Differential Cryptanalysis in the Fixed-Key Model

Tim Beyne[1] and Vincent Rijmen[1,2]

[1] imec-COSIC, KU Leuven, Belgium
name.lastname@esat.kuleuven.be
[2] University of Bergen, Bergen, Norway

Abstract. A systematic approach to the fixed-key analysis of differential probabilities is proposed. It is based on the propagation of 'quasidifferential trails', which keep track of probabilistic linear relations on the values satisfying a differential characteristic in a theoretically sound way. It is shown that the fixed-key probability of a differential can be expressed as the sum of the correlations of its quasidifferential trails.
The theoretical foundations of the method are based on an extension of the difference-distribution table, which we call the quasidifferential transition matrix. The role of these matrices is analogous to that of correlation matrices in linear cryptanalysis. This puts the theory of differential and linear cryptanalysis on an equal footing.
The practical applicability of the proposed methodology is demonstrated by analyzing several differentials for RECTANGLE, KNOT, Speck and Simon. The analysis is automated and applicable to other SPN and ARX designs. Several attacks are shown to be invalid, most others turn out to work only for some keys but can be improved for weak-keys.

Keywords: Differential cryptanalysis · Hypothesis of stochastic equivalence · Correlation matrices · RECTANGLE · KNOT · Speck · Simon

1 Introduction

At CRYPTO 1990, Biham and Shamir [5] published the first reduced-round differential attacks on the block cipher DES. Differential cryptanalysis is now one of the cornerstones of the security analysis of block ciphers and hash functions. Its central problem is to count the number of inputs of a function for which a given input difference results in a particular output difference or, what amounts to the same, to compute the probability of a differential.

For functions that can be written as a composition of simple operations, the standard procedure is to analyze sequences of intermediate differences or *characteristics*. The probability of a characteristic is then heuristically estimated by multiplying the probabilities of the intermediate differentials. In the context of block ciphers, Lai, Massey and Murphy [16] showed that this procedure yields the correct value of the *key-averaged probability* for Markov ciphers.

However, since the key is fixed throughout a differential attack, even the average data-complexity cannot be computed from the average probability of differentials alone. Hence, Lai *et al.* [16] introduced an additional assumption

known as the *hypothesis of stochastic equivalence*. It states that the probability for each key is close to the average probability.

In practice, it turns out that the probability can vary significantly between keys. Hence, standard assumptions may lead to incorrect conclusions. Furthermore, averages may hide weak-key attacks that can considerably degrade security. Finally, the same formalism is used even when there is no key, such as for cryptographic permutations, or when the cryptanalyst has full control over the key, such as in many hash functions.

From a theoretical viewpoint, it can be argued that the standard approach to differential cryptanalysis is incomplete, since it does not offer any tools to compute probabilities beyond the average case. This is in contrast to linear cryptanalysis [20], where key-dependence is much better understood. In particular, the correlation matrix approach of Daemen *et al.* [10] shows that the correlation of a linear approximation is precisely equal to the sum of the correlations of all its linear trails.

Previous work. Knudsen [15] already observed significant deviations from the hypothesis of stochastic equivalence for the characteristics used in the differential analysis of DES. Experiments such as those of Ankele and Kölbl [2] and Heys [14] further suggest that such deviations are the norm rather than the exception.

Daemen and Rijmen [11] showed that the fixed-key probability of two-round characteristics of AES is either zero or 2^h, with h an integer independent of the key. Such characteristics are called *plateau characteristics*, and have been used in several other contexts [9,19,21,25]. Although plateau characteristics are the only systematic method to analyze fixed-key probabilities for S-box-based ciphers, their scope remains limited. They assume that the input or output values satisfying a differential over the S-box form an affine space. In addition, their analysis becomes difficult for more than two rounds.

For constructions relying on modular additions, several techniques were developed in the context of collision attacks on hash functions. These methods keep track of additional information about the values satisfying a characteristic. For example, the breakthrough results of Wang *et al.* [26] rely on *signed differences*. De Cannière and Rechberger [12] extended these to *generalized differences*, allowing arbitrary constraints to be imposed on individual bits. Leurent [18] proposed a framework for ARX-constructions based on two-bit conditions. Xu *et al.* [27] recently introduced *signed sums*, which are single-bit conditions. Despite their merit, these techniques have significant limitations. Imposing conditions directly on values becomes difficult for keyed functions, since key-additions result in conditions that potentially depend on many unknown bits. Hence, these methods are limited to keyless functions except for local, key-independent effects in ciphers such as XTEA that use modular additions between dependent values. Furthermore, the conditions that are imposed cannot fully explain the probability of a characteristic, and the right choice of the type of conditions to use depends on the function under analysis.

Contribution. We develop a general methodology to analyze the fixed-key probabilities of differentials. It allows propagating probabilistic linear relations on the values satisfying differential characteristics in a theoretically sound way. The theoretical foundations of the proposed approach are inspired by the correlation matrix framework [10] and its recent generalization [4] that provide a natural description of linear cryptanalysis.

Section 3 builds up an extension of the difference-distribution table that we call the *quasidifferential transition matrix*. It is obtained by performing a change-of-basis on the permutation matrices describing the propagation of probability distributions of pairs through a function, analogous to the construction of correlation matrices using the Fourier transformation. Our choice of basis ensures that the difference distribution table is obtained as a submatrix, and simultaneously diagonalizes the transition matrices corresponding to round-key additions.

By construction, quasidifferential transition matrices satisfy similar properties as correlation matrices. For example, composition of functions corresponds to multiplication of quasidifferential transition matrices. This property leads to quasidifferential trails, the central notion of our methodology. In Section 4, we prove that the sum of the correlations of all quasidifferential trails in a characteristic is equal to its exact probability. Likewise, the probability of a differential is the sum of the correlations of all quasidifferential trails. A few quasidifferential trails often capture the essence of the key-dependence. For example, the key-dependence in the DES characteristics observed by Knudsen [15] is explained by taking into account one additional one-round quasidifferential trail.

To demonstrate the practical applicability of our methodology, we apply it to four primitives. To this end, an algorithm to compute the quasidifferential transition matrix of general functions in time proportional (up to logarithmic factors) to the size of the matrix is given in Section 5. In addition, the quasidifferential transition matrix of bitwise-and and modular addition are determined.

Section 6 presents an automated search tool for quasidifferential trails in RECTANGLE [28]. The implementation is provided as supplementary material[3], and can also be used for the analysis of other, similar ciphers. Our analysis shows that the best published key-recovery attack on round-reduced RECTANGLE does not work, but we show how to modify it to obtain a valid weak-key attack.

In Section 7 we apply the same tool to KNOT [29], a second-round candidate in the NIST lightweight cryptography competition. We show that previously proposed reduced-round forgery and collision attacks do not work, because the characteristics they rely on have probability zero. At the same time, we show that their probabilities are two orders of magnitude larger for some choices of the round constants.

Section 8 reevaluates the best published attacks on Speck. The analysis relies on an automated search tool that is provided as supplementary material. It can easily be modified for other ARX designs. We find that most of the attacks we analyzed only work for a subset of keys. However, we also show that for

[3] All of our source code can be found at https://github.com/TimBeyne/quasidifferential-trails.

weak keys, attacks with lower data-complexity can be obtained. In the extended version of the paper, we provide a similar search tool for Simon.

2 Preliminaries and Related Work

Most of the notations used in this paper are standard, or will be introduced where necessary. Throughout this paper, random variables are denoted in boldface. The average of a random variable $\boldsymbol{x}$ will be denoted by $\mathsf{E}\,\boldsymbol{x}$, and its variance by $\mathsf{Var}\,\boldsymbol{x}$. All key-dependent probabilities given in this paper are with respect to a fixed key, unless it is explicity mentioned that they are averages.

2.1 Differential Cryptanalysis

Differential cryptanalysis [5,6] is a technique to analyze the propagation of differences through a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$. Typically, the cryptanalyst attempts to find a differential $(a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$ such that the difference equation

$$\mathsf{F}(x) + \mathsf{F}(x + a) = b\,, \tag{1}$$

has a large number of solutions in x. The ordered pairs $(x, x+a)$ for which Equation (1) holds are called *right pairs* for the differential (a, b). The number of right pairs divided by 2^n is called the probability of the differential. The *difference-distribution table* $\mathsf{DDT}^{\mathsf{F}}$ 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 right pairs for a particular differential:

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

with $\boldsymbol{x}$ uniform random on $\mathbb{F}_2^n$. A differential with probability $p \gg 2^{-n}$ results in a distinguisher with data-complexity $\mathcal{O}(1/p)$.

Characteristics. Computing or estimating the probability of a differential for a general function with many inputs can be computationally difficult. However, differential cryptanalysis is typically applied to functions F of the form $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$, where the functions F_i admit differentials with relatively high probability 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 simplicity of notation, assume that $m = n$ and the functions F_i are all n-bit functions. It holds that

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

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

$$\Pr\left[\textstyle\bigwedge_{i=1}^{r} \mathsf{F}_i(\boldsymbol{x}_i) + \mathsf{F}_i(\boldsymbol{x}_i + a_i) = a_{i+1}\right] = \prod_{i=1}^{r} \Pr\left[\mathsf{F}_i(\boldsymbol{z}_i) + \mathsf{F}_i(\boldsymbol{z}_i + a_i) = a_{i+1}\right].$$

Under the same independence heuristic, combining the equations above yields

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

We stress that Equation (2) is an approximation, and it is easy to come up with examples such as $\mathsf{F} = \mathsf{F}_2 \circ \mathsf{F}_1$ with $\mathsf{F}_2 = \mathsf{F}_1^{-1}$ where it fails spectacularly.

Key-averaged probabilities. If the functions $\mathsf{F}_1, \ldots, \mathsf{F}_r$ depend on keys $k_1, \ldots, k_r$, then the heuristic Equation (2) can be motivated using the *Markov cipher* assumption [16]. In particular, it can be shown that if all round keys are uniform random and independent, then the key-averaged probability of a characteristic is indeed equal to the product of the intermediate key-averaged probabilities.

Aside from the fact that most ciphers are not true Markov ciphers due to round-key dependencies introduced by the key-schedule, one is ultimately interested in fixed-key rather than key-averaged probabilities. Importantly, this is true even when computing the key-averaged data-complexity of an attack. After all, in general $\mathsf{E}\,[1/p_{\boldsymbol{k}}] \neq 1/\,\mathsf{E}\,[p_{\boldsymbol{k}}]$ with $p_{\boldsymbol{k}}$ the probability for a random key $\boldsymbol{k}$.

Hence, to bridge this gap, an additional hypothesis was introduced by Lai, Massey and Murphy [16, §2]. Informally, the *hypothesis of stochastic equivalence* states that the key-averaged probability of a characteristic is close to its fixed-key probability for any particular key. As discussed in the introduction, previous work has shown that this assumption is often unrealistic.

2.2 Linear Cryptanalysis

Although the average probability of characteristics and differentials is relatively well understood, few techniques are known to analyze fixed-key probabilities. This contrasts with linear cryptanalysis, where linear trails give a complete description of the correlation of linear approximations even in the fixed-key setting.

A natural way to describe linear cryptanalysis is by means of correlation matrices. These matrices were first introduced by Daemen *et al.* [10]. Although the scope of the present paper is limited to differential cryptanalysis only, it is useful to introduce these matrices as they provide an important motivation for the quasidifferential transition matrices that will be introduced in Section 3.

From the viewpoint introduced in [3, 4], correlation matrices represent linear operators that act on functions $\mathbb{F}_2^n \to \mathbb{R}$ such as probability distributions. In the following, let $\mathbb{R}[\mathbb{F}_2^n]$ denote the vector space of such functions. The functions δ_x such that $\delta_x(y) = 1$ if $y = x$ and zero elsewhere form an orthonormal basis for $\mathbb{R}[\mathbb{F}_2^n]$ with respect to the inner product $\langle f, g\rangle = \sum_{x\in\mathbb{F}_2^n} f(x)\, g(x)$. Below, this basis will be referred to as the *standard basis*.

Another convenient basis for $\mathbb{R}[\mathbb{F}_2^n]$ consists of the group characters of $\mathbb{F}_2^n$. These are homomorphisms from $\mathbb{F}_2^n$ to the multiplicative group $\mathbb{C}\setminus\{0\}$. Any such homomorphism is of the form $\chi_u(x) = (-1)^{u^\mathsf{T} x}$ with $u \in \mathbb{F}_2^n$. The characters χ_u form an orthogonal basis for $\mathbb{R}[\mathbb{F}_2^n]$. Specifically, $\langle \chi_u, \chi_v\rangle = 2^n\, \delta_u(v)$. Hence, any function $f \in \mathbb{R}[\mathbb{F}_2^n]$ can be expressed as a linear combination of the characters

χ_u. This leads to the Fourier transformation, which is defined in Definition 2.1. The basis $\{\chi_u \mid u \in \mathbb{F}_2^n\}$ will be called the *character basis*.

Definition 2.1 (Fourier transformation). *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be a function. The Fourier transformation of f is the function $\mathscr{F}_n f : \mathbb{F}_2^n \to \mathbb{R}$ defined by $(\mathscr{F}_n f)(u) = \langle \chi_u, f \rangle$. That is, $(\mathscr{F}_n f)(u)/2^n$ is the coordinate corresponding to the basis function χ_u when f is expressed in the character basis.*

The motivation for using the character basis is that it simplifies the effect of translating functions by a constant. In particular, if $g(x) = f(x+t)$, then $(\mathscr{F}_n g)(u) = \chi_u(t)\,(\mathscr{F}_n f)(u)$ because $\chi_u(x+t) = \chi_u(t)\chi_u(x)$ by the definition of characters as group homomorphisms.

Correlation matrices describe how a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ transforms functions in $\mathbb{R}[\mathbb{F}_2^n]$ to functions in $\mathbb{R}[\mathbb{F}_2^m]$. In the standard basis, the relation is expressed by a permutation matrix that is called the *transition matrix* in Definition 2.2. The same linear transformation can be expressed in the Fourier basis and this yields Definition 2.3.

Definition 2.2 (Transition matrix [4, Definition 3.2]). *Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a function. Define $T^{\mathsf{F}} : \mathbb{R}[\mathbb{F}_2^n] \to \mathbb{R}[\mathbb{F}_2^m]$ as the unique linear operator defined by $\delta_x \mapsto \delta_{\mathsf{F}(x)}$ 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 $\mathbb{R}[\mathbb{F}_2^m]$.*

Definition 2.3 (Correlation matrix [4, Definition 3.3]). *Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a function. 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, $C^{\mathsf{F}} = \mathscr{F}_m\, T^{\mathsf{F}}\, \mathscr{F}_n^{-1}$. The correlation matrix of F is the coordinate representation of C^{F} with respect to the standard bases of $\mathbb{R}[\mathbb{F}_2^n]$ and $\mathbb{R}[\mathbb{F}_2^m]$.*

The coordinates of the correlation matrix C^{F} correspond to the correlations of linear approximations over F. In particular, $C^{\mathsf{F}}_{v,u} = 2\Pr[v^{\mathsf{T}}\mathsf{F}(\boldsymbol{x}) + u^{\mathsf{T}}\boldsymbol{x} = 0] - 1$ with $\boldsymbol{x}$ uniform random. In fact, the original definition of correlation matrices due to Daemen *et al.* [10] starts from this equivalence.

Correlation matrices satisfy several natural properties, most of which are direct consequences of the properties of transition matrices and Definition 2.3. In particular, for a function $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$, 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:

$$C^{\mathsf{F}}_{u_{r+1},u_1} = \sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} C^{\mathsf{F}_i}_{u_{i+1},u_i}\,. \tag{3}$$

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 $u_2, \ldots, u_r$. This result should be compared with Equation (2) for differentials. However, there is a fundamental difference: whereas Equation (2) is heuristic and at best true on

average with respect to independent uniform random round keys, Equation (3) holds exactly without any assumptions.

As argued in the introduction, closing the gap between Equation (2) and Equation (3) is essential to achieve a more complete understanding of differential cryptanalysis. To this end, Section 3 introduces quasidifferential transition matrices as a differential analog of correlation matrices.

3 Quasidifferential Transition Matrices

The probability of differentials can be described exactly by tracking the distribution of pairs of state values. Such a distribution can be described by a function $p : \mathbb{F}_2^n \times \mathbb{F}_2^n \to [0,1] \subseteq \mathbb{R}$. There exists a transition matrix which describes the propagation of such probability distributions through a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$.

However, keeping track of pairs directly is inconvenient because it does not provide a simple description of translations – which are essential to understand key-dependence. In Section 3.1, we define a new basis that is nicer to work with. In Section 3.2, it is then shown that expressing transition matrices in this new basis leads to matrices with similar properties as correlation matrices. These *quasidifferential transition matrices* will be used in Section 4 to give a natural fixed-key description of differential cryptanalysis.

3.1 Quasidifferential Basis

As discussed in Section 2, the Fourier transformation simplifies the effect of translations on functions. However, the character basis is not suitable to describe differences between the halves of pairs in a straightforward way. The basis proposed in Definition 3.1 below is a hybrid solution. Up to scaling, it contains the probability distributions of uniform random pairs with a fixed difference and, as shown below, it simplifies the effect of translations.

Definition 3.1 (Quasidifferential basis). *Let n be a positive integer. For any $u, a \in \mathbb{F}_2^n$, the function $\beta_{u,a} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{R}$ is defined by*

$$\beta_{u,a}(x,y) = \chi_u(x)\, \delta_a(x+y)\,.$$

The set of all $\beta_{u,a}$ will be called the quasidifferential basis for $\mathbb{R}[\mathbb{F}_2^n \times \mathbb{F}_2^n]$.

The functions $\beta_{u,a}$ are not only linearly independent, but also orthogonal. This is shown in Theorem 3.1, which also states the important translation-invariance property.

Theorem 3.1. *The quasidifferential basis defined in Definition 3.1 is translation-invariant and orthogonal. Specifically:*

(1) For all $(u,a),(v,b) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$, it holds that $\langle \beta_{v,b}, \beta_{u,a} \rangle = 2^n\, \delta_v(u)\, \delta_b(a)$.
(2) For all $(u,a) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ and $t \in \mathbb{F}_2^n$, it holds that

$$\beta_{u,a}(x+t, y+t) = \chi_u(t)\, \beta_{u,a}(x,y)\,.$$

Proof. The first property follows from the expression

$$\langle \beta_{v,b},\, \beta_{u,a} \rangle = \sum_{(x,y)\in \mathbb{F}_2^n \times \mathbb{F}_2^n} \chi_v(x)\, \delta_b(x+y)\, \chi_u(x)\, \delta_a(x+y)\,.$$

Indeed, if $a \neq b$, then $x+y=a$ and $x+y=b$ never hold simultaneously. If $a=b$, then the result follows from the orthogonality of the characters χ_u. The translation-invariance follows from the fact that $\chi_u(x+t) = \chi_u(t)\,\chi_u(x)$. □

Similar to the Fourier transformation, we define the change-of-basis operator $\mathscr{Q}_n : \mathbb{R}[\mathbb{F}_2^n \times \mathbb{F}_2^n] \to \mathbb{R}[\mathbb{F}_2^n \times \mathbb{F}_2^n]$ by $(\mathscr{Q}_n f)(u,a) = \langle \beta_{u,a},\, f\rangle$. By Theorem 3.1 (1), $(\mathscr{Q}_n f)(u,a)/2^n$ is then indeed the coordinate corresponding to basis function $\beta_{u,a}$ when f is expressed in the quasidifferential basis.

3.2 Quasidifferential Transition Matrix

Recall from Section 2.2 that the correlation matrix of a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ with transition matrix T^{F} is defined as $C^{\mathsf{F}} = \mathscr{F}_m\, T^{\mathsf{F}}\, \mathscr{F}_n^{-1}$. Below, we define the *quasidifferential transition matrix* similarly using the change-of-basis operator $\mathscr{Q}_n$ and the transition matrix for pairs of values. The latter matrix can be succinctly written as the Kronecker (or tensor) product $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$, which is defined as a $2^{2m} \times 2^{2n}$ matrix with coordinates

$$(T^{\mathsf{F}} \otimes T^{\mathsf{F}})_{(y_1,y_2),(x_1,x_2)} = T^{\mathsf{F}}_{y_1,x_1} T^{\mathsf{F}}_{y_2,x_2} = \delta_{y_1}(\mathsf{F}(x_1))\delta_{y_2}(\mathsf{F}(x_2))\,.$$

Note that we index the coordinates of $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$ directly by pairs of bitvectors. This avoids choosing an arbitrary convention for converting between integers and bitvector pairs.

Definition 3.2 (Quasidifferential transition matrix). *Let n and m be positive integers and $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ a function. The quasidifferential transition matrix D^{F} is defined as the matrix-representation of $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$ with respect to the quasidifferential basis defined in Definition 3.1. That is, $D^{\mathsf{F}} = \mathscr{Q}_m (T^{\mathsf{F}} \otimes T^{\mathsf{F}}) \mathscr{Q}_n^{-1}$.*

To make Definition 3.2 more concrete, we compute the coordinates of D^{F}. Like for $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$, the coordinates of D^{F} will be indexed by pairs $(u,a) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ and $(v,b) \in \mathbb{F}_2^m \times \mathbb{F}_2^m$. By the orthogonality of the quasidifferential basis (Theorem 3.1 (1)), it holds that $\mathscr{Q}_n^{-1} = \mathscr{Q}_n^{\mathsf{T}}/2^n$ and consequently

$$D^{\mathsf{F}}_{(v,b),\,(u,a)} = \langle \delta_{(v,b)}\,, \mathscr{Q}_n (T^{\mathsf{F}} \otimes T^{\mathsf{F}}) \mathscr{Q}_n^{\mathsf{T}}\, \delta_{(u,a)} \rangle / 2^n = \langle \beta_{v,b},\, (T^{\mathsf{F}} \otimes T^{\mathsf{F}}) \beta_{u,a} \rangle / 2^n\,.$$

Working this out yields the following expression:

$$\begin{aligned} D^{\mathsf{F}}_{(v,b),\,(u,a)} &= \frac{1}{2^n} \sum_{(x,y)\in \mathbb{F}_2^n \times \mathbb{F}_2^n} \chi_u(x)\chi_v(\mathsf{F}(x))\, \delta_a(x+y)\delta_b(\mathsf{F}(x)+\mathsf{F}(y)) \\ &= \frac{1}{2^n} \sum_{\substack{x\in \mathbb{F}_2^n \\ \mathsf{F}(x+a)=\mathsf{F}(x)+b}} (-1)^{u^{\mathsf{T}}x + v^{\mathsf{T}}\mathsf{F}(x)}\,. \end{aligned} \tag{4}$$

For $u = v = 0$, Equation (4) reduces to the probability of the differential with input difference a and output difference b. That is, $D^{\mathsf{F}}_{(0,b),(0,a)} = 2^{-n}\,\mathsf{DDT}^{\mathsf{F}}_{a,b}$. For $a = b = 0$, one obtains the coordinates of the correlation matrix of F. Specifically, $D^{\mathsf{F}}_{(v,0),(u,0)} = C^{\mathsf{F}}_{v,u}$. More generally, the right hand side of Equation (4) can be interpreted as a kind of correlation matrix for the function F when restricted to the right pair set of the differential (a, b). That is, the coordinates of D^{F} express the correlations of probabilistic linear relations ('linear approximations') between the input and output values of the right pairs.

The following result summarizes some of the basic properties of quasidifferential transition matrices. Properties (1) to (3) are identical to those of correlation matrices [4, Theorem 3.1], and their proofs are nearly identical. For Theorem 3.2 (2), the Kronecker product of two quasidifferential transition matrices is defined by

$$(D^{\mathsf{F}_1} \otimes D^{\mathsf{F}_2})_{(v_1\|v_2,b_1\|b_2),(u_1\|u_2,a_1\|a_2)} = D^{\mathsf{F}_1}_{(v_1,b_1),(u_1,a_1)} D^{\mathsf{F}_2}_{(v_2,b_2),(u_2,a_2)}\,,$$

with $x\|y$ the concatenation of bitvectors x and y.

Theorem 3.2. *Let n and m be positive integers and $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ a function. The matrix D^{F} has the following properties:*

(1) If F is a bijection, then D^{F} is an orthogonal matrix.
(2) If $\mathsf{F} = (\mathsf{F}_1, \ldots, \mathsf{F}_m)$, then $D^{\mathsf{F}} = \bigotimes_{i=1}^{m} D^{\mathsf{F}_i}$. *(boxed maps)*
(3) If $\mathsf{F} = \mathsf{F}_2 \circ \mathsf{F}_1$, then $D^{\mathsf{F}} = D^{\mathsf{F}_2} D^{\mathsf{F}_1}$. *(composition)*
(4) If $\mathsf{F}(x) = x + t$ for some $t \in \mathbb{F}_2^m$, then $D^{\mathsf{F}}_{(v,b),(u,a)} = \chi_v(t)\,\delta_v(u)\,\delta_b(a)$.
(5) If F is a linear function, then $D^{\mathsf{F}}_{(v,b),(u,a)} = \delta_u(\mathsf{F}^{\mathsf{T}}(v))\,\delta_b(\mathsf{F}(a))$.

Proof. Property (1) follows from the fact that $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$ is a permutation matrix when F is a bijection and the fact that $\mathscr{Q}_n/\sqrt{2^n}$ is an orthogonal matrix by Theorem 3.1 (1). Property (2) follows immediately from the analogous result for $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$ and the separability of the basis. Property (3) also follows from the same property for $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$. The fourth property is due to the translation invariance and orthogonality of the quasidifferential basis (Theorem 3.1). Finally, Property (5) can be deduced from Equation (4):

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

where the second equality follows from the orthogonality of characters and the fact that $\mathsf{F}(x + a) = \mathsf{F}(x) + b$ if and only if $b = \mathsf{F}(a)$. □

Consider the S-box $\mathsf{S} : \mathbb{F}_2^4 \to \mathbb{F}_2^4$ of the lightweight block cipher RECTANGLE, given by $\mathsf{S} = (3\ a)(0\ 6\ 7\ 9)(1\ 5\ e\ 4)(2\ c\ 8\ b\ d\ f)$ in cycle notation. The 256×256 quasidifferential transition matrix of S is shown in Figure 1, with colors representing the absolute value of the entries. The integer indices correspond to pairs (u, a) by the map $(u, a) \mapsto \mathsf{int}(u) + 16 \times \mathsf{int}(a)$, where $\mathsf{int}(u) = \sum_{i=1}^{4} u_i 2^{4-i}$.

Figure 1 immediately reveals a number of properties of quasidifferential transition matrices. The top-left square in Figure 1 corresponds to the correlation matrix of S. Each block shows the correlations of probabilistic linear relations between the input and output values for the right pairs. Hence, Figure 1 looks like a 'magnified' version of the difference-distribution table of S.

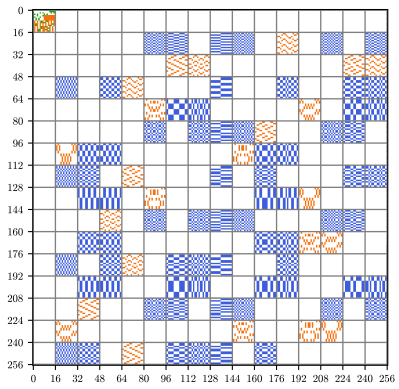

Fig. 1: The quasidifferential transition matrix D^{S} of the RECTANGLE S-box S. Blue cells correspond to values of absolute value 1/8, orange cells to 1/4, and green cells to 1/2. Empty cells correspond to zeros.

4 Quasidifferential Trails

Motivated by the notion of *linear trails* and Equation (3) from Section 2.2, the following definition defines *quasidifferential trails*. In Section 4.1, it will be shown that exact expressions for the probabilities of differentials can be given in terms of the correlations of quasidifferential trails.

Definition 4.1. *A quasidifferential trail for a function* $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ *is a sequence* $\varpi_1, \varpi_2, \ldots, \varpi_{r+1}$ *of mask-difference pairs* $\varpi_i = (u_i, a_i)$*. The correlation of this quasidifferential trail is defined as* $\prod_{i=1}^{r} D^{\mathsf{F}_i}_{\varpi_{i+1},\varpi_i}$.

Quasidifferential trails with $u_1 = u_2 = \ldots = u_{r+1} = 0$ correspond to characteristics. Their correlation is equal to the product of the one-round probabilities of the characteristic with differences $a_1, \ldots, a_{r+1}$:

$$\prod_{i=1}^{r} D^{\mathsf{F}}_{(0,a_{i+1}),(0,a_i)} = \prod_{i=1}^{r} \Pr\left[\mathsf{F}_i(\boldsymbol{x} + a_i) = \mathsf{F}_i(\boldsymbol{x}) + a_{i+1}\right],$$

with $\boldsymbol{x}$ uniform random on $\mathbb{F}_2^n$. This follows from Equation (4) and Definition 4.1.

4.1 Exact Probabilities from Quasidifferential Trails

Theorem 3.2 (2) implies that 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, is equal to the exact probability of the differential with input difference a_1 and output difference a_{r+1}. Specifically, expanding the coordinate of $D^{\mathsf{F}} = \prod_{i=1}^{r} D^{\mathsf{F}_i}$ corresponding to this differential yields

$$D^{\mathsf{F}}_{\varpi_{r+1},\varpi_1} = \sum_{\varpi_2,\ldots,\varpi_r} \prod_{i=1}^{r} D^{\mathsf{F}_i}_{\varpi_{i+1},\varpi_i} \,. \tag{5}$$

This expression also holds when the input or output mask is nonzero. Furthermore, as shown in Theorem 4.1, quasidifferential trails also allow computing the probability of a characteristic. This result should be compared with Equation (3).

Theorem 4.1. *Let $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a function such that $\mathsf{F} = \mathsf{F}_r \circ \ldots \circ \mathsf{F}_1$. The probability of a characteristic with differences $a_1, \ldots, a_{r+1}$ is equal to the sum of the correlations of all quasidifferential trails with the same intermediate differences:*

$$\Pr\left[\bigwedge_{i=1}^{r} \mathsf{F}_i(\boldsymbol{x}_i + a_i) = \mathsf{F}_i(\boldsymbol{x}_i) + a_{i+1}\right] = \sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} D^{\mathsf{F}_i}_{(u_{i+1},a_{i+1}),(u_i,a_i)} \,,$$

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

Proof. Substituting Equation (4) in the right-hand side above yields

$$\prod_{i=1}^{r} D^{\mathsf{F}_i}_{(u_{i+1},a_{i+1}),(u_i,a_i)} = \frac{1}{2^{nr}} \sum_{\substack{x_1,\ldots,x_r \\ \mathsf{F}(x_i+a_i)=\mathsf{F}(x_i)+a_{i+1}}} \prod_{i=1}^{r} (-1)^{u_i^{\mathsf{T}} x_i + u_{i+1}^{\mathsf{T}} \mathsf{F}_i(x_i)} \,.$$

Summing over $u_2, \ldots, u_r$ then results in the equation

$$\begin{aligned}
\sum_{u_2,\ldots,u_r} \prod_{i=1}^{r} D^{\mathsf{F}_i}_{(u_{i+1},a_{i+1}),(u_i,a_i)} &= \frac{1}{2^{nr}} \sum_{\substack{x_1,\ldots,x_r \\ \mathsf{F}(x_i+a_i)=\mathsf{F}(x_i)+a_{i+1}}} \prod_{i=1}^{r} \sum_{u_i} (-1)^{u_i^{\mathsf{T}}(x_{i+1}+\mathsf{F}_i(x_i))} \\
&= \frac{1}{2^n} \sum_{\substack{x_1,\ldots,x_r \\ \mathsf{F}(x_i+a_i)=\mathsf{F}(x_i)+a_{i+1}}} \prod_{i=1}^{r} \delta_{x_{i+1}}(\mathsf{F}_i(x_i)) \,.
\end{aligned}$$

Writing the right-hand side in terms of probabilities gives desired the result. □

Theorem 4.1 can also be obtained using the following intuitive argument, illustrated in Figure 2. Let $\mathsf{G} = (\mathsf{F}_1, \mathsf{F}_2 \circ \mathsf{F}_1, \ldots, \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1)$. A differential for G with input difference a_1 and output difference $(a_2, \ldots, a_{r+1})$ is equivalent to a characteristic for $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ with intermediate differences $a_2, \ldots, a_r$. For the linear function $\mathsf{L}(x) = x \| x$, Theorem 3.2 (5) yields $D^{\mathsf{L}}_{(v,b),(u,a)} = \delta_u(v_1 + v_2)\, \delta_{b_1}(a)\, \delta_{b_2}(a)$ with $v = v_1 \| v_2$ and $b = b_1 \| b_2$. Hence, all trails through G are of the form shown in Figure 2 and the result follows from Equation (5).

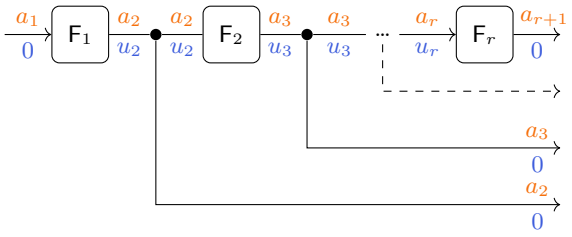

Fig. 2: Quasidifferential trail through the function G. Differences are indicated in orange (above), masks in blue (below).

4.2 Example: Differential Cryptanalysis of DES

As a first example of quasidifferential trails and Theorem 4.1, we consider the effect of key-dependence on the differential cryptanalysis of DES by Biham and Shamir [5, 6]. The example in this section is particularly simple, but more advanced applications will be discussed in Sections 6 and 8.

Recall that the differential cryptanalysis of DES is based on an iterative characteristic of the form shown in Figure 3. There exist two differences that achieve the same maximal average probability of approximately $2^{-7.87}$. For simplicity (the other case is similar), we will consider the difference $a = \texttt{0x19600000}$. The key-dependence of this characteristic was already noted by Knudsen [15, §5], who explained it using an argument specific to DES. Below, it will be shown that the general methodology of quasidifferential trails automatically provides a simple explanation.

The round function F_k of DES consists of a linear expansion function $\mathsf{E} : \mathbb{F}_2^{32} \to \mathbb{F}_2^{48}$, which duplicates certain bits, followed by the key addition and a nonlinear layer S consisting of eight 6-bit to 4-bit S-boxes. Finally, the S-box layer is followed by a bit-permutation P. The key-averaged probability of the characteristic in Figure 3 is easily computed from the difference-distribution tables of the first three S-boxes: $14/64 \times 8/64 \times 10/64 = 1120/64^3$.

However, the structure of the round function of DES leads to one-round quasidifferential trails, as shown on the right side of Figure 3. In particular, since E is not surjective, there exist masks $u \neq 0$ such that $\mathsf{E}^\mathsf{T}(u) = 0$. For the difference a mentioned above, there exists one such quasidifferential trail with $u = \texttt{0x001400000000}$. The correlation of this trail can be computed from the quasidifferential transition matrix for the first three S-boxes and equals $\chi_u(k_2)\, 14/64 \times -8/64 \times 6/64 = -\chi_u(k_2)\, 672/64^3$. It follows that a full description of the probability of the characteristic over $2r$ rounds is given by

$$p_k = \prod_{i=1}^{r} \left(\frac{1120}{64^3} - (-1)^{k_{2i,12}+k_{2i,14}} \frac{672}{64^3} \right).$$

Although for every two rounds only two trails are especially important, these trails can be combined in many ways. In particular, the expression above is

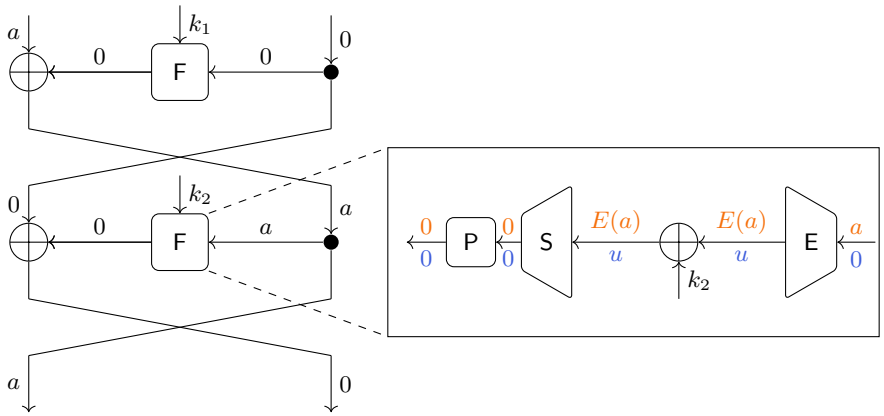

Fig. 3: Iterative characteristic for two rounds of DES.

equivalent to a sum over 2^r quasidifferential trails. This is a typical way in which a relatively small local effect can result in significant variations in the overall probability of a characteristic.

Due to the above, the probability of the thirteen round differential used in the differential attack of Biham and Shamir [6] is roughly 17 times larger for one in 64 keys and more than 244 times smaller than the average probability for an equal number of keys, as previously observed by Knudsen [15].

It is natural to wonder if there exist other quasidifferential trails with large absolute correlation. For example, a more general three-round effect can occur when $\mathsf{E}^\mathsf{T}(u) \neq 0$. However, most quasidifferential trails activating four or less additional S-boxes have correlation zero because the correlation of a linear approximation with input mask $\mathtt{1}$ or $\mathtt{32}$ and output mask $\mathtt{1}, \mathtt{2}, \mathtt{4}$ or $\mathtt{8}$ is zero for all S-boxes. This follows from the fact that the S-boxes are permutations when the first and last input bits are fixed. It can be checked that the best three-round quasidifferential trail of this type has absolute correlation at most $2^{-19.41}$.

4.3 Interpretation of Quasidifferential Trails

As discussed in Section 3.2, the coordinates of D^F can be interpreted as the correlations of linear approximations between the input and output values for the right pairs of a differential. Quasidifferential trails provide a way to connect such approximations through a sequence of functions.

Since $|D^\mathsf{F}_{(v,b),(u,a)}|$ never exceeds the probability of the differential (a, b), the quasidifferential trails with the highest correlation tend to have nonzero masks in only a few rounds. We refer to these quasidifferential trails as 'local'. In general, the best quasidifferential trails typically activate as few S-boxes as possible. An S-box is active if the output mask or the input difference is nonzero.

Quasidifferential trails with absolute correlation equal to the correlation of the corresponding differential trail are of particular interest. They correspond to

deterministic linear relations on the intermediate values of right pairs. Perhaps surprisingly, many ciphers admit such quasidifferential trails. One reason for this is that the differentials of many popular S-boxes are *planar* [11]. That is, the right values form an affine space. Propagating this affine space is the basis of the plateau characteristics approach [11], but is difficult to do for more than two rounds. Theorem 4.2 is related to these quasidifferential trails and will be useful in Sections 6 to 8.

Theorem 4.2. *For a function* $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ *and a characteristic* $a_1, \ldots, a_{r+1}$ *with correlation* p *(as quasidifferential trail), it holds that:*

(1) If $(u_1, a_1), \ldots, (u_{r+1}, a_{r+1})$ *is a quasidifferential trail with correlation* $(-1)^b\, p$ *where* $b \in \{0, 1\}$*, then for any quasidifferential trail* $(v_1, a_1), \ldots, (v_{r+1}, a_{r+1})$ *with correlation* c*, the correlation of the quasidifferential trail* $(u_1 + v_1, a_1), \ldots, (v_{r+1} + u_{r+1}, a_{r+1})$ *is* $(-1)^b\, c$*.*

(2) If the correlations of any number of quasidifferential trails with differences $a_1, \ldots, a_{r+1}$ *and correlation* $\pm p$ *sum to zero, then the probability of the characteristic* $a_1, \ldots, a_{r+1}$ *is zero.*

Proof. By Theorem 4.1 the second property follows from the first one, since it implies that the set of all quasidifferential trails can be partitioned into subsets whose correlations sum to zero. For the first property, note that the correlation of the quasidifferential trail $(u_1, a_1), \ldots, (u_{r+1}, a_{r+1})$ equals $\pm p$ if and only if $D^{\mathsf{F}_i}_{(u_{i+1}, a_{i+1}), (u_i, a_i)} = \pm D^{\mathsf{F}_i}_{(0, a_{i+1}), (0, a_i)}$ for $i = 1, \ldots, r-1$.

By Equation (4), this implies that $u_{i+1}^{\mathsf{T}} \mathsf{F}_i(x) = u_i^{\mathsf{T}} x + b_i$ for all x such that $\mathsf{F}_i(x + a_i) = \mathsf{F}_i(x) + a_{i+1}$. Hence, again by Equation (4), the correlation of the i^{th} transition of the quasidifferential trail $(u_1 + v_1, a_1), \ldots, (u_{r+1} + v_{r+1}, a_{r+1})$ is multiplied by a factor $(-1)^{b_i}$. The result then follows from $b = \sum_{i=1}^{r} b_i$. □

Finally, we briefly consider how strong quasidifferential trails can exist for a large number of rounds of a cipher. For every active S-box in a quasidifferential trail that is not active in the corresponding characteristic, the correlation of the trail contains a factor equal to the correlation of an ordinary linear approximation over that S-box. These approximations never have correlation ± 1, since the S-box is a nonlinear function. Hence, to avoid activating too many differentially inactive S-boxes, the masks of the quasidifferential trail should follow the differences as closely as possible. By Theorem 3.2 (5), one structural property that makes this more likely is if the linear layer L of the cipher satisfies $\mathsf{L}^{-1} = \mathsf{L}^{\mathsf{T}}$. Such 'self-dual' linear layers, including all bit-permutations, are in common use. Insights such as these can be used by designers to avoid strong key-dependency or, should they choose to do so, to amplify key-dependent effects on purpose.

4.4 Key-Alternating Ciphers

For key-alternating ciphers, quasidifferential trails with nonzero masks have an intuitive interpretation. Let $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ with $\mathsf{F}_i(x) = \mathsf{G}_i(x) + k_i$. By Equa-

tion (5) and Theorem 3.1 (2), it holds that

$$D^{\mathsf{F}}_{\varpi_{r+1},\varpi_1} = \sum_{\varpi_2,\ldots,\varpi_r} \prod_{i=1}^{r} (-1)^{u_{i+1}^{\mathsf{T}} k_i} D^{\mathsf{G}_i}_{\varpi_{i+1},\varpi_i}\,, \tag{6}$$

where $\varpi_i = (u_i, a_i)$ for $i = 1, \ldots, r+1$. It is easy to see that for $u_0 = u_r = 0$, the average of the above with respect to independent uniform random round keys $k_1, \ldots, k_r$ is equal to the sum of the average probabilities of all characteristics. More generally, one has the following result.

Theorem 4.3. *Let* $\mathsf{F} = \mathsf{F}_r \circ \cdots \circ \mathsf{F}_1$ *with* $\mathsf{F}_i(x) = \mathsf{G}_i(x) + k_i$. *If* $\boldsymbol{k} = (\boldsymbol{k}_1, \ldots, \boldsymbol{k}_r)$ *is a uniform random variable on a set* $\mathcal{K}$, *then*

$$\Pr\left[\mathsf{F}(\boldsymbol{x}+a) = \mathsf{F}(\boldsymbol{x}) + b\right] = \sum_{\substack{u_2,\ldots,u_r \\ a_2,\ldots,a_r \\ (u_2,\ldots,u_r)\perp\mathcal{K}}} \prod_{i=1}^{r} D^{\mathsf{G}_i}_{(u_{i+1},a_{i+1}),(u_i,a_i)}\,,$$

where $u_1 = u_{r+1} = 0$ *and the probability is over a uniform random* $\boldsymbol{x}$ *and over the keys* $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_r$. *In particular, for* $\mathcal{K} = \mathbb{F}_2^n$, *only quasidifferential trails with zero masks contribute to the key-averaged probability of the differential.*

Proof. Taking the average of both sides of Equation (6) with respect to $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_r$ yields the result, since $\sum_{i=1}^{r} u_{i+1}^{\mathsf{T}} \boldsymbol{k}_i$ is zero when $(u_2, \ldots, u_r) \in \mathcal{K}^{\perp}$ and uniform random otherwise. □

A result similar to Theorem 4.3 but for characteristics follows from Theorem 4.1. Furthermore, Equation (6) allows computing the variance of the probability of a differential:

$$\mathsf{E}\big[D^{\mathsf{F}}_{\varpi_{r+1},\varpi_1}\big]^2 + \mathsf{Var}\big[D^{\mathsf{F}}_{\varpi_{r+1},\varpi_1}\big] = \sum_{\varpi_2,\ldots,\varpi_r} \prod_{i=1}^{r} \big(D^{\mathsf{G}_i}_{\varpi_{i+1},\varpi_i}\big)^2\,.$$

This result is analogous to a well-known result of Nyberg [22] in the context of linear cryptanalysis, which states that the variance of the correlation of a linear approximation is equal to the sum of the squared correlations of the linear trails in the approximation.

5 Computing the Quasidifferential Transition Matrix

The differential cryptanalysis of specific primitives using quasidifferential trails requires calculating the quasidifferential transition matrix for each round transformation. For affine functions, Theorem 3.2 (4) and (5) show how to compute the quasidifferential transition matrix.

In general, calculating the quasidifferential transition matrix is nontrivial because the dimensions of the matrix D^{F} scale exponentially with the number of input and output bits of F. In the following two sections, we show that this is not an issue for most primitives: we provide efficient methods to compute the quasidifferential transition matrix for small (such as 4- or 8-bit) S-boxes, for the bitwise-and operations and for modular additions.

5.1 S-boxes

The matrix D^{F} can be computed using a number of operations roughly proportional to its number of elements. Specifically, for a function $\mathsf{F} : \mathbb{F}_2^n \to \mathbb{F}_2^m$, the matrix D^{F} can be computed in $\mathcal{O}((n+m)2^{2n+2m})$ time using a method similar to the fast Fourier transform. Specifically, the matrix $\mathscr{Q}_n$ with columns $\beta_{u,a}$ satisfies $\mathscr{Q}_n = \mathscr{Q}_1^{\otimes n}$. It follows that there exists an efficient divide-and-conquer algorithm for multiplication with $\mathscr{Q}_n$ or its transpose, analogous to the fast Fourier transform. Hence, since $D^{\mathsf{F}} = \mathscr{Q}_m \, (T^{\mathsf{F}} \otimes T^{\mathsf{F}}) \, \mathscr{Q}_n^{\mathsf{T}} / 2^n$ by Definition 3.2, the matrix D^{F} can be computed by applying this divide-and-conquer multiplication algorithm to both the rows and columns of $T^{\mathsf{F}} \otimes T^{\mathsf{F}}$. A Sage implementation of this algorithm is provided as supplementary material. It is also possible to compute the quasidifferential transition matrix from the correlation matrix of F using essentially the same approach. This is discussed in the extended version of the paper.

5.2 Bitwise-And and Modular Addition

Several ciphers use bitwise-and or modular addition as their nonlinear components. Although these functions potentially have many input and output bits, they are highly structured. This makes it possible to express the entries of their quasidifferential transition matrix in terms of relatively simple logical constraints. These constraints can be used to model the propagation of quasidifferential trails in such ciphers as an MILP, SAT or SMT problem, *cf.* Section 8.

In the following, the bitwise-and of $x, y \in \mathbb{F}_2^n$ will be denoted by $x \wedge y$, the bitwise or by $x \vee y$. We also define $\mathsf{and}(x \| y) = x \wedge y$. The bitwise complement of x will be written as $\bar{x}$. The addition of the integers represented by x and y modulo 2^n will be denoted by $\mathsf{add}(x \| y)$. Finally, we write $x \preccurlyeq y$ when $x_i \leq y_i$ for $i = 1, \ldots, n$.

Bitwise-And. The quasidifferential transition matrix of and is easy to compute because it acts on each bit independently. Hence, Theorem 3.2 (2) can be used. This results in the following theorem. The proof can be found in the extended version of the paper.

Theorem 5.1. *Let $a, b, c \in \mathbb{F}_2^n$ be differences and $u, v, w \in \mathbb{F}_2^n$ masks. It holds that $D^{\mathsf{and}}_{(w,c),\,(u\|v,a\|b)} \neq 0$ if and only if $c \preccurlyeq a \vee b$, $u \vee v \preccurlyeq a \vee b \vee w$ and $a \wedge u + b \wedge v = c \wedge w$. Furthermore, if these conditions hold, then*

$$D^{\mathsf{and}}_{(w,c),\,(u\|v,a\|b)} = 2^{-\operatorname{wt}(a \vee b) - \operatorname{wt}(w \wedge \bar{a} \wedge \bar{b})} \, (-1)^{u^{\mathsf{T}}(\bar{a} \wedge c) + v^{\mathsf{T}}(a \wedge c) + u^{\mathsf{T}}(a \wedge b)} \, .$$

Modular Addition. The quasidifferential transition matrix of add can be computed using its CCZ equivalence to a quadratic function [23] that is nearly the same as bitwise-and. This results in Theorem 5.2. The proof is given in the extended version of the paper. In Theorem 5.2, the linear map $M : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is defined by $M(x)_1 = 0$ and $M(x)_i = \sum_{j=1}^{i-1} x_j$ for $i > 1$ and its 'pseudoinverse' is $M^{\dagger}(x) = [x + (x \ll 1)] \gg 1$, where $\ll$ and $\gg$ denote left and right shifts respectively.

Theorem 5.2. *Denote the map of modular addition with modulus 2^n by* $\mathsf{add} : \mathbb{F}_2^{2n} \to \mathbb{F}_2^n$. *Furthermore, let $a, b, c \in \mathbb{F}_2^n$ be differences and $u, v, w \in \mathbb{F}_2^n$ masks. It holds that $D^{\mathsf{add}}_{(w,c),(u\|v,a\|b)} \neq 0$ if and only if*

$$\begin{aligned} c'_1 &= 0 \\ M^\dagger c' &\preccurlyeq a' \vee b' \\ u' \vee v' &\preccurlyeq a' \vee b' \vee M^\mathsf{T} w' \\ a' \wedge u' + b' \wedge v' &= c' \wedge M^\mathsf{T} w' \\ (a'_n = b'_n = 0) \vee (a'_n u'_n + b'_n v'_n \neq w'_n) &\vee (a'_n v'_n = \bar{a}'_n u'_n)\,, \end{aligned}$$

where $(a', b', c') = (b+c, a+c, a+b+c)$ and $(u', v', w') = (u+w, v+w, u+v+w)$. Furthermore, if the above conditions hold, then

$$D^{\mathsf{add}}_{(w,c),\,(u\|v,a\|b)} = 2^{z - \mathrm{wt}(a' \vee b') - \mathrm{wt}(M^\mathsf{T} w' \wedge \bar{a}' \wedge \bar{b}')}\, (-1)^{(\bar{a}' \wedge M^\dagger c' + a' \wedge b')^\mathsf{T} u' + (a' \wedge M^\dagger c')^\mathsf{T} v'}\,,$$

where $z = (a'_n \vee b'_n) \wedge (a'_n u'_n + b'_n v'_n = w'_n) \wedge (a'_n v'_n \neq \bar{a}'_n u'_n)$.

6 Application to RECTANGLE

RECTANGLE [28] is a 64-bit substitution-permutation network, with a nonlinear layer consisting of 4-bit S-boxes and a bit-permutation as the linear layer. The state is represented by a 4×16 array of bits. For the specification of RECTANGLE, we refer the reader to the extended version of the paper.

There are several reasons why RECTANGLE is an interesting target to illustrate the use of quasidifferential trails. The linear layer is a bit-permutation and simpler compared to similar ciphers such as PRESENT [8]. In particular, it rotates the second, third and fourth rows of the state by 1, 12 and 13 bits respectively. As discussed in Section 4.3, the self-duality of bit-permutations potentially results in quasidifferential trails with high absolute correlation relative to the probability of the corresponding differential trail. In addition, differential cryptanalysis is the dominant attack for RECTANGLE. The optimal differentials for RECTANGLE also have a limited differential effect, *i.e.* they contain few characteristics. This simplifies the analysis.

To perform the analysis in this section, we developed an SMT-based program to automate the search for quasidifferential trails in RECTANGLE. This tool is provided as supplementary material and can easily be adapted to similar ciphers such as PRESENT. Additional details can be found in the extended version of the paper.

6.1 Differentials

Table 1 lists several differentials for RECTANGLE. Differential i is a 14-round differential used in the best published key-recovery attack on RECTANGLE [28].

Although its probability is suboptimal, its input and output differences are betted suited for key-recovery. The corresponding 18-round key-recovery attack requires 2^{64} data and enough memory to hold 2^{72} counters. The time-complexity amounts to $2^{78.67}$ (80-bit key) or $2^{126.66}$ (128-bit key) 18-round encryptions. A success probability of 67% is claimed.

Differential ii has a dominant characteristic with average probability 2^{-61}. Based on the analysis of the designers (which included differential effects), this differential is believed to have a maximal average probability. Up to rotational equivalence, there are a total of 32 such differentials. However, as discussed below, these differentials all have similar behavior.

The average probability of differential iii is suboptimal, but the analysis in Section 6.2 shows that its probability is much larger for some keys.

Table 1: Differentials (a, b) for 14 rounds of RECTANGLE. The column p_{avg} gives an estimate of the average differential probability for independent round-keys.

a	b	p_{avg}	Comment	№
0020000600000000	0004000000000020	$2^{-63} + 2^{-66}$	18-round key-recovery	i
0100007000000000	0861008400000010	$2^{-61} + 2 \cdot 2^{-64}$	'Optimal' (1 of 32).	ii
00000000c0000600	0004000000000020	$2 \cdot 2^{-65} + 13 \cdot 2^{-68}$	'Suboptimal'.	iii

6.2 Analysis

In order to search for optimal quasidifferential trails, we model the propagation of the masks for a fixed difference as a 'Satisfiability Modulo Theories' (SMT) problem. Using Theorem 4.1, quasidifferential trails allow us to compute the probability of a characteristic. The extended version of the paper contains additional information about the SMT model and its implementation.

Differential i. For completeness, we list the two dominant characteristics for this differential in the extended version of the paper. The first two columns of Table 2 list the number of quasidifferential trails of each correlation for these two characteristics.

Any characteristic has at least one quasidifferential trail with correlation equal to its average probability p_{avg}, namely the trail with all-zero masks. The fact that the first characteristic has two quasidifferential trails with correlation $\pm p_{\mathsf{avg}}$ and the second four, is special. Table 3 shows two of these trails (one for each characteristic) with the same masks. Only rounds 9 to 12 are shown, since the masks are zero in all other rounds. Hence, these two trails describe a local, three-round effect. This is already an interesting outcome of our approach by itself, since previous techniques such as plateau characteristics are not able to describe such three-round effects.

Table 2: Number of quasidifferential trails for 14 rounds of RECTANGLE.

$\lvert c\rvert/p_{\mathsf{avg}}$	Differential i		Differential ii			Differential iii	
	2^{-63}	2^{-66}	2^{-61}	2^{-64}	2^{-64}	2^{-65}	2^{-65}
1	2	4	2	2	4	32	32
2^{-1}	2	4	2	2	4	32	32
2^{-2}	26	52	24	24	48	352	352
2^{-3}	26	60	24	24	56	480	480
2^{-4}	182	396	176	176	384	2656	2656

Table 3: Differences and masks for two three-round quasidifferential trails with absolute correlation 2^{-13} and 2^{-19}. The masks are the same for both trails.

Differences ($p_{\mathsf{trail}} = 2^{-63}$)	Differences ($p_{\mathsf{trail}} = 2^{-66}$)	Masks (both)
.....2....6.	.....2....6.	.....
.....c....2.	.....c....2.	.....c.....
.....86..	.....86..	.....84..
.....12..	.....92..	.....12..
.....3...	.....3..8	.....3...
.....8...	.....8..1	.....

Note that the propagation of the masks closely follows that of the differences. As discussed in Section 4.3, this is beneficial to obtain quasidifferential trails with high correlation. The correlation for the quasidifferential trail corresponding to the first characteristic in rounds 9 to 12 is equal to

$$\begin{aligned}&(-1)^{\kappa_1} \times D^{\mathsf{S}}_{(\mathsf{c},\mathsf{c}),(2,0)} D^{\mathsf{S}}_{(2,0),(6,0)} \times D^{\mathsf{S}}_{(1,1),(8,8)} D^{\mathsf{S}}_{(2,2),(6,4)} \times D^{\mathsf{S}}_{(8,0),(3,3)} \\ &= (-1)^{\kappa_1} \times \frac{-1}{8} \times \frac{1}{4} \times \frac{1}{8} \times \frac{1}{4} \times \frac{1}{8} = (-1)^{1+\kappa_1}\, 2^{-13}\,,\end{aligned}$$

where $\kappa_1 = k_{10,10} + k_{10,15} + k_{11,12} + k_{11,13}$. Similarly, for the second characteristic, the correlation of the quasidifferential trail is equal to

$$\begin{aligned}&(-1)^{\kappa_1} \times D^{\mathsf{S}}_{(\mathsf{c},\mathsf{c}),(2,0)} D^{\mathsf{S}}_{(2,0),(6,0)} \times D^{\mathsf{S}}_{(9,1),(8,8)} D^{\mathsf{S}}_{(2,2),(6,4)} \times D^{\mathsf{S}}_{(8,0),(3,3)} D^{\mathsf{S}}_{(1,0),(8,0)} \\ &= (-1)^{\kappa_1} \times \frac{-1}{8} \times \frac{1}{4} \times \frac{-1}{8} \times \frac{1}{4} \times \frac{1}{8} \times \frac{1}{8} = \ (-1)^{\kappa_1}\, 2^{-19}\,.\end{aligned}$$

Note the sign difference compared to the first characteristic. As shown below, it implies that the two characteristics are incompatible: for each key, one of them must have probability zero. Taking into account the first four quasidifferential trails, the probability of the first characteristic is

$$p_{i,1} \approx \big(1 - (-1)^{\kappa_1}\big)\big(1 + (-1)^{\lambda}/2\big)2^{-63} = \mathbb{1}_{\kappa_1 = 1}\big(1 + (-1)^{\lambda}/2\big)2^{-62}\,,$$

where λ is a linear combination of round-key bits. Although we did not include all quasidifferential trails in the analysis, Theorem 4.2 (2) allows concluding

that the characteristic has probability zero when $\kappa_1 = 0$. Furthermore, it can be argued that lower-correlation trails are typically less significant. Although it is possible that for example the 26 trails with correlation 2^{-65} contribute a term of magnitude $2^{-63.3}$, this only happens for a small fraction of keys since it requires the signs of all these trails to point in the same direction. For the second characteristic, considering the first 8 trails results in

$$\begin{aligned} p_{i,2} &\approx \big(1 + (-1)^{\kappa_1} - (-1)^{\kappa_2} - (-1)^{\kappa_1+\kappa_2}\big)\big(1 + (-1)^{\lambda}/2\big)2^{-66} \\ &= \mathbb{1}_{\kappa_1=0}\mathbb{1}_{\kappa_2=1}\big(1 + (-1)^{\lambda}/2\big)2^{-64}\,. \end{aligned}$$

Impact on the key-recovery attack. The time-complexity of the 18-round key-recovery attack based on differential i is determined by the number of remaining pairs for the right key after filtering the data. For the maximal number of input structures, the number of remaining unordered pairs will be $p_i\, 2^{63}$.

If $\kappa_1 = 0$, then the number of pairs is $\mathbb{1}_{\kappa_2=1}(2 + (-1)^{\lambda})/4$ on average over the remaining key bits. Since this is less than one for all values of κ_2 and λ, the key-recovery advantage will be too low to improve over brute-force.

For $\kappa_1 = 1$, the average number of unordered pairs is $2 + (-1)^{\lambda}$. Using a threshold of one pair as in the original attack, this gives a time-complexity of $2^{77.65}$ (80-bit key) or $2^{125.65}$ (128-bit key) assuming that the cost of evaluating the key-schedule is negligible compared to the cost of evaluating the cipher. Assuming that the number of right pairs follows a Poisson distribution within each key class, the success probability is then approximately $(1 - e^{-1})/2 + (1 - e^{-3})/2 \approx 79\%$. Hence, for this case, the attack still marginally improves over exhaustive search. However, achieving this improvement requires filtering for weak keys using the condition $\kappa_1 = 1$ during the key-recovery phase. Otherwise, no improvement over exhaustive search is obtained. The observations above can be summarized as follows.

Result 1. *The key-recovery attack on 18-round* RECTANGLE *from [28] using differential i does not improve over exhaustive search. For keys with* $k_{10,10} + k_{10,15} + k_{11,12} + k_{11,13} = 1$*, the attack can be modified to filter out keys not satisfying this condition and then achieves a success probability of approximately 79% with a time-complexity of* $2^{77.65}$ *(80-bit key) or* $2^{125.65}$ *(128-bit key) 18-round encryptions. The attack requires* 2^{64} *data and enough memory to store* 2^{72} *counters.*

By Result 1, there is a rectified 18-round key-recovery attack on RECTANGLE with *average* success probability 39.5% and (marginally) better time-complexity than exhaustive search.

Differential ii. The analysis of differential ii is very similar to that of i. The three dominant characteristics are given in the extended version of the paper. Based on the first four trails for the first two characteristics and the first eight trails

for the second, the characteristic probabilities are

$$p_{ii,1} \approx \mathbb{1}_{\kappa_1=1}\left(1+(-1)^{\lambda}/2\right)2^{-60}$$
$$p_{ii,2} \approx \mathbb{1}_{\kappa_1=1}\left(1+(-1)^{\lambda}/2\right)2^{-63}$$
$$p_{ii,3} \approx \mathbb{1}_{\kappa_1=0}\mathbb{1}_{\kappa_2=0}\left(1+(-1)^{\lambda}/2\right)2^{-62}\,.$$

That is, for half of the keys, the dominant characteristic actually has no right pairs. For the other keys, its probability is roughly twice as large. The second characteristic shows similar behavior. Also note that the third characteristic is not compatible with the first two.

A similar analysis was performed for all other (up to rotational equivalence) 14-round differentials with a dominant characteristic of average probability 2^{-61}. The results were essentially the same.

Differential iii. Both characteristics with probability 2^{-65} are given in the extended version of the paper. Based on the 32 quasidifferential trails with correlation 2^{-65}, we find that the first characteristic has a nonzero probability if and only if 5 linearly independent equations in the round keys hold. The average probability over these keys is 2^{-60}. For the second characteristic, we find a similar effect with slightly different conditions on the round keys. Like for the first characteristic, the average probability over these keys is 2^{-60}. Furthermore, the conditions for the two characteristics to have nonzero probability (given in the extended version) are incompatible. Hence, the sum of the probabilities of the first two characteristics is 2^{-60} for 1/16 keys and zero for all other keys.

In addition, there are 13 characteristics with an average probability of 2^{-68}. We find that each of these characteristics has nonzero probability zero for only 1/64 or 1/128 keys. The conditions for this to happen may partially overlap or be inconsistent with the conditions for the first two characteristics.

7 Application to KNOT

In order to illustrate the relevance of our techniques to the analysis of permutations, we analyze several differential attacks on the KNOT family of permutations and their authenticated-encryption and hashing modes [29]. KNOT is a large-state variant of RECTANGLE and was a second-round candidate in the NIST lightweight cryptography project. In this paper, we only consider the primary variant, which is a 256-bit permutation. The state is represented by a 4×64 rectangular array. The round function operations are similar to those of RECTANGLE, but a different S-box is used and the third and fourth row of the state are rotated by 8 and 25 positions respectively. Additional details may be found in the extended version of the paper.

7.1 Differentials

At the 2020 NIST lightweight cryptography workshop, Zhang *et al.* [30] presented several differential attacks on round-reduced KNOT authenticated encryption and hashing modes. The differentials used in these attacks are listed in

Table 4, along with their estimated probabilities (without taking into account quasidifferential trails). In this section, it will be shown that these attacks do not work because the probability of the differentials in Table 4 is much smaller than expected. Furthermore, it will be shown that there exist round constants for which their probabilities are two orders of magnitude larger. All relevant characteristics are listed in the extended version of the paper.

Table 4: Differentials for r rounds of KNOT-256. The column p_{avg} gives an estimate of the 'average' differential probability (for independent uniform random round constants). The differences are given in the extended version of the paper.

r	p_{avg}	Application	№
10	5×2^{-56}	Hash collision and AEAD forgery.	i
12	10×2^{-66}	Hash collision and AEAD forgery.	ii

7.2 Analysis

The analysis of the differentials in Table 4 is similar to the analysis for RECTANGLE. The SMT-model for RECTANGLE can easily be modified to efficiently search for quasidifferential trails in KNOT.

Differential i. Based on the quasidifferential trails with correlation 2^{-56} for each of the five characteristics with $p_{\mathsf{avg}} = 2^{-56}$, we conclude that all of them have probability zero for the standard round constants of KNOT-256. Hence, the differential probability is much lower than what might be expected from the 'average'. Even if there exist other characteristics with unexpectedly large probability (a scenario considered below), this is a significant issue for the collision attack on the KNOT hash function. Indeed, the collision search consists of finding a right pair for one of the best few characteristics, since this is significantly easier than finding a right pair for the differential by random search.

Despite the observations above, it is possible that there exists a low-probability characteristic with an unexpectedly high probability for the default round constants. The differential contains four characteristics with 'average' probability 2^{-60}. However, by analyzing the corresponding quasidifferential trails, we find that they too have probability zero. Next, there are 17 characteristics with 'average' probability 2^{-62}. Again, we find that all of them have probability zero. We also considered 24 characteristics with 'average' probabilities 2^{-63} and 2^{-65} and found that they have probability zero. Although we did not analyze all characteristics with probability 2^{-66} or lower, they can only have a high nonzero probability for a very small fraction of round constants. Given the number of such characteristics, it is unlikely that a high probability characteristic exists.

On the flip side, there exist round constants for which one or more of the five characteristics have probability 2^{-50}. This is due to the existence of 64

quasidifferential characteristics with absolute correlation 2^{-56}. A careful inspection of the conditions on the round constants shows that there exist variants of KNOT with modified constants for which the probability of differential i is approximately $5 \cdot 2^{-50} = 2^{-47.7}$. Further improvements are possible by taking into account additional characteristics and quasidifferential trails.

Differential ii. The analysis of the 12-round differential is similar to the 10-round differential, and leads to similar conclusions. This is not surprising given that both characteristics follow a similar pattern up to rotational symmetry. We find that each of the 10 dominant characteristics has probability zero for the default round constants. In addition, we did not find any characteristics with 'average' probability 2^{-70} or higher with a nonzero probability. Hence, it is unlikely that the 12-round forgery and collision attacks presented by Zhang *et al.* are valid. Finally, we can identify round constants for which one or more of the 10 characteristics has a probability of 2^{-59}.

8 Application to Speck

In this section we investigate the key-dependency of several differentials for Speck from the literature. The bitvector constraints for modular addition from Theorem 5.2 are the main ingredient of our SMT-model. The same approach can be applied to any ARX block cipher or permutation. The implementation of our model is provided as supplementary material.

In Section 8.1 we provide a simple explanation (using a single quasidifferential trail) for an experimental observation of Ankele and Kölbl [2] on Speck-64. In Sections 8.2 and 8.3 we analyze the differentials used in the best published attacks on all variants of Speck. In the extended version of the paper, Speck is briefly reviewed.

8.1 Explaining Observations of Ankele and Kölbl on Speck-64

Ankele and Kölbl [2] experimentally estimated the probability of a 7-round differential for Speck-64 for 10000 random keys and found that the distribution of the number of rights pairs is bimodal. Their results are reproduced in Figure 4, but colored to indicate two key classes that follow from the analysis below.

The fact that the histogram in Figure 4 is bimodal already suggests the presence of an important quasidifferential trail with nonzero masks. Automatic search reveals that the best such quasidifferential trail has correlation 2^{-23}. The dominant characteristic (with probability 2^{-21}) and the masks of the quasidifferential trail with correlation 2^{-23} are shown in Table 5.

The quasidifferential trail from Table 5 only involves the modular additions of the first two rounds. Figure 5 shows the propagation of the mask-difference pairs for these rounds in more detail. Following Section 4.3, the interpretation of this trail is that there exists a linear combination of the output of the first modular addition which is biased for the right pairs. This implies that a rotated

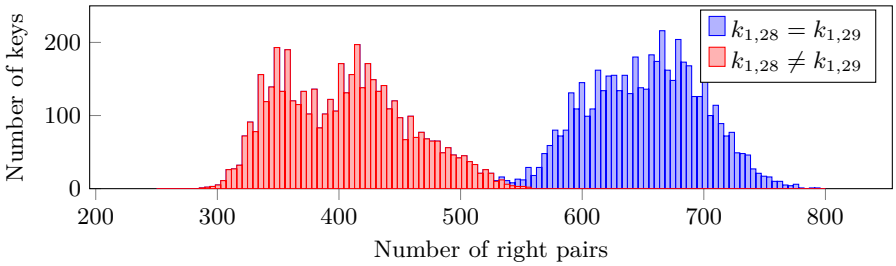

Fig. 4: Number of right pairs for the SPECK-64 differential from [2], for a total of 10000 keys. For each key, 2^{30} pairs were sampled uniformly at random.

Differences	Masks
4...4.92 1.42..4.	....
82.2.... ..12.2..	18...... ....
..9..... ....1...	....
....8...	....
......8. ......8.	....
8.....8. 8....48.	....
..8..48. ..8.2.84	....
8.8.a.8. 8481a4a.	....

Table 5: Differential trail with probability 2^{-21} for 7 rounds of Speck-64, and the masks of a quasidifferential trail with correlation 2^{-23}.

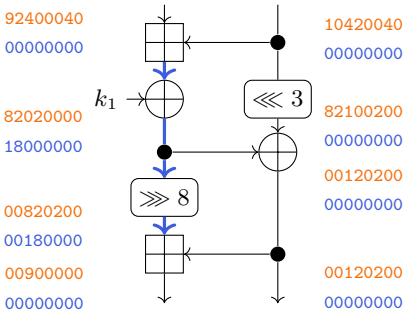

Fig. 5: Quasidifferential trail with correlation $2^{-5} \times 2^{-6} = 2^{-11}$ through two rounds of Speck-64, with differences in orange and masks in blue.

linear combination of the left input of the second modular addition is biased. This bias results in a smaller or larger number of right pairs, depending on the value of a linear combination of bits of k_1. Specifically, the probability can be estimated as

$$2^{-21} + (-1)^{k_{1,28}+k_{1,29}}\, 2^{-23}\,.$$

For 2^{30} random input pairs, the average number of right pairs is still $2^9 = 512$. However, the above formula predicts that the average is $512 + 2^7 = 640$ if $k_{1,28} = k_{1,29}$ and $512 - 2^7 = 384$ otherwise. This explains most of the variation in the experimental results shown in Figure 4. Additional effects, such as the more limited bimodal behavior for $k_{1,28} \neq k_{1,29}$, can be explained by taking into account additional quasidifferential trails.

8.2 Analysis of Differential Attacks on Speck-32

The best published attacks on reduced-round Speck are differential attacks using the enumeration key-recovery strategy proposed by Dinur [13]. Given an r-round

differential, an $r+3$ round attack is obtained by prepending one round (for free) and appending two rounds. For variants with longer key lengths, one performs the same attack for each guess of the last few round keys.

In this section, we analyze the best published attacks on Speck-32 reduced to 11-14 rounds. These attacks rely on the 6-9 round differentials shown in Table 6. Lee *et al.* [17] report on a 10-round differential with average probability $2^{-30.39}$, but it does not lead to a 15-round key-recovery attack because the time-complexity would be $2^{31.39}$ for a success probability of $1 - 1/e \approx 63\%$.

Table 6: Differentials (a, b) for r-round Speck-32.

r	a	b	p_{avg}	Reference	№
6	0211	0a04 850a 9520	2^{-13}	Abed *et al.* [1]	i
7	0a60	4205 850a 9520	2^{-18}	Abed *et al.* [1]	ii
8	1488	1008 850a 9520	$2^{-24} + 2^{-27}$	Abed *et al.* [1]	iii
9	8054	a900 0040 0542	$2^{-30} + 2 \cdot 2^{-33\dagger}$	Biryukov *et al.* [7], Song *et al.* [24]	iv

† $3060307 \cdot 2^{-47} \approx 2^{-29.45}$ with characteristics of average probability $\leq 2^{-49}$

Differentials i and ii. The six round differential i is dominated by a characteristic with average probability 2^{-13}, given in the extended version of the paper. The next-best characteristic has average probability 2^{-23} and will be ignored in our analysis. We find two quasidifferential trails with correlation $\pm 2^{-15}$ and two with correlation $\pm 2^{-17}$. There also exist trails with absolute correlation 2^{-19} and lower, but their effect on the probability is limited except for a small fraction of keys. Grouping these trails appropriately, the following estimate is obtained:

$$p_i \approx (1 + (-1)^{0003^\mathsf{T} k_5}/4)(1 + (-1)^{0180^\mathsf{T} k_5}/4)2^{-13}\,,$$

where, for simplicity, only one trail of correlation $\pm 2^{-17}$ is included.

The analysis of the seven round differential is similar. The dominant differential trail has average probability 2^{-18} and is the same as the six round trail with one additional round at the beginning. Hence,

$$p_{ii} \approx (1 + (-1)^{0003^\mathsf{T} k_6}/4)(1 + (-1)^{0180^\mathsf{T} k_6}/4)2^{-18}\,.$$

Differential iii. The differential is dominated by two characteristics. The first has average probability 2^{-24}. Since the last part of these characteristics is the same as for the dominant characteristics of differentials i and ii, some of the same quasidifferential trails are obtained. However, there also exist quasidifferential trails with correlation equal to the probability of the trail. This implies that there exists keys for which these characteristics have probability zero. Specifically, for the first characteristic, we find that

$$p_{iii,1} \approx \mathbb{1}_{0600^\mathsf{T} k_2 = 0} \mathbb{1}_{1800^\mathsf{T} k_3 = 0} (1 + (-1)^{0003^\mathsf{T} k_7}/4)(1 + (-1)^{0180^\mathsf{T} k_7}/4)\, 2^{-22}\,.$$

That is, its probability is zero for 3/4 keys, but four times larger for the other keys. For the second characteristic, we have

$$p_{iii,2} \approx \mathbb{1}_{\mathtt{0600}^\mathsf{T}k_2=0}\mathbb{1}_{\mathtt{1800}^\mathsf{T}k_3=0}\mathbb{1}_{\mathtt{0a00}^\mathsf{T}k_2=0}(1+(-1)^{\mathtt{0003}^\mathsf{T}k_7}/4)(1+(-1)^{\mathtt{0180}^\mathsf{T}k_7}/4)\,2^{-24}\,.$$

Hence, the second characteristic has nonzero probability only when the first probability is nonzero *and* $\mathtt{0a00}^\mathsf{T}k_2 = 0$.

Differential iv. The probability is dominated by three characteristics (listed in the extended version of the paper). Additional characteristics only increase the overall probability, but more detailed analysis reveals that many additional characteristics have probability zero for most keys, and high probability for a relatively small fraction of keys.

The first characteristic has average probability 2^{-30}. Based on all quasidifferential trails with absolute correlation $\geq 2^{-32}$, we obtain

$$p_{iv,1} \approx \mathbb{1}_{\mathtt{000c}^\mathsf{T}k_5=0}(1-(-1)^{\mathtt{0180}^\mathsf{T}k_1}/4)\,2^{-29}\,.$$

For the second characteristic (with average probability 2^{-33}), the quasidifferential trails with absolute correlation $\geq 2^{-34}$ yield

$$p_{iv,2} \approx \mathbb{1}_{\mathtt{6000}^\mathsf{T}k_2=1}(1+(-1)^{\mathtt{000c}^\mathsf{T}k_5}/2+(-1)^{\mathtt{0300}^\mathsf{T}k_4+\mathtt{000c}^\mathsf{T}k_5}/2)\,2^{-32}\,.$$

Note that one of the two quasidifferential trails with absolute correlation 2^{-34} involves three modular additions. By Theorem 4.2, the condition $\mathtt{6000}^\mathsf{T}k_2 = 1$ is necessary to obtain a nonzero probability. However, the conditions $\mathtt{0300}^\mathsf{T}k_4 = 0$ and $\mathtt{000c}^\mathsf{T}k_5 = 1$ only imply a small but possibly nonzero correlation. For the third characteristic, we consider all quasidifferential trails with absolute correlation $\geq 2^{-35}$ and obtain

$$p_{iv,3} \approx \mathbb{1}_{\mathtt{0c00}^\mathsf{T}k_2=1}\mathbb{1}_{\mathtt{000c}^\mathsf{T}k_5=0}(1-(-1)^{\mathtt{0180}^\mathsf{T}k_1}/2)\,2^{-31}\,.$$

Note that the condition $\mathtt{000c}^\mathsf{T}k_5 = 0$ is shared with the first characteristic. Since the probability of the second characteristic is too low, this implies that previous key-recovery attacks on 14 rounds of Speck-32 work for only half of the keys.

Impact on key-recovery attacks. The above analysis allows us to reevaluate the best published attacks on reduced-round Speck-32. The attack on 13 rounds only works for one in four keys. Likewise, the attack on 14 rounds works only for half of the keys. Another way to formulate this is that the (key-averaged) success probability of these attacks is much lower than expected. For eleven and twelve rounds, the success probability is also slightly lower, but less so. Unfortunately, restoring the previous success-probability is not possible except by using alternative differentials.

However, if the results of our analysis are taken into account, weak-key attacks with lower data requirements are obtained. These attacks can be optimized either with respect to the number of weak keys, or with respect to the data-complexity. To minimize the data-complexity, we make assumptions on the key

to maximize the probability of the differential. To maximize the number of keys for which the attack works, only conditions to ensure nonzero probabilities are imposed. Assuming that the adversary stops requesting data once the key has been found[4], these attacks require less data than what would be expected based on the average-case analysis.

Table 7: Rectified attacks on r-round Speck-32.

r	Time *encryptions*	Data *chosen plaintexts*	Weak-keys *density*	Comment
11	$2^{45.36}$	$2^{13.36}$	2^{-2}	Optimized for data
	$2^{45.88}$	$2^{13.88}$	1	Optimized for number of keys
12	$2^{50.36}$	$2^{18.36}$	2^{-2}	Optimized for data
	$2^{50.88}$	$2^{18.88}$	1	Optimized for number of keys
13	$2^{54.03}$	$2^{22.03}$	2^{-5}	Optimized for data
	$2^{56.20}$	$2^{24.20}$	2^{-2}	Optimized for number of keys
14	$2^{61.84}$	$2^{29.84}$	2^{-1}	Optimized for number of keys

The results are shown in Table 7. For example, the 6-round differential (11 round attack) has a probability at most $(1 + 1/4)^2\, 2^{-13} \approx 2^{-12.36}$. With early stopping, the average number of pairs required is $2^{13}(1/(1 - 1/4)^2 + 2/(1 - 1/4^2) + 1/(1 + 1/4)^2)/4 \approx 2^{12.88}$. For 14 rounds, we omit the attack optimizing the data-complexity, since it requires more time than exhaustive search over a key space of size 2^{64-1} for a similar success probability.

8.3 Analysis of Differential Attacks on Larger Variants of Speck

The techniques to analyze Speck-32 in Section 8.2 carry over to the larger variants of Speck. In this section, we reevaluate the best published attacks on these variants. They rely on the key-recovery technique of Dinur [13] and are based on the differentials shown in Table 8 below. For 16 rounds of Speck-96, Song *et al.* [24] also propose a differential with average probability $2^{-94.94}$. However, we do not include it as its probability is too low to improve over exhaustive search.

Most of the differentials in Table 8 rely on a significant differential effect. Nevertheless, the analysis below will be limited to a few characteristics in each case. This is done only to simplify the analysis, since each characteristic has its own key-dependent behaviour that is not independent of other characteristics. Note that including additional characteristics can only increase the probability of the differential. In addition, it will be shown that key-dependence is much more significant than the differential effect for all differentials in Table 8. A detailed analysis of the differentials in Table 8 is given in the extended version.

[4] This is possible due to the way the key-recovery attack works.

Table 8: Differentials for r-round Speck-n (differences in the extended version). The average differential probability is p_{avg}, the average probability of the analyzed characteristics is p_{char}. The values p_{min} and p_{max} are the minimum and maximum value of the probability of the analyzed characteristics.

n	r	p_{avg}	p_{char}	p_{min}	p_{max}	Reference	№
48	11	$2^{-44.31}$	$2^{-46} + 2^{-47}$	0	2^{-43}	Song *et al.* [24]	i
64	15	$2^{-60.56}$	2^{-62}	0	2^{-59}	Song *et al.* [24]	ii
96	15	$2^{-81.00}$	2^{-81}	0	$2^{-73.68}$	Song *et al.* [24]	iii
128	20	$2^{-124.35}$	$4 \cdot 2^{-128}$	0	$2^{-120.36}$	Song *et al.* [24]	iv

Impact on key-recovery attacks. The analysis above directly impacts the key-recovery attacks based on the differentials from Table 8. Like for Speck-32, all of these attacks have lower success probability than previously expected. Nevertheless, the analysis also leads to weak-key attacks with lower data-complexity. The results are summarized in Table 9.

Note that for Speck-128, our analysis shows that the key-recovery attacks probably do not improve over exhaustive search over the reduced key-space. Improvements may be possible if checking the weak-key conditions can be made comparatively cheap, provided that checking candidate keys dominates the cost. Since a detailed analysis of the time-complexity is outside of the scope of this paper, Table 9 only lists a distinguisher for this case. Although our analysis did not include all characteristics, these would only increase the *average* differential probability by $2^{-124.9}$. Further analysis shows that the probabilities of these characteristics are strongly key-dependent. Hence, the key-recovery attacks on Speck-128 from [24] most likely do not improve over exhaustive search.

Acknowledgements. We thank Anne Canteaut and Jean-René Reinhard for responding to our questions about their attacks on PRINCE. Tim Beyne is supported by a PhD Fellowship from the Research Foundation – Flanders (FWO). This work was partially supported by the Research Council KU Leuven, grant C16/18/004 on New Block Cipher Structures.

References

1. Abed, F., List, E., Lucks, S., Wenzel, J.: Differential cryptanalysis of round-reduced Simon and Speck. In: FSE 2014. LNCS, vol. 8540, pp. 525–545
2. Ankele, R., Kölbl, S.: Mind the gap - A closer look at the security of block ciphers against differential cryptanalysis. In: SAC 2018. LNCS, vol. 11349, pp. 163–190
3. Beyne, T.: Block cipher invariants as eigenvectors of correlation matrices. In: ASIACRYPT 2018, Part I. LNCS, vol. 11272, pp. 3–31
4. Beyne, T.: A geometric approach to linear cryptanalysis. In: ASIACRYPT 2021, Part I. LNCS, vol. 13090, pp. 36–66
5. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. In: CRYPTO'90. LNCS, vol. 537, pp. 2–21

Table 9: Rectified attacks on r-round Speck.

Variant	r	Time *encryptions*	Data *chosen plaintexts*	Weak-keys *density*	Comment
48/72	15	2^{68}	2^{44}	2^{-3}	Optimized for data
		$2^{68.58}$	$2^{44.58}$	2^{-2}	Optimized for number of keys
48/96	16	2^{92}	2^{44}	2^{-3}	Optimized for data
		$2^{92.58}$	$2^{44.58}$	2^{-2}	Optimized for number of keys
64/96	19	2^{92}	2^{60}	2^{-3}	——
64/128	20	2^{124}	2^{60}	2^{-3}	——
96/96	18	$2^{74.68}$	$2^{74.68}$	2^{-9}	Optimized for data
		$2^{77.25}$	$2^{77.25}$	2^{-6}	Optimized for number of keys
96/144	19	$2^{122.68}$	$2^{74.68}$	2^{-9}	Optimized for data
		$2^{125.25}$	$2^{77.25}$	2^{-6}	Optimized for number of keys
128/m	20	$2^{121.36}$	$2^{121.36}$	2^{-7}	Distinguisher (data-optimized)
		$2^{125.36}$	$2^{125.36}$	2^{-3}	Distinguisher (key-optimized)

6. Biham, E., Shamir, A.: Differential cryptanalysis of the full 16-round DES. In: CRYPTO'92. LNCS, vol. 740, pp. 487–496
7. Biryukov, A., Roy, A., Velichkov, V.: Differential analysis of block ciphers SIMON and SPECK. In: FSE 2014. LNCS, vol. 8540, pp. 546–570
8. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An ultra-lightweight block cipher. In: CHES 2007. LNCS, vol. 4727, pp. 450–466
9. Canteaut, A., Lambooij, E., Neves, S., Rasoolzadeh, S., Sasaki, Y., Stevens, M.: Refined probability of differential characteristics including dependency between multiple rounds. IACR Trans. Symm. Cryptol. **2017**(2), 203–227
10. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In: FSE'94. LNCS, vol. 1008, pp. 275–285
11. Daemen, J., Rijmen, V.: Plateau characteristics. IET Inf. Secur. **1**(1), 11–17
12. De Cannière, C., Rechberger, C.: Finding SHA-1 characteristics: General results and applications. In: ASIACRYPT 2006. LNCS, vol. 4284, pp. 1–20
13. Dinur, I.: Improved differential cryptanalysis of round-reduced Speck. In: SAC 2014. LNCS, vol. 8781, pp. 147–164
14. Heys, H.M.: Key dependency of differentials: Experiments in the differential cryptanalysis of block ciphers using small S-boxes. ePrint, Report 2020/1349 (2020)
15. Knudsen, L.R.: Iterative characteristics of DES and s^2-DES. In: CRYPTO'92. LNCS, vol. 740, pp. 497–511
16. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: EUROCRYPT'91. LNCS, vol. 547, pp. 17–38
17. Lee, H., Kim, S., Kang, H., Hong, D., Sung, J., Hong, S.: Calculating the approximate probability of differentials for ARX-based cipher using SAT solver. Journal of the Korea Institute of Information Security & Cryptology **28**(1), 15–24
18. Leurent, G.: Analysis of differential attacks in ARX constructions. In: ASIACRYPT 2012. LNCS, vol. 7658, pp. 226–243
19. Liu, Y., Zhang, W., Sun, B., Rijmen, V., Liu, G., Li, C., Fu, S., Cao, M.: The phantom of differential characteristics. Des. Codes Cryptogr. **88**(11), 2289–2311

20. Matsui, M.: Linear cryptanalysis method for DES cipher. In: EUROCRYPT'93. LNCS, vol. 765, pp. 386–397
21. Mendel, F., Rijmen, V., Toz, D., Varici, K.: Differential analysis of the LED block cipher. In: ASIACRYPT 2012. LNCS, vol. 7658, pp. 190–207
22. Nyberg, K.: Linear approximation of block ciphers (rump session). In: EUROCRYPT'94. LNCS, vol. 950, pp. 439–444
23. Schulte-Geers, E.: On CCZ-equivalence of addition mod 2^n. Des. Codes Cryptogr. **66**(1-3), 111–127
24. Song, L., Huang, Z., Yang, Q.: Automatic differential analysis of ARX block ciphers with application to SPECK and LEA. In: ACISP 16, Part II. LNCS, vol. 9723, pp. 379–394
25. Sun, L., Wang, W., Wang(66), M.: More accurate differential properties of LED64 and Midori64. IACR Trans. Symm. Cryptol. **2018**(3), 93–123
26. Wang, X., Yu, H.: How to break MD5 and other hash functions. In: EUROCRYPT 2005. LNCS, vol. 3494, pp. 19–35
27. Xu, Z., Li, Y., Jiao, L., Wang, M., Meier, W.: Do NOT misuse the Markov cipher assumption – Automatic search for differential and impossible differential characteristics in ARX ciphers. ePrint, Report 2022/135 (2022)
28. Zhang, W., Bao, Z., Lin, D., Rijmen, V., Yang, B., Verbauwhede, I.: RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms. Sci. China Inf. Sci. **58**(12), 1–15
29. Zhang, W., Ding, T., Yang, B., Bao, Z., Xiang, Z., Ji, F., Zhao, X.: KNOT: Algorithm specifications and supporting document. Submission to NIST lightweight cryptography project
30. Zhang, W., Ding, T., Zhou, C., Ji, F.: Security analysis of KNOT-AEAD and KNOT-Hash. NIST Lightweight Cryptography Workshop (2020)