta_3)\} \Big).\end{aligned}$$

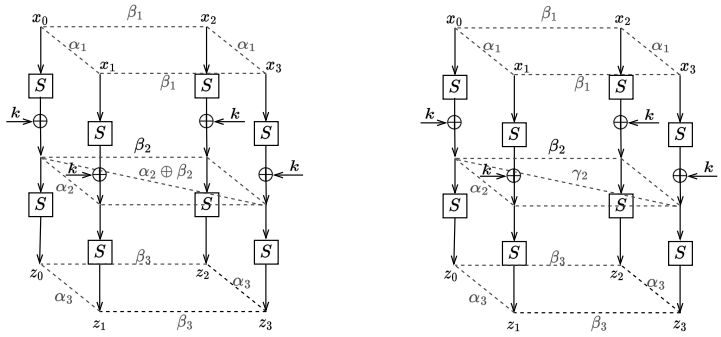

Fig. 6: The sketch map of DBCT (left) and DBCT* (right)

It is easy to see that the maximal value of DBCT* is greater than or equal to that of DBCT. As a comparison, we computed the maximal values of DBCT* for several sboxes and list them in Table 1, in which the maximal values of DBCT are taken from [35]. We find that the maximal value of DBCT* is much bigger than that of DBCT, which may impact the probabilities of boomerang attacks significantly. It is noticed that, due to the existence of linear layer between successive two S-boxes, the maximal values and distributions of DBCT and DBCT* may vary from cipher to cipher even for the same S-box if we consider the effect of linear layers, which implies that DBCT and DBCT* are different from other tables in essence.

Table 1: Comparing the maximal values of DBCT and DBCT*

S-box	CRAFT [5]	QARMA [1]	PRESENT [12]	GIFT [4]
DBCT	128	48	40	40
DBCT*	160	120	80	72
S-box	LBlock-s0 [34]	LBlock-s1 [34]	MIBS [22]	TWNIE [30]
DBCT	40	40	32	28
DBCT*	64	80	68	64

At last, we stress that from the view of 3-differentials it is natural to introduce all kinds of tables to calculate the probability of E_m. From Theorem 6, the probability of E_m can be expressed as the *sum-of-product* of the probabilities of one-round 3-differentials under the assumption that the intermediate differentials are independent. When computing the value of the sum-of-product, we can extract common factors and collect the similar terms, which can be precomputed and stored as a look-up table, then all kinds of tables will appear naturally. Note that introducing all kinds of tables to compute the probability cannot improve

its accuracy. In order to improve accuracy, it needs the theory proposed in this paper, and we will give an example in the following section.

6 Application to GIFT

GIFT is a family of block ciphers proposed by Banik *et al.* at CHES 2017 [4]. It consists of two versions, *i.e.*, GIFT-64 and GIFT-128, while in this paper we only focus on GIFT-64, a 64-bit block cipher with a 128-bit key and with 28 rounds, and its details will be given in next subsection.

The reasons why we choose GIFT as the target to illustrate the theory proposed in the present paper are the following: The intuition tells us that the AddRoundKey operation of GIFT, adopting half round key XOR, is very different from other block ciphers, and there maybe exist some problems in the differential-like attacks. In addition, the linear layer is a bit-permutation, and it potentially results in quasi-3-differential trails with high absolute correlations relative to the probability of the corresponding 3-differential characteristic.

6.1 Description of GIFT-64

Round Function. The round function of GIFT-64 consists of three operations: SubCells, PermBits, and AddRoundKey. For convenience, we consider the 64-bit state as 16 4-bit nibbles. The three operations of the round function are as follows:

1. **SubCells:** Nonlinear S-box substitutions are applied to each nibble, as is shown in Table 2. Its DDT is given in Table 6.

Table 2: The S-box of GIFT

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$GS(x)$	1	a	4	c	6	f	3	9	2	d	b	7	5	0	8	e

2. **PermBits:** This operation maps the bit from the position i of the cipher state to the position $P_{64}(i)$ as

$$b_{P_{64}(i)} \leftarrow b_i, i \in \{0, 1, \ldots, 63\},$$

where $P_{64}(i)$ can be calculated as

$$63-\left\{4\left\lfloor\frac{63-i}{16}\right\rfloor+16\left[3\left\lfloor\frac{(63-i) \bmod 16}{4}\right\rfloor+(63-i) \bmod 16\right]+(63-i) \bmod 4\right\} \bmod 64.$$

3. **AddRoundKey:** This step adds the round key and the round constant. In the r-th round, a 32-bit round key RK_r is extracted from the key state and is further partitioned into two 16-bit words as $RK_r = U||V =$

$u_{15} \cdots u_1 u_0 || v_{15} \cdots v_1 v_0$, Then, U and V are XORed to the cipher state as $b_{4i+1} \leftarrow b_{4i+1} \oplus u_i, b_{4i} \leftarrow b_{4i} \oplus v_i, i \in \{0, 1, \ldots, 15\}$. A single "1" and a 6-bit constant $C = c_5 c_4 c_3 c_2 c_1 c_0$ are added to each state at bit position 63, 23, 19, 15, 11, 7, 3 respectively, *i.e.*, $b_{63} \leftarrow b_{63} \oplus 1, b_{23} \leftarrow b_{23} \oplus c_5, b_{19} \leftarrow b_{19} \oplus c_4, b_{15} \leftarrow b_{15} \oplus c_3, b_{11} \leftarrow b_{11} \oplus c_2, b_7 \leftarrow b_7 \oplus c_1, b_3 \leftarrow b_3 \oplus c_0$.

Key Schedule. Split the master key K into 8 16-bit subkeys $k_7||k_6||...||k_1||k_0 \leftarrow K$. For each round, the round key consists of the last two significant subkeys, and then the key state is updated following $k_7||k_6||...||k_1||k_0 \leftarrow k_1 \ggg 2||k_0 \ggg 12||k_7||...||k_2$, where $\ggg i$ is an i-bit right rotation within a 16-bit word.

Different from other differential-like attacks, in our work we need to consider the round constants.

Round Constants. The values of the round constants are given in table 3, encoded to byte values for each round, with c_0 being the least significant bit.

Table 3: The round constants of GIFT-64

Rounds	Constants
0 - 13	01,03,07,0F,1F,3E,3D,3B,37,2F,1E,3C,39,33
14 - 27	27,0E,1D,3A,35,2B,16,2C,18,30,21,02,05,0B

We refer readers to [4] for more details of GIFT.

6.2 Analysis for 3-differentials

In order to search for optimal quasi-3-differential trails, we model the propagation of the masks for a characteristic as a Satisfiability Modulo Theories (SMT) problem, which is similar to existing SMT-based models for finding linear trails. To solve the SMT problem, we use Boolector [28] through its Python interface pyboolector. The model used in this section is also similar as that provided in [8].

In [14], Chen *et al.* mounted a 23-round key recovery attack for GIFT-64 based on a 19-round related key rectangle distinguisher, whose probability was claimed to be 1. However, Ji *et al.* recalculated the probability of the distinguisher and found that its probability is only 2^{-18} [23]. The distinguisher is listed in the following table 4.

For this distinguisher, there are in total about $2^{25.57}$ 3-differential characteristics. The number of optimal 3-differential characteristics is 2^{13}, and we list all of them in table 5, where $\clubsuit = \{1,2,4,7,a,b,e,f\}$, $\diamondsuit = \{2,a\}$, $\heartsuit = \{1,3,5,7,9,b,c,e\}$, $\spadesuit = \{2,4\}$; and $\tilde{\clubsuit} = \{1,6\}$, $\tilde{\diamondsuit} = \tilde{\spadesuit} = \{5,6\}$, $\tilde{\heartsuit} =$

Table 4: The propagation of E_m of the 19-round related-key rectangle distinguisher for GIFT-64 in [14]

Rounds	E_0					E_1				
10	β	0100	0000	0102	0200					
	β'	0800	0000	060a	0600	γ''	0000	0906	0000	0085
11	β''	00a2	0000	8020	0044	γ'	0000	050c	0a00	0000
						γ	0000	0802	0100	0000

$\beta' = S(\beta)$, $\beta'' = K \circ P(\beta')$, $\gamma' = S^{-1}(\gamma)$, $\gamma'' = P^{-1} \circ K^{-1}(\gamma')$.

$\{3,7,b,f\}$. For a differential of E_0, such as $0100000001020200 \rightarrow 08000000060a0600 \rightarrow 00a2000080200044 \rightarrow 0015000030500077$, and a differential of E_1, such as $0000010200000012 \rightarrow 0000090600000085 \rightarrow 0000050c0a000000 \rightarrow 0000080201000000$, the corresponding 3-differential characteristic is

$$(0100000001020200, 0000010200000012, 0100010201020212) \xrightarrow{\text{SubCells}}$$
$$(08000000060a0600, 0000090600000085, 08000906060a0685) \xrightarrow{\text{AddRoundKey}\circ\text{PermBits}}$$
$$(00a2000080200044, 0000050c0a000000, 00a2050c8a200044) \xrightarrow{\text{SubCells}}$$
$$(0015000030500077, 0000080201000000, 0015080231500077).$$

Table 5: The all possible optimal 3-differential trails of E_m of the 19-round related-key rectangle distinguisher for GIFT-64 in [14]

Rounds	E_0					E_1				
10	β	0100	0000	0102	0200		0000	0♣0♢	0000	00♡♠
	β'	0800	0000	060a	0600	γ''	0000	0906	0000	0085
11	β''	00a2	0000	8020	0044	γ'	0000	050c	0a00	0000
		00♣♢	0000	♡0♠0	0077	γ	0000	0802	0100	0000

$\beta' = S(\beta)$, $\beta'' = K \circ P(\beta')$, $\gamma' = S^{-1}(\gamma)$, $\gamma'' = P^{-1} \circ K^{-1}(\gamma')$.

It is amazing that the probability of every optimal 3-differential characteristic is zero. Such as, for the above 3-differential characteristic, there are in total 16 quasi-3-differential trails with absolute correlations equaling to the optimal 3-differential probability $2^{-38.83}$, but the sum of the correlations of the 16 quasi-3-differential trails is always zero, independently with the round keys. Although we did not include all quasi-3-differentials in the analysis, Theorem 4 (2) allows concluding that the characteristic has probability zero. That is to say, this characteristic has no contribution to the boomerang distinguisher. This is an interesting outcome of the approach itself since previous techniques are not able to find the fact. We remind that the above distinguisher is a related-key distinguisher, so for different encryption/decryption oracles the round keys are different, but it can be easily handled using formula (7).

Table 6: DDT of GIFT's 4-bit S-box

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	2	2	0	2	2	2	2	2	0	0	2
2	0	0	0	0	0	4	4	0	0	2	2	0	0	2	2	0
3	0	0	0	0	0	2	2	0	2	0	0	2	2	2	2	2
4	0	0	0	2	0	4	0	6	0	2	0	0	0	2	0	0
5	0	0	2	0	0	2	0	0	2	0	0	0	2	2	2	4
6	0	0	4	6	0	0	0	2	0	0	2	0	0	0	2	0
7	0	0	2	0	0	2	0	0	2	2	2	4	2	0	0	0
8	0	0	0	4	0	0	0	4	0	0	0	4	0	0	0	4
9	0	2	0	2	0	0	2	2	2	0	2	0	2	2	0	0
a	0	4	0	0	0	0	4	0	0	2	2	0	0	2	2	0
b	0	2	0	2	0	0	2	2	2	2	0	0	2	0	2	0
c	0	0	4	0	4	0	0	0	2	0	2	0	2	0	2	0
d	0	2	2	0	4	0	0	0	0	0	2	2	0	2	0	2
e	0	4	0	0	4	0	0	0	2	2	0	0	2	2	0	0
f	0	2	2	0	4	0	0	0	0	2	0	2	0	0	2	2

The sum of the probabilities of all the optimal 3-differential characteristics is $2^{-38.83} \times 2^{13} = 2^{-25.83}$ calculated under the assumption that the intermediate differentials are independent, which is included in the probability of E_m estimated by previous techniques. Although it is indeed zero, compared with the probability 2^{-18} given in [23], it has little impact on the result. Therefore, in order to accurately calculate the probability of the middle part E_m of a sandwich attack, we have to sum the correlations of all the quasi-3-differential trails, that is, using the formula (10), but the computing complexity is so high that we cannot complete it in a reasonable time even for a small example.

The reason why the above situation occurs is that the AddRoundKey operation of GIFT is the so-called half-round key XOR. Recently, Baksi [2] also noticed that half-round key XOR leads to the undesired consequence that the security against the differential/linear attacks is overestimated. Therefore, for the new block cipher − BAKSHEESH [3], a GIFT-like cipher, but with improved efficiency, the designers replaced half-round key XOR with full-round key XOR.

In the following, we will explain that in the case of full round key XOR, for each 3-differential characteristic, say $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1})$, there at least exists one round key such that its probability is nonzero under the assumption that the round keys are independent. Let $\mathsf{E} = \mathsf{E}_r \circ \cdots \circ \mathsf{E}_1$ and $\mathsf{E}_i(x) = \mathsf{G}_i(x) \oplus k_i$ for $1 \le i \le r$. Then by Theorem 3 and Theorem 2 (4), we have

$$\Pr[\wedge_{i=1}^{r} \boldsymbol{\alpha}_i \xrightarrow[\boldsymbol{x}_i]{\mathsf{E}} \boldsymbol{\alpha}_{i+1}] = \sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} (-1)^{u_{i+1}^{\top} k_i} D^{\mathsf{G}_i}_{(u_{i+1},\boldsymbol{\alpha}_{i+1}),(u_i,\boldsymbol{\alpha}_i)}$$

with $u_1 = u_{r+1} = 0$, $\boldsymbol{x}_i = \mathsf{E}_{i-1}(\boldsymbol{x}_{i-1})$ for $2 \le i \le r$, and $\boldsymbol{x}_1$ uniform random on $\mathbb{F}_2^n$. Assume there are in total s quasi-3-differential trails for the 3-differential characteristic $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{r+1})$. For the ℓ-th quasi-3-differential trail, denoting

the corresponding masks vector as $u^\ell = u_2^\ell || u_3^\ell || \cdots || u_r^\ell$ (note that $u_1^\ell = u_{r+1}^\ell = 0$), we use c_ℓ to represent the sign of the correlation of this quasi-3-differential trail, $1 \leq \ell \leq s$, *i.e.*,

$$c_\ell = \begin{cases} 0 & \text{if } \prod_{i=1}^{r-1} D^{\mathsf{G}_i}_{(u_{i+1}^\ell, \boldsymbol{\alpha}_{i+1}),(u_i^\ell, \boldsymbol{\alpha}_i)} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Let $k = k_1 || k_2 || \cdots || k_{r-1}$. Then, we set

$$\begin{cases} {u^1}^\top k = c_1 \\ {u^2}^\top k = c_2 \\ \quad \cdots \\ {u^s}^\top k = c_s. \end{cases}$$

The above linear equation system has at least one solution since the round keys are linear independent. In this case, the correlation of every quasi-3-differential trail is $\prod_{i=1}^{r} (-1)^{u_{i+1}^\top k_i} D^{\mathsf{G}_i}_{(u_{i+1}, \boldsymbol{\alpha}_{i+1}),(u_i, \boldsymbol{\alpha}_i)} > 0$, so $\Pr[\wedge_{i=1}^{r} \boldsymbol{\alpha}_i \xrightarrow[\boldsymbol{x}_i]{\mathsf{E}} \boldsymbol{\alpha}_{i+1}] > 0$.

7 Conclusions

In this paper, we generalized the notion of QDTMs to d-differential, and got the d-QDTMs. Consequently, we can give the exact formula of the probability of d-differential in the fixed-key model without any assumptions, even in the related-key setting. In particular, we can partially answer the open problems proposed by Tiessen and Dunkelman *et al.* at EUROCRYPT 2016 and EUROCRYPT 2020 theoretically, although the computing complexity is very high in practice. Moreover, we have revisited the boomerang attack and found that there is a gap between the real probability and the value calculated under the framework of sandwich attack. Then, we applied our theory to lightweight block cipher GIFT. Unfortunately, we cannot improve the probability of the boomerang distinguisher we considered because of the constraint of computing power. It is interesting that we found the probability of every optimal 3-differential included in the middle part E_m of the boomerang distinguisher given in [14] is zero, which can be seen as an evidence of the intuition that the security of block ciphers adopting half-round key XOR against the differential-like attacks is overestimated in previous works.

Finally, we hope that the theoretic result presented in this paper can help the community of symmetric ciphers to have a better understanding of the boomerang attack, and even can help researchers to design more efficient heuristic algorithms for estimating the probability of boomerang distinguisher guaranteed by theory.

Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments which have highly improved the manuscript. The first author

would also like to thank Fukang Liu and Qianqian Yang for helpful discussions. Libo Wang is supported by the National Natural Science Foundation of China under Grants 62102167, 62032025, and by the Guangdong Basic and Applied Basic Research Foundation under Grants 2022A1515010299, 2020A1515110364. Ling Song is supported by the National Natural Science Foundation of China under Grants 62372213, 62022036, and 62132008. Baofeng Wu is supported by the National Natural Science Foundation of China under Grant 61972393. Takanori Isobe is supported by JST, PRESTO Grant Number JPMJPR2031 and No.05801. This research was in part conducted under a contract of "Research and development on new generation cryptography for secure wireless communication services" among "Research and Development for Expansion of Radio Wave Resources (JPJ000254)", which was supported by the Ministry of Internal Affairs and Communications, Japan.

References

1. R. Avanzi. The QARMA block cipher family. almost MDS matrices over rings with zero divisors, nearly symmetric even-mansour constructions with non-involutory central rounds, and search heuristics for low-latency s-boxes. *IACR Trans. Symmetric Cryptol.*, 2017(1):4–44, 2017.
2. A. Baksi. The problem of half round key xor. *Cryptology ePrint Archive*, 2023.
3. A. Baksi, J. Breier, A. Chattopadhyay, T. Gerlich, S. Guilley, N. Gupta, K. Hu, T. Isobe, A. Jati, P. Jedlicka, et al. Baksheesh: Similar yet different from gift. *Cryptology ePrint Archive*, 2023.
4. S. Banik, S. K. Pandey, T. Peyrin, Y. Sasaki, S. M. Sim, and Y. Todo. GIFT: A small present - towards reaching the limit of lightweight encryption. In W. Fischer and N. Homma, editors, *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, volume 10529 of *Lecture Notes in Computer Science*, pages 321–345. Springer, 2017.
5. C. Beierle, G. Leander, A. Moradi, and S. Rasoolzadeh. CRAFT: lightweight tweakable block cipher with efficient protection against DFA attacks. *IACR Trans. Symmetric Cryptol.*, 2019(1):5–45, 2019.
6. T. Beyne. Block cipher invariants as eigenvectors of correlation matrices. In T. Peyrin and S. D. Galbraith, editors, *Advances in Cryptology - ASIACRYPT 2018 - 24th International Conference on the Theory and Application of Cryptology and Information Security, Brisbane, QLD, Australia, December 2-6, 2018, Proceedings, Part I*, volume 11272 of *Lecture Notes in Computer Science*, pages 3–31. Springer, 2018.
7. T. Beyne. A geometric approach to linear cryptanalysis. In M. Tibouchi and H. Wang, editors, *Advances in Cryptology - ASIACRYPT 2021 - 27th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 6-10, 2021, Proceedings, Part I*, volume 13090 of *Lecture Notes in Computer Science*, pages 36–66. Springer, 2021.
8. T. Beyne and V. Rijmen. Differential cryptanalysis in the fixed-key model. In Y. Dodis and T. Shrimpton, editors, *Advances in Cryptology - CRYPTO 2022 - 42nd Annual International Cryptology Conference, CRYPTO 2022, Santa Barbara, CA, USA, August 15-18, 2022, Proceedings, Part III*, volume 13509 of *Lecture Notes in Computer Science*, pages 687–716. Springer, 2022.

9. E. Biham, O. Dunkelman, and N. Keller. Related-key boomerang and rectangle attacks. In *Eurocrypt*, volume 3494, pages 507–525. Springer, 2005.
10. E. Biham and A. Shamir. Differential cryptanalysis of des-like cryptosystems. In A. Menezes and S. A. Vanstone, editors, *Advances in Cryptology - CRYPTO '90, 10th Annual International Cryptology Conference, Santa Barbara, California, USA, August 11-15, 1990, Proceedings*, volume 537 of *Lecture Notes in Computer Science*, pages 2–21. Springer, 1990.
11. A. Biryukov and D. Khovratovich. Related-key cryptanalysis of the full AES-192 and AES-256. In M. Matsui, editor, *Advances in Cryptology - ASIACRYPT 2009, 15th International Conference on the Theory and Application of Cryptology and Information Security, Tokyo, Japan, December 6-10, 2009. Proceedings*, volume 5912 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2009.
12. A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. B. Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: an ultra-lightweight block cipher. In P. Paillier and I. Verbauwhede, editors, *Cryptographic Hardware and Embedded Systems - CHES 2007, 9th International Workshop, Vienna, Austria, September 10-13, 2007, Proceedings*, volume 4727 of *Lecture Notes in Computer Science*, pages 450–466. Springer, 2007.
13. H. Boukerrou, P. Huynh, V. Lallemand, B. Mandal, and M. Minier. On the feistel counterpart of the boomerang connectivity table introduction and analysis of the FBCT. *IACR Trans. Symmetric Cryptol.*, 2020(1):331–362, 2020.
14. L. Chen, G. Wang, and G. Zhang. Milp-based related-key rectangle attack and its application to gift, khudra, mibs. *The Computer Journal*, 62(12):1805–1821, 2019.
15. C. Cid, T. Huang, T. Peyrin, Y. Sasaki, and L. Song. Boomerang connectivity table: A new cryptanalysis tool. In J. B. Nielsen and V. Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part II*, volume 10821 of *Lecture Notes in Computer Science*, pages 683–714. Springer, 2018.
16. J. Daemen, R. Govaerts, and J. Vandewalle. Correlation matrices. In B. Preneel, editor, *Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings*, volume 1008 of *Lecture Notes in Computer Science*, pages 275–285. Springer, 1994.
17. S. Delaune, P. Derbez, and M. Vavrille. Catching the fastest boomerangs application to SKINNY. *IACR Trans. Symmetric Cryptol.*, 2020(4):104–129, 2020.
18. O. Dunkelman, N. Keller, E. Ronen, and A. Shamir. The retracing boomerang attack. In A. Canteaut and Y. Ishai, editors, *Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part I*, volume 12105 of *Lecture Notes in Computer Science*, pages 280–309. Springer, 2020.
19. O. Dunkelman, N. Keller, and A. Shamir. A practical-time related-key attack on the KASUMI cryptosystem used in GSM and 3g telephony. In T. Rabin, editor, *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, volume 6223 of *Lecture Notes in Computer Science*, pages 393–410. Springer, 2010.
20. O. Dunkelman, N. Keller, and A. Shamir. A practical-time related-key attack on the KASUMI cryptosystem used in GSM and 3g telephony. *J. Cryptol.*, 27(4):824–849, 2014.
21. H. Hadipour, N. Bagheri, and L. Song. Improved rectangle attacks on SKINNY and CRAFT. *IACR Trans. Symmetric Cryptol.*, 2021(2):140–198, 2021.

22. M. Izadi, B. Sadeghiyan, S. S. Sadeghian, and H. A. Khanooki. MIBS: A new lightweight block cipher. In J. A. Garay, A. Miyaji, and A. Otsuka, editors, *Cryptology and Network Security, 8th International Conference, CANS 2009, Kanazawa, Japan, December 12-14, 2009. Proceedings*, volume 5888 of *Lecture Notes in Computer Science*, pages 334–348. Springer, 2009.
23. F. Ji, W. Zhang, C. Zhou, and T. Ding. Improved (related-key) differential cryptanalysis on GIFT. In O. Dunkelman, M. J. J. Jr., and C. O'Flynn, editors, *Selected Areas in Cryptography - SAC 2020 - 27th International Conference, Halifax, NS, Canada (Virtual Event), October 21-23, 2020, Revised Selected Papers*, volume 12804 of *Lecture Notes in Computer Science*, pages 198–228. Springer, 2020.
24. A. B. Kidmose and T. Tiessen. A formal analysis of boomerang probabilities. *IACR Trans. Symmetric Cryptol.*, 2022(1):88–109, 2022.
25. X. Lai, J. L. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In D. W. Davies, editor, *Advances in Cryptology - EUROCRYPT '91, Workshop on the Theory and Application of of Cryptographic Techniques, Brighton, UK, April 8-11, 1991, Proceedings*, volume 547 of *Lecture Notes in Computer Science*, pages 17–38. Springer, 1991.
26. C. Li, B. Wu, and D. Lin. Generalized boomerang connectivity table and improved cryptanalysis of GIFT. In Y. Deng and M. Yung, editors, *Information Security and Cryptology - 18th International Conference, Inscrypt 2022, Beijing, China, December 11-13, 2022, Revised Selected Papers*, volume 13837 of *Lecture Notes in Computer Science*, pages 213–233. Springer, 2022.
27. S. Murphy. The return of the cryptographic boomerang. *IEEE Trans. Inf. Theory*, 57(4):2517–2521, 2011.
28. A. Niemetz, M. Preiner, C. Wolf, and A. Biere. Btor2, btormc and boolector 3.0. In *Computer Aided Verification: 30th International Conference, CAV 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 14-17, 2018, Proceedings, Part I*, pages 587–595. Springer, 2018.
29. L. Song, X. Qin, and L. Hu. Boomerang connectivity table revisited. application to SKINNY and AES. *IACR Trans. Symmetric Cryptol.*, 2019(1):118–141, 2019.
30. T. Suzaki, K. Minematsu, S. Morioka, and E. Kobayashi. TWINE: A lightweight block cipher for multiple platforms. In L. R. Knudsen and H. Wu, editors, *Selected Areas in Cryptography, 19th International Conference, SAC 2012, Windsor, ON, Canada, August 15-16, 2012, Revised Selected Papers*, volume 7707 of *Lecture Notes in Computer Science*, pages 339–354. Springer, 2012.
31. T. Tiessen. Polytopic cryptanalysis. In M. Fischlin and J. Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, volume 9665 of *Lecture Notes in Computer Science*, pages 214–239. Springer, 2016.
32. D. A. Wagner. The boomerang attack. In L. R. Knudsen, editor, *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999, Proceedings*, volume 1636 of *Lecture Notes in Computer Science*, pages 156–170. Springer, 1999.
33. H. Wang and T. Peyrin. Boomerang switch in multiple rounds. application to AES variants and deoxys. *IACR Trans. Symmetric Cryptol.*, 2019(1):142–169, 2019.
34. W. Wu and L. Zhang. Lblock: A lightweight block cipher. In J. López and G. Tsudik, editors, *Applied Cryptography and Network Security - 9th International Conference, ACNS 2011, Nerja, Spain, June 7-10, 2011. Proceedings*, volume 6715 of *Lecture Notes in Computer Science*, pages 327–344, 2011.

35. Q. Yang, L. Song, S. Sun, D. Shi, and L. Hu. New properties of the double boomerang connectivity table. *IACR Trans. Symmetric Cryptol.*, 2022(4):208–242, 2022.