Bounds in Shallows and in Miseries* **

Céline Blondeau[1], Andrey Bogdanov[2], and Gregor Leander[3]

[1] Aalto University, School of Science, Finland
celine.blondeau@aalto.fi
[2] Technical University of Denmark, Denmark
anbog@dtu.dk
[3] Ruhr University Bochum, Germany
gregor.leander@rub.de

Abstract. Proving bounds on the *expected differential probability* (EDP) of a characteristic over all keys has been a popular technique of arguing security for both block ciphers and hash functions. In fact, to a large extent, it was the clear formulation and elegant deployment of this very principle that helped Rijndael win the AES competition. Moreover, most SHA-3 finalists have come with explicit upper bounds on the EDP of a characteristic as a major part of their design rationale. However, despite the pervasiveness of this design approach, there is no understanding of what such bounds actually mean for the security of a primitive once a key is fixed — an essential security question in practice.

In this paper, we aim to bridge this fundamental gap. Our main result is a quantitative connection between a bound on the EDP of differential characteristics and the highest number of input pairs that actually satisfy a characteristic for a fixed key. This is particularly important for the design of permutation-based hash functions such as sponge functions, where the EDP value itself is not informative for the absence of rekeying. We apply our theoretical result to revisit the security arguments of some prominent recent block ciphers and hash functions. For most of those, we have good news: a characteristic is followed by a small number of pairs only. For Keccak, though, currently much more rounds would be needed for our technique to guarantee any reasonable maximum number of pairs.

Thus, our work — for the first time — sheds light on the fixed-key differential behaviour of block ciphers in general and substitution-permutation networks in particular which has been a long-standing fundamental problem in symmetric-key cryptography.

Keywords: block cipher, hash function, differential cryptanalysis, differential characteristic, expected differential probability, Grøstl.

---

* *"There is a tide in the affairs of men / Which, taken at the flood, leads on to fortune; / Omitted, all the voyage of their life / **Is bound in shallows and in miseries**. / On such a full sea are we now afloat; / And we must take the current when it serves, / Or lose our ventures".* The Tragedy of Julius Caesar by William Shakespeare. Act 4, Scene 3.

** This is final version of the paper which would be presented at CRYPTO 2013

1 Introduction

Block Ciphers and Hash Functions. Block ciphers and hash functions are at the very core of cryptography today, being accountable for absolutely most data encryption and authentication occurring in the field. It is not by accident that block ciphers (AES) and hash functions (SHA) are among the few cryptographic algorithms standardized by NIST, the U.S. National Institute of Standards and Technology. The security properties of block ciphers and hash functions are largely interconnected. The traditional way of building a hash function has been to employ a block cipher in a mode of operation, such as Davies-Meyer, Matyas-Meyer-Oseas, Miyaguchi-Preneel or Hirose. Rather lately, it has become popular to build hash functions from permutations which are usually obtained by fixing a key in a well-understood block cipher. While SHA-1 and SHA-2 conform to the former design principle (having the SHACAL block ciphers at their foundation), SHA-3 (Keccak) — building upon the sponge construction [3] — adopts the latter one. In the paper, we will deal with the cryptographic fixed-key properties of block ciphers (or, equivalently, with properties of permutations) — a research field that, rather unduly as we think, has not received much attention recently.

Differential Cryptanalysis, Differential Characteristics, Probabilities. Any sound newly developed block cipher or hash function comes with strong arguments against *differential cryptanalysis*. It was introduced in 1990 by Biham and Shamir [7] for recovering the key of round-reduced DES — the former U.S. Data Encryption Standard. Later they used it to attack the full DES [8]. Differential cryptanalysis was known to the designers of DES (IBM with NSA involvement) back in the 1970s though [15]. Based on the seminal idea of differential cryptanalysis, plenty of extensions have been proposed [5,12,27,28,35]. In fact, it was a variant of differential cryptanalysis that resulted in the first key recovery for the full AES, though in a weak related-key model [9].

Since its publication for the case of DES, differential cryptanalysis has been applied to numerous *iterative block ciphers*, that is, block ciphers whose data transform consists of consecutive application of similar simpler maps (*rounds*).

In the differential cryptanalysis context, given a pair of inputs to the cipher with a certain difference Δ_0, one tracks the propagation of this difference through the r round transforms resulting in intermediate differences Δ_i, $i = 1, \ldots, r-1$, and an output difference Δ_r. The sequence of all these $r+1$ differences $(\Delta_0, \ldots, \Delta_r)$ is called a *differential characteristic*. In a block cipher, once the key K is fixed, the fixed-key *differential characteristic probability* (DP) π_K is the probability for a pair of inputs to follow this differential characteristic. When π_K is averaged over all round keys, one obtains the *expected differential characteristic probability* (EDP) π. The paper at hand contributes to the fundamental understanding of the links between DP and EDP.

Hypothesis of Stochastic Equivalence and Plateau Characteristics. As regards the connection between the fixed-key DP π_K and the EDP π for a characteristic in a cipher, the common *hypothesis of stochastic equivalence* [29] states

that $\pi \approx \pi_K$ for almost all keys. However, there is an essential gap between EDP π and DP π_K, since there can be a significant discrepancy between those values. Probably the most prominent example of a strictly non-equivalent behaviour is actually constituted by the AES and its *plateau characteristics*, that is, differential characteristics that, depending on the key, have a probability of either 0 or 2^{h-n}, where h is the *height* of the plateau characteristic and n is the block size. One presumes that most characteristics over the full AES are plateau of height 1 [21,33] with $n = 128$. At the same time, it is well known that the value of EDP for, say, AES-128 does not exceed 2^{-330}.

Bounds on Differential Characteristic Probability as Security Argument. While the computation of the DP value or even its informative upper-bounding is known to be a difficult problem in symmetric-key cryptography, it turns out that it is possible to compute an upper bound on the value of EDP for a characteristic — at least for some suitable ciphers and under some assumptions. Both block cipher and hash function designers tend to compute such a bound on the EDP whenever possible. This bound is widely accepted as a valid security argument not only for block ciphers [14, 18, 26] but also for hash functions [2, 4, 6, 24, 36].

The starting point of computing an upper bound on EDP is the *Markov cipher* assumption which requires the iterative cipher to be such that the transition probability $\pi_{\Delta_i,\Delta_{i+1}}$ for differences $\Delta_i \to \Delta_{i+1}$ over a round does not depend on the actual input value, where the probability is taken over the keys. Under the Markov cipher assumption and if round keys are independent, it can be shown [29] that the product of all transition probabilities equals the EDP, $\pi = \prod_i \pi_{\Delta_i,\Delta_{i+1}}$.

This approach gives the designers of new block ciphers and hash functions a formal way of arguing the resistance of their primitives towards differential cryptanalysis. Eventually, it took the symmetric-key community several years since the introduction of differential cryptanalysis to come up with the paradigm that finally manifested itself as a winning approach and stipulated the spread of substitution-permutation networks: *the wide trail design strategy* [19] (*decorrelation theory* [34] being another example of a similar but somewhat more general approach). Here, one builds a primitive such that the minimum number of *active S-boxes* (i.e. nonlinear components with nonzero input and output differences in a differential characteristic) over all nontrivial differential characteristics is maximized. Then an upper bound on the difference propagation probability through an S-box is used to compute the EDP for the full characteristic. In fact, to a large extent, it was the clear formulation and elegant deployment of the wide trail design strategy that helped Rijndael win the NIST AES competition.

The application of such approaches is by far not limited to block ciphers though. Also three out of five SHA-3 finalists (Keccak [4], Grøstl [24] and JH [36]) – including the SHA-3 winner – have come with wide-trail type security arguments, providing some bounds on the EDP. However, the exact meaning – quantitative or even qualitative – of these bounds appears to have been unclear.

Criticism of the Expected Differential Characteristic Probability. Indeed, while an upper bound on EDP does contain information on the behaviour of a fixed-key differential characteristic on average, interpreting it can be rather confusing, even for block ciphers. It is not clear what such a bound says when the key is fixed, like it is the case in almost any block cipher based encryption procedure or in permutation based hash functions.

As already mentioned above, an upper bound on EDP can get rather low. For instance, consider the permutation P (or Q) of Grøstl-256. This is a permutation on 512 bits and the designers show that, under the round independency assumption, any differential characteristic has a probability of less than 2^{-972}, by using the wide-trail design strategy. Since 2^{511} (unordered) pairs are possible, this *might seem* to indicate that any given characteristic is fulfilled by zero pairs. But then, it is trivial to find many characteristics that are fulfilled by at least one pair – just take two inputs with some difference Δ_I, apply the permutation P to it and note down the intermediate differences Δ_i after each round.

Of course, this situation is not specific to Grøstl since similar arguments are provided by, among many others, the designers of the SHA-3 winner Keccak, SHA-3 finalist JH as well as lightweight hash functions such as SPONGENT [13] and PHOTON [25]. Here, lower bounds on the number of differentially active S-boxes can only be translated to upper bounds on the *expected* differential characteristic probability (in other words, on the *expected* number of pairs following a characteristic), averaged over *all keys*. At the same time, such designs rely on a permutation which is a substitution-permutation network for a *fixed key*, usually supplied in form of round constants. The assumption that the rounds are independent can only be argued strictly if we have independent round keys. Here this is clearly not the case as *the key (resp. the round constants) is fixed*.

The Motivation. The question of what bounds on expected differential characteristic probabilities actually imply for permutation-based hash functions (like Keccak, Grøstl and JH as well as many more recent hash functions such as SPONGENT and PHOTON) – or even for a block cipher with a fixed key – remains unanswered. Thus, there is a fundamental lack of understanding of what those bounds mean. The significance of this problem is emphasized by the large number of designs that use such bounds without discussing their impact. So, given that there will always be characteristics that are fulfilled by at least one pair, what can we hope for? From a designer's point of view — focusing on differential characteristics — a reasonable goal will be that with high probability, there is no characteristic that is fulfilled by more than one pair.

Now, the critical question is: How small should a bound be on the EDP of a differential characteristic to guarantee this goal above? Concretely, in the example of Grøstl-256, is 2^{-972} enough, too big or already far too small?

Somewhat more specifically, the research problem we aim to address in this work is the following:
Given an upper bound on the expected differential characteristic probability (EDP) π for a block cipher over all keys, what is the probability to have at most B pairs of inputs following a differential characteristic for one fixed key?

Our Main Contribution. In this paper, we answer this question formally and shed some light on what those bounds mean for constructions with a fixed key or a fixed constant. The only assumption we are making in our work is that the number of (unordered) pairs that satisfy a given characteristic follows a binomial distribution over all possible keys (resp. round constants).

Now we formulate our main result. Let B be a bound on the number of pairs of input that fulfill a differential characteristic in a block cipher (or a permutation), once the key is fixed. Let q_B be the probability that all nontrivial characteristics are fulfilled by at most B pairs of inputs. Recall that we denote the block size in bits by n. The main result of our paper is summarized in the next theorem.

Theorem (Main Result). *If π is an upper bound on the expected differential characteristic probability (EDP) for a block cipher (or a permutation) over all keys, the probability q_B that all nontrivial characteristics are fulfilled by at most B input pairs is lower-bounded by:*

$$q_B \geq 1 - \frac{\pi^B}{(B+1)!2^B} 2^{(B+2)n}.$$

Given this result, one can now obtain the greatest sufficient value of an upper bound on EDP to achieve the design goal of a permutation having at most $B = 1$: An upper bound on π of 2^{-3n-7} or lower suffices to attain this goal with probability $q_1 \geq 0.99$. Using this theorem reciprocally for already existing designs, we can state, for instance, that the designers' upper bound of $\pi = 2^{-972}$ on the EDP for the P or Q permutations in Grøstl is sufficient to have at most 3 pairs with a probability of at least $q_3 = 1 - 2^{-363.58}$.

Though we do not consider the EDP and DP of differentials, our work does shed light on a fundamental, previously ignored, problem in the design of (round-based) fixed permutations. The only requirement to apply our result is to provide a bound on the EDP of a characteristic, which is the only indicator of security against differential cryptanalysis that designers are usually able to give.

2 Preliminaries

In this section, we introduce our notation and subsequently the statistical model along with its single assumption. We want our model to deal with all differential characteristics of a cipher. For this purpose, we introduce what we coin as the *differential characteristic spectrum* of a cipher. In a nutshell, this is a list of probabilities of characteristics along with their quantity. Note that, while this spectrum is very suitable to model the differential behaviour of a given cipher, for any real-world cipher it is completely out of reach to compute this spectrum. We therefore will later, cf. Section 3.1, explain how to bypass the need to compute the entire spectrum. In this sense, the introduction of the spectrum is a way of eliminating most of it in a sound manner.

2.1 The Model

Let n denote the bit size of the primitive (permutation size or block size). Let p denote the probability of a differential characteristic and X_p denote the random variable (taken over independent round keys) which corresponds to the number of pairs that fulfill a differential characteristic with probability p. Note that we are always taking the *whole input space* into consideration. In order to simplify the treatment, we furthermore always considered *unordered pairs*. The sole assumption underlying our model, and thus our results, is that X_p follows a binomial distribution. More precisely:

Assumption 1 *X_p follows (over the independent round keys) a binomial distribution with parameters (N, p), i.e.*

$$X_p \sim \mathcal{B}(N, p)$$

where $N = 2^{n-1}$ is the total number of unordered pairs with a fixed difference.

This assumption has been used frequently in the literature, cf. Section 2.2 for more details. The only class of characteristics that clearly do not follow a binomial distribution we are aware of are plateau characteristics, most prominently present in the AES [21,22,33]. We discuss plateau characteristics in Appendix A and explain in which cases our result extends to those characteristics as well.

Now, as outlined above, we do not only deal with a single characteristic, but with all characteristics at the same time. There are usually characteristics with many different probabilities, and conversely many characteristics for a given probability. Following [10] for a similar concept, we capture this information in what we refer to as the *differential spectrum* of a cipher, cf. also the characteristic weight counting function of [20]. For this, denote by $(p_i)_i$ the set of all occurring probabilities and by A_i the number of characteristics with probability p_i. For convenience we assume that $p_i \geq p_{i+1}$, i.e. the probabilities are ordered in descending order. Note that we explicitly exclude the trivial characteristics, i.e. the characteristic with input difference 0.

Definition 1 (Differential Spectrum). *The vector of pairs $\mathcal{S} = ((p_i, A_i))_i$ is called the* differential spectrum *of a cipher.*

The complete list of differential characteristics is modeled, according to Assumption 1, by random variables $X_j^{(p_i)} \sim \mathcal{B}(p_i, 2^{n-1})$, where $1 \leq j \leq A_i$. Clearly, it holds that

$$\sum_i p_i A_i = 2^n - 1 \tag{1}$$

simply as every pair follows some (non-trivial) characteristic. Actually, even more is true, namely

$$\Pr\left(\sum_i \sum_j X_j^{(p_i)} = 2^{n-1}(2^n - 1)\right) = 1.$$

From this perspective it seems reasonable to assume that the vector $(X_j^{(p_i)})_{i,j}$ is multinomial distributed. However, this way one would assume that no other dependency between the individual characteristic exists. This is especially doubtable in the case of characteristics that are identical for a large number of rounds and only diverge in the very last (or first) round.

Let us turn to our main focus, i.e. studying the question of what the maximal number of pairs is that follow a characteristic. In the above model, this corresponds to studying the distribution of the random variable $Z^{\mathcal{S}}$ defined as

$$Z^{\mathcal{S}} = \max_{i,j}\{X_j^{(p_i)}\}. \tag{2}$$

The relevance of $Z^{\mathcal{S}}$ is the following. If, for example, we can argue (within our model) that $\Pr(Z^{\mathcal{S}} \leq B) = 0.99$, we are guaranteed that choosing random round constants (resp. round keys), will result in 99 out of 100 cases in a permutation such that no characteristic is followed by more than B pairs.

In particular, for ensuring that any characteristic is fulfilled by at most one pair, $\Pr(Z^{\mathcal{S}} \leq 1)$ should be close to one.

In the sequel, as in Assumption 1, $N = 2^{n-1}$ will denote the number of unordered pairs with a fixed difference. Furthermore, B will denote the bound on the number of pairs we consider and $q_B = P(Z^{\mathcal{S}} \leq B)$ denotes the probability that no characteristic is fulfilled by more than B pairs. For a given spectra we denote $p = p_0$, i.e. the maximal probability of a characteristic (or an upper bound on it).

2.2 Binomial Distribution

Since differential cryptanalysis is one of the major cryptanalytic techniques in symmetric-key cryptography, the distributions of random variables associated with differential characteristics and differentials have been extensively studied over the past 20 years. In [1], examples with different EDP and fixed-key DP are considered and it is noted that computing the average values does not in general allow one to draw conclusions about the shape of the distributions. In a similar direction, the work [23] develops a model derived from binomial distribution and performs experiments in the case of a random permutation. Moreover, it provides instances of ciphers for which this model holds and does not hold. In [22], it is shown that for only 2 rounds of the AES, this model is not correct due to the existence of plateau characteristics (see the discussion in Appendix A).

In [11], experiments show that — at least for some relevant ciphers — the distribution is actually binomial, that is, as stated in Assumption 1. To establish a clear independent empirical basis for our theoretical study and to support the meaningfulness of the assumption, we conducted experiments of our own on a 16-bit reduced version of PRESENT [14]. The experiments depicted in Figure 1 correspond to 5, 6 and 7 rounds of SmallPresent: As the number of rounds increases and the EDP of the best characteristic decreases, the deviation from the binomial distribution is almost non-existant and clearly indicates that Assumption 1 is realistic in that case.

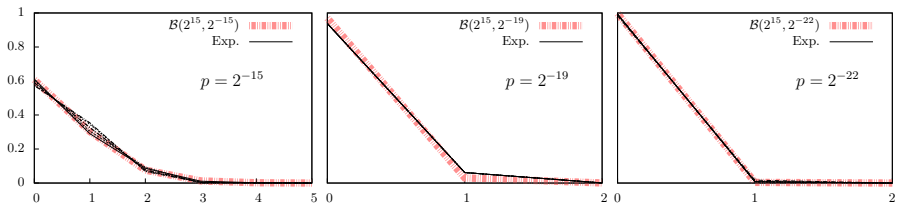

Fig. 1. Distribution for the characteristics of SmallPresent (n=16): Comparison between the distribution of different characteristics with EDP p and the binomial distribution $\mathcal{B}(2^{n-1}, p)$.

3 The Link: Bounding the Bound

In this section, we present and prove the main result of this work. First, we show how to avoid computing the whole spectrum and still be able to make useful statements about the distribution of $Z^{\mathcal{S}}$, as introduced in (2). Second, we provide the proof of our main theorem outlined in the introduction.

3.1 Cutting the Spectrum of a Cipher

As introduced in Section 2, we assume that the number of pairs fulfilling a given characteristic with a certain probability follows a binomial distribution. As we are interested in primitives where no characteristic is satisfied by many pairs, we consider the bound on the EDP which is smaller or equal to 2^{-n}, i.e. $p = p_0 \leq 2^{-n}$ in the sequel.

For a fixed B, we show, in this section, that an underestimate of the $q_B = P(Z^{\mathcal{S}} \leq B)$ can be obtained with partial knowledge of the spectrum. More precisely, we study what we call a *cut* of a given spectrum.

Definition 2. *Given a spectrum* $\mathcal{S} = (p_i, A_i)_{i=0}^{w}$ *we define a* cut spectrum *of order* t $(1 \leq t \leq w)$ *by* $\mathcal{S}_t = (p_i, A'_i)_{i=0}^{t}$ *where*

$$A'_i = A_i \text{ for } 0 \leq i \leq t-1$$

and

$$A'_t = \frac{(2^n - 1) - \sum_{i=0}^{t-1} A_i p_i}{p_t}.$$

The cut spectrum *of order* 0 *is defined by* $\mathcal{S}_0 = (p_0, A'_0)$ *with* $A'_0 = \frac{2^n - 1}{p_0}$.

Thus, by studying a cut of order t of a spectrum, one assumes that all characteristics with probability less than p_{t-1} actually have probability p_t. The definition of A'_t then follows directly from (1).

The following theorem shows that, by studying cuts of a given spectrum, we obtain an underestimate of the cumulative distribution of $Z^{\mathcal{S}}$, cf. (2). While this

might be intuitively reasonable, strictly proving it is rather technical, as we will see below.

Note that all results presented here hold independently of any possible correlation between the individual characteristics. Depending of their relations, tightness of the results can differ but the results remain valid.

Theorem 1. *Given a spectrum $\mathcal{S}$ and its cuts $(\mathcal{S}_t)_t$, the following holds for any $B \geq 1$.*

$$q_B = \Pr(Z^{\mathcal{S}} \leq B) \geq \Pr(Z^{\mathcal{S}_t} \leq B) \geq \cdots \geq \Pr(Z^{\mathcal{S}_0} \leq B).$$

In order to prove the statement of the theorem, we first introduce two technical lemmata.

Lemma 1. *Let $2^{-n} \geq p_1 \geq p_2$ be two probabilities. For all, $i \geq 2$ we have*

$$p_1^{i-1}(1-p_1)^{N-i} - p_2^{i-1}(1-p_2)^{N-i} \geq 0.$$

Proof. To prove this inequality, we shall prove that

$$T = \left(\frac{p_1}{p_2}\right)^{i-1} \left(\frac{1-p_1}{1-p_2}\right)^{N-i} \geq 1.$$

Assuming that $p_1 \geq p_2$, we have

$$\frac{p_1 - p_2}{p_1} \leq -\log(1 - \frac{p_1 - p_2}{p_1}) = \log(\frac{p_1}{p_2}).$$

Now the proof follows by a succession of inequalities:

$$\begin{aligned}
& N \leq \frac{1}{2p_1} \leq \frac{1-p_1}{p_1 - p_2}\frac{p_1 - p_2}{p_1} \\
\Rightarrow\ & (N-i) \leq N \leq \frac{1-p_1}{p_1 - p_2}\log(p_1/p_2) \\
\Rightarrow\ & (N-i)\frac{p_1 - p_2}{1-p_1} \leq \log(p_1/p_2) \leq (i-1)\log(p_1/p_2) \\
\Rightarrow\ & (N-i)\log(1 + \frac{p_1 - p_2}{1-p_1}) \leq (N-i)\frac{p_1 - p_2}{1-p_1} \leq (i-1)\log(p_1/p_2) \\
\Rightarrow\ & 0 \leq (i-1)\log(p_1/p_2) - (N-i)\log(\frac{1-p_2}{1-p_1}) \\
\Leftrightarrow\ & 1 \leq exp\left[(i-1)\log(p_1/p_2) + (N-i)\log(\frac{1-p_1}{1-p_2})\right] \\
\Leftrightarrow\ & 1 \leq T.
\end{aligned}$$

□

Lemma 2. *Given two random variables X_1, X_2 with $X_i \sim \mathcal{B}(N, p_i)$, $i = 1$ or 2, and $2^{-n} \geq p_1 > p_2$, for all $B \geq 1$ it holds that*

$$\frac{1}{p_1}\Pr(X_1 > B) - \frac{1}{p_2}\Pr(X_2 > B) \geq 0.$$

Proof. We consider

$$\begin{aligned} C &= \frac{1}{p_1}\Pr(X_1 > B) - \frac{1}{p_2}\Pr(X_2 > B) \\ &= \sum_{i=B+1}^{N} \binom{N}{i} \left[p_1^{i-1}(1-p_1)^{N-i} - p_2^{i-1}(1-p_2)^{N-i} \right]. \end{aligned}$$

From Lemma 1, if $B \geq 1$, we have

$$p_1^{i-1}(1-p_1)^{N-i} - p_2^{i-1}(1-p_2)^{N-i} \geq 0.$$

And we conclude that $C \geq 0$. □

Now, using these two lemmata, we can prove that the partial knowledge of the spectrum allows the computation of an underestimate of $q_B = \Pr(Z^{\mathcal{S}} \leq B)$.

Proof (Proof of Theorem 1). To simplify the notation we denote by X_i a random variable that follows a binomial distribution with parameters N and p_i, i.e. $X_i \sim \mathcal{B}(N, p_i)$.

Using the fact the probability of a union of events is smaller than the sum of the probabilities of the different events, independently of the correlation between the different variables X_i, the following holds for any B and any cut spectrum $\mathcal{S}_t$:

$$\begin{aligned} \Pr(Z^{\mathcal{S}_t} \leq B) = 1 - \Pr(Z^{\mathcal{S}_t} > B) &\geq 1 - \sum_{i=0}^{t} \sum_{j=1}^{A'_i} \Pr(X_j^{(p_i)} > B) \\ &\geq 1 - \sum_{i=0}^{t} A'_i \Pr(X_i > B) \end{aligned}$$

For the cut spectra $\mathcal{S}_t = (p_i, A'_i)_{i=0}^{t}$ and $\mathcal{S}_{t-1} = (p_i, \Gamma'_i)_{i=0}^{t-1}$, we first prove that:

$$\forall t > 1 \quad \sum_{i=0}^{t} A'_i \Pr(X_i > B) \leq \sum_{i=0}^{t-1} \Gamma'_i \Pr(X_i > B). \tag{3}$$

Note that, as a consequence of (1), we have

$$A'_t = \frac{(2^n - 1) - \sum_{i=0}^{t-2} A'_i p_i - A'_{t-1} p_{t-1}}{p_t} \text{ and } \quad \Gamma'_{t-1} = \frac{(2^n - 1) - \sum_{i=0}^{t-2} \Gamma'_i p_i}{p_{t-1}}.$$

As for $i < t-1$ we have $A'_i = \Gamma'_i$, we obtain

$$\begin{aligned} C &= \sum_{i=0}^{t} A'_i \Pr(X_i > B) - \sum_{i=0}^{t-1} \Gamma'_i \Pr(X_i > B) \\ &= A'_{t-1}\Pr(X_{t-1} > B) + A'_t \Pr(X_t > B) - \Gamma'_{t-1}\Pr(X_{t-1} > B) \\ &= \left[A'_{t-1} - \frac{(2^n - 1) - \sum_{i=0}^{t-2} A'_i p_i}{p_{t-1}} \right] \Pr(X_{t-1} > B) + A'_t \Pr(X_t > B) \\ &= \left[-\frac{(2^n - 1) - \sum_{i=0}^{t-2} A'_i p_i - A'_{t-1} p_{t-1}}{p_{t-1}} \right] \Pr(X_{t-1} > B) + A'_t \Pr(X_t > B) \\ &= p_t A'_t \left[\frac{1}{p_t}\Pr(X_t > B) - \frac{1}{p_{t-1}}\Pr(X_{t-1} > B) \right]. \end{aligned}$$

Given $p_{t-1} > p_t$, from Lemma 2 it follows that $C \leq 0$. The remaining of the proof follows then applying (3) iteratively for all t. □

Example. Figure 2 illustrates Theorem 1 for a reduced variant of PRESENT. For SMALLPRESENT [30] with a 16-bit block size, it is still feasible to compute the spectrum using a branch and bound approach. Clearly, not all characteristics have the same probability, but different probabilities occur with varying frequency.

As predicted by Theorem 1, taking only part of spectrum into account, i.e. studying $Z^{\mathcal{S}_t}$, leads to an underestimate of $q_B = \Pr(Z^{\mathcal{S}} \leq B)$. In this example, studying $\mathcal{S}_7$ already gives a rather tight estimate of the actual value of q_B.

As mentioned above, for real-world ciphers even computing such a cut spectrum is hard. Indeed, for many primitives, designers can often only provide (a bound on) the EDP of the most probable characteristic.

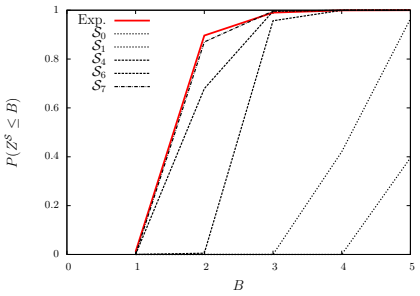

Fig. 2. Influence of cut spectra on $q_B = P(Z^{\mathcal{S}} \leq B)$: Experiments on 6 rounds of SMALLPRESENT with fixed input difference.

3.2 At the Limit: Considering the Bound only

In practice, for many primitives, designers are only able to compute (a bound on) the EDP. In that case, the spectrum is cut to its maximum and is defined by $\mathcal{S}_0 = (\pi, A_0)$, where π is an upper bound (or exact value) of the EDP and, following (1), $A_0 = \frac{2^n-1}{\pi}$ is the total number of these potential characteristics. To simplify the notation, in this section the random variable associated to characteristics of this sort is denoted by X.

We recall here the main result of this paper presented in the introduction, now stated formally:

Theorem 2 (Main Result). *Under Assumption 1, if π is an upper bound on the expected differential characteristic probability (EDP) for a block cipher over all keys (or a permutation over all round constants), the probability q_B that all nontrivial characteristics are fulfilled by at most B input pairs is lower-bounded by:*

$$q_B \geq 1 - \frac{\pi^B}{(B+1)!2^B} 2^{(B+2)n}. \tag{4}$$

Proof. From Theorem 1 and (3), we have:

$$\Pr(Z^{\mathcal{S}} \leq B) \geq \Pr(Z^{\mathcal{S}_0} \leq B) \geq 1 - A_0 P(X > B) \geq 1 - \frac{2^n}{\pi} P(X > B).$$

Meaning that when only π is known, the tail of cumulative function of $Z^{\mathcal{S}}$ can be bounded by:

$$\Pr(Z^{\mathcal{S}} \leq B) \geq 1 - \frac{2^n}{\pi} \Pr(X > B). \tag{5}$$

As we have

$$\begin{aligned}\Pr(X > B) &= \sum_{i=B+1}^{N} \binom{N}{i} \pi^i (1-\pi)^{N-i} \leq \sum_{i=B+1}^{N} \binom{N}{i} \pi^i \\ &\leq \sum_{i=B+1}^{N} \frac{N^i}{(B+1)!} \pi^i \leq 2 \frac{(N\pi)^{B+1}}{(B+1)!},\end{aligned}$$

we conclude that

$$\Pr(Z^S \leq B) \geq 1 - \frac{4}{(B+1)!} N^{B+2} \pi^B.$$

□

4 The Impact: Shallows and Miseries?

In this section, we inspect the impact of our main result on the practical constructions, both block ciphers with a fixed key and permutation-based hash functions.

4.1 Sufficient Condition

From our main result stated in Theorem 2, we can deduce that the upper bound on the EDP of a characteristic in a primitive, should be of order of magnitude $\pi \approx 2^{-[(B+2)/B]n}$, in order for our result to guarantee that no characteristic is fulfilled by more than B pairs. More precisely, from Theorem 2, we immediately obtain the following estimate of π:

Corollary 1 (Sufficient Condition). *Let C_B defined by*

$$C_B = \left[(1 - q_B)(B+1)! 2^B\right]^{1/B}.$$

To guarantee with probability $q_B = P(Z^S \leq B)$ that no characteristic is fulfilled by more than B pairs, it suffices to have the maximal probability π of a characteristic such that

$$\pi \leq C_B 2^{-[(B+2)/B)]n}.$$

By computing the exact value of C_B introduced in Corollary 1, in order to guarantee that for 99% of the keys or fixed constants, no characteristic is fulfilled by more than one pair, we should consider permutations where the maximal EDP is lower than 2^{-3n-7}. For $B = 2$, the maximum EDP can be up to 2^{-2n-2}. For larger values of B, $C_B \approx 1$ and the same security claims can be achieved if $\pi \leq 2^{-[(B+2)/B]n}$. In Figure 3, we illustrate that the sufficient

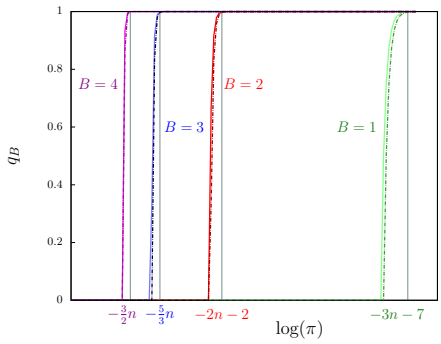

Fig. 3. Bringing it all together: q_B as a function B and EDP bound for characteristics. The dotted curves are computed with (4). The continuous curves are computed using (5) in the case where $n = 64$.

condition of Corollary 1 is rather tight. Note that though computations have been performed for $n = 64$, results are similar for larger values of n. Most importantly, Figure 3 shows that the distributions are extremely steep at around $2^{-[(B+2)/B)]n}$. Thus, a value of π slightly below this threshold does not guarantee anything any more, while a value slightly larger guarantee the non-existence of characteristics followed by more then B pairs almost certainly.

4.2 Revisiting the Security Arguments of Prominent Primitives

In here, we consider only primitives that come with informative bounds on the EDP of their differential characteristics which excludes ARX-based designs, for instance, where no efficient way of arguing tight bounds on EDP is known. However, we prominently note here that this by no means indicate per se that those designs are stronger or weaker in cryptanalytic terms. This only says that we can state much less about those constructions using the state-of-the-art techniques.

The major findings of this section are presented in Table 1. Not all bounds provided by the respective designers of the primitives mentioned in the table are actually tight. Proving better bounds on the EDP would improve the probabilities q_B for those constructions.

As it usually gets harder to provide strong diffusion with the increase of the block size, bigger variants of primitives often have a higher value of B_0. The notable exceptions are constituted by the lightweight hash functions PHOTON and SPONGENT: For PHOTON, while the maximum number of fulfilling pairs

Table 1. Lower bound on the probability to have at most B pairs following a differential characteristic for various primitives. The primitives are either block ciphers with a fixed key (public or secret) or fixed permutations. n: block size. q_B: the probability for all nontrivial characteristics to be fulfilled by at most B input pairs. π: an upper bound on the the expected differential characteristic probability averaged over all keys. − means that Theorem 2 does not provide any informative indication for the parameter set. B_0 is the minimum value of B such that q_B is close to 1.

Primitives	n	π	$q_1 \geq$	$q_2 \geq$	$q_3 \geq$	B_0
AES-128	128	2^{-330}	-	$1-2^{-152.6}$	$1-2^{-357.6}$	2
AES-192	128	2^{-450}	$1-2^{-68}$	$1-2^{-392.6}$	$1-2^{-717.6}$	1
AES-256	128	2^{-480}	$1-2^{-98}$	$1-2^{-452.6}$	$1-2^{-807.6}$	1
Rijndael-192/192	192	2^{-450}	-	$1-2^{-136.6}$	$1-2^{-397.6}$	2
Rijndael-192 /256	192	2^{-480}	-	$1-2^{-196.6}$	$1-2^{-487.6}$	2
Rijndael-256	256	2^{-480}	-	-	$1-2^{-167.6}$	3
PRESENT-80/-128	64	2^{-122}	-	-	$1-2^{-53.6}$	3
LED-64	64	2^{-400}	$1-2^{-210}$	$1-2^{-548.6}$	$1-2^{-887.6}$	1
LED-128	64	2^{-600}	$1-2^{-410}$	$1-2^{-948.6}$	$1-2^{-1487.6}$	1
SPONGENT-88/80/8	88	2^{-176}	-	$1-2^{-4.6}$	$1-2^{-95.6}$	2
SPONGENT-128/128/8	136	2^{-272}	-	$1-2^{-4.6}$	$1-2^{-143.6}$	2
SPONGENT-160/160/16	176	2^{-352}	-	$1-2^{-4.6}$	$1-2^{-183.6}$	2
SPONGENT-224/224/16	240	2^{-480}	-	$1-2^{-4.6}$	$1-2^{-247.6}$	2
SPONGENT-256/256/16	272	2^{-544}	-	$1-2^{-4.6}$	$1-2^{-279.6}$	2
PHOTON-80	100	2^{-216}	-	$1-2^{-36.6}$	$1-2^{-155.6}$	2
PHOTON-128	144	2^{-294}	-	$1-2^{-16.6}$	$1-2^{-169.6}$	2
PHOTON-160	196	2^{-384}	-	-	$1-2^{-179.6}$	3
PHOTON-224	256	2^{-486}	-	-	$1-2^{-185.6}$	3
PHOTON-256	288	2^{-882}	$1-2^{-20}$	$1-2^{-616.6}$	$1-2^{-1213.6}$	1
Grøstl-224/256	512	2^{-972}	-	-	$1-2^{-363.6}$	3
Grøstl-384/512	1024	2^{-1469}	-	-	-	5
JH-224/-256/-384/-512	1024	2^{-1184}	-	-	-	13

grows first with the size, it gets much lower for its biggest version, which is due to the special case design of the largest permutation [25]. For SPONGENT, the distribution of B_0 is very smooth because of its clearly stated design goal: the EDP bound of 2^{-2n} for an n-bit permutation [13].

SHA-3 finalists certainly deserve special treatment. The standard behavior of B_0 (its increase as n grows) is clearly visible in Grøstl. The EDP bounds are not tight for either version of Grøstl though. With the designers' bounds, one can state that it is good news for Grøstl since not more than 5 pairs can satisfy a

characteristic here (and only at most 3 pairs for the smaller variant). The bound of JH is only sufficient to show a maximum of 13 satisfying pairs.

Table 1 does not contain Keccak. The reason is that the best existing bound for the 24 rounds of KECCAK-f[1600] (with a permutation size of $n = 1600$ bits) is actually only 2^{-296} [16]. So, if one aims to attain the goal of having at most two satisfying pairs for a Keccak-type permutation given this bound, 10 times more rounds (240 rounds) would be needed in KECCAK-f[1600]. To achieve $B_0 = 18$, it suffices to take 6 times more rounds (144 rounds). However, if the special case of 1- to 8-symmetric characteristics is considered, a bound of 2^{-1648} can be proven for 18 rounds [4], which leads to $B_0 = 60$ with $q_{60} = 1 - 2^{-18.1}$ for the 18 rounds. Again, we emphasize that this by no means indicates any type of weakness in Keccak. Having high upper bounds on the EDP of a differential characteristic prohibits our model from providing any informative bounds on the number of pairs following a differential characteristic.

5 Conclusion and Future Work

In this paper, we establish the fundamental link of a bound on EDP of a differential characteristic to its fixed-key DP, i.e. the maximum number of input pairs that follow a characteristic. We apply our framework to prominent hash function and block cipher designs. Once the key is fixed (which is almost always the case in practice), our result is the only formal foundation for arguing the crucial differential security properties of symmetric-key primitives available so far.

Having said that, we also clearly state some important open problems which are out of scope of this paper. First, though we constrain ourselves to considering the EDP and DP of differential characteristics here, a much more interesting object of study would be the connection between EDP and DP for *differentials*, which are sets of differential characteristics with certain input and output differences. However, since even bounding EDP for those is notoriously difficult (though not impossible, at least for several rounds of suitable constructions [17]), studying similar questions for differentials seems out of reach given the current techniques (note that [23] considered the differential behaviour of random permutation — a work that technically bears some similarities with ours). Second, though the basic differential cryptanalysis is certainly the most essential differential attack to consider, more advanced techniques such as the *rebound attack* [31] often pose a more critical threat, even in the case where the probabilities of any differential characteristic over the full permutation is very small. However, for rebound attacks, considering differential characteristics over smaller parts of a permutation makes a lot of sense: A good bound over a fraction of rounds will imply that there are only a small number of pairs satisfying a characteristic, so those values are not easy to find. Multiple inbound stages might undermine this reasoning though and it is still an open problem to argue provable security against rebound attacks. We think however that our result opens up the possibil-

ity of making at least some basic security arguments for rebound attacks where it has not been feasible so far to come up with a sound bound considerations.

We believe that those are some of the most important fundamental problems in symmetric-key cryptography open today and would like to see the results of this paper as a first step towards their solution.

References

1. Kazumaro Aoki. On maximum non-averaged differential probability. In Stafford E. Tavares and Henk Meijer, editors, *Selected Areas in Cryptography*, volume 1556 of *LNCS*, pages 118–130. Springer, 1998.
2. Ryad Benadjila, Olivier Billet, Henri Gilbert, Gilles Macario-Rat, Thomas Peyrin, Matt Robshaw, and Yannick Seurin. SHA-3 Proposal: ECHO, 2010. Version 2.0.
3. Guido Bertoni, Joan Daemen, Michael Peeters, and Gilles Van Assche. Sponge functions. In *Ecrypt Hash Workshop 2007*, 2007.
4. Guido Bertoni, Joan Daemen, Michael Peeters, and Gilles van Assche. The Keccak SHA-3 submission, 2011. Version 3.
5. Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In Jacques Stern, editor, *EUROCRYPT*, volume 1592 of *LNCS*, pages 12–23. Springer, 1999.
6. Eli Biham and Orr Dunkelman. The SHAvite-3 Hash Function, 2009. Tweaked Version.
7. Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. In Alfred Menezes and Scott A. Vanstone, editors, *CRYPTO*, volume 537 of *LNCS*, pages 2–21. Springer, 1990.
8. Eli Biham and Adi Shamir. Differential Cryptanalysis of the Full 16-Round DES. In Ernest F. Brickell, editor, *CRYPTO*, volume 740 of *LNCS*, pages 487–496. Springer, 1992.
9. Alex Biryukov, Dmitry Khovratovich, and Ivica Nikolic. Distinguisher and Related-Key Attack on the Full AES-256. In Shai Halevi, editor, *CRYPTO*, volume 5677 of *LNCS*, pages 231–249. Springer, 2009.
10. Céline Blondeau, Anne Canteaut, and Pascale Charpin. Differential properties of power functions. *International Journal of Information and Coding Theory*, 1(2):149–170, 2010.
11. Céline Blondeau and Benoît Gérard. Links Between Theoretical and Effective Differential Probabilities: Experiments on PRESENT. In *Ecrypt Workshop on Tools for Cryptanalysis*, June 2010.
12. Céline Blondeau, Benoît Gérard, and Kaisa Nyberg. Multiple Differential Cryptanalysis Using LLR and χ^2 Statistics. In Ivan Visconti and Roberto De Prisco, editors, *SCN*, volume 7485 of *LNCS*, pages 343–360. Springer, 2012.
13. Andrey Bogdanov, Miroslav Knezevic, Gregor Leander, Deniz Toz, Kerem Varici, and Ingrid Verbauwhede. SPONGENT: A Lightweight Hash Function. In Preneel and Takagi [32], pages 312–325.
14. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and Charlotte Vikkelsø. PRESENT: An ultra-lightweight block cipher. In *Cryptographic Hardware and Embedded Systems - CHES*, volume 4727 of *Springer LNCS*, pages 450–466, 2007.
15. Don Coppersmith. The Data Encryption Standard (DES) and its strength against attacks. *IBM Journal of Research and Development*, 38(3):243–250, 1994.

16. Joan Daemen and Gilles Van Assche. Differential Propagation Analysis of Keccak. In Anne Canteaut, editor, *FSE*, volume 7549 of *LNCS*, pages 422–441. Springer, 2012.
17. Joan Daemen, Mario Lamberger, Norbert Pramstaller, Vincent Rijmen, and Frederik Vercauteren. Computational aspects of the expected differential probability of 4-round AES and AES-like ciphers. *Computing*, 85(1-2):85–104, 2009.
18. Joan Daemen and Vincent Rijmen. The Block Cipher Rijndael. In Jean-Jacques Quisquater and Bruce Schneier, editors, *CARDIS*, volume 1820 of *LNCS*, pages 277–284. Springer, 1998.
19. Joan Daemen and Vincent Rijmen. The Wide Trail Design Strategy. In Bahram Honary, editor, *IMA Int. Conf.*, volume 2260 of *LNCS*, pages 222–238, 2001.
20. Joan Daemen and Vincent Rijmen. Probability distributions of Correlation and Differentials in Block Ciphers. IACR Eprint Report 2005/212, 2005.
21. Joan Daemen and Vincent Rijmen. Understanding Two-Round Differentials in AES. In *SCN*, volume 4116 of *LNCS*, pages 78–94. Springer, 2006.
22. Joan Daemen and Vincent Rijmen. Plateau characteristics. *Iet Information Security*, 1:11–17, 2007.
23. Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology*, 1(3):221–242, 2007.
24. Praveen Gauravaram, Lars R. Knudsen, Krystian Matusiewicz, Florian Mendel, Christian Rechberger, Martin Schläffer, and Søren S. Thomsen. Grøstl – a SHA-3 candidate, 2011.
25. Jian Guo, Thomas Peyrin, and Axel Poschmann. The PHOTON Family of Lightweight Hash Functions. In Phillip Rogaway, editor, *CRYPTO*, volume 6841 of *LNCS*, pages 222–239. Springer, 2011.
26. Jian Guo, Thomas Peyrin, Axel Poschmann, and Matthew J. B. Robshaw. The LED Block Cipher. In Preneel and Takagi [32], pages 326–341.
27. Goce Jakimoski and Yvo Desmedt. Related-Key Differential Cryptanalysis of 192-bit Key AES Variants. In Mitsuru Matsui and Robert J. Zuccherato, editors, *Selected Areas in Cryptography*, volume 3006 of *LNCS*, pages 208–221. Springer, 2003.
28. Lars R. Knudsen. Truncated and Higher Order Differentials. In Bart Preneel, editor, *FSE*, volume 1008 of *LNCS*, pages 196–211. Springer, 1994.
29. Xuejia Lai, James L. Massey, and Sean Murphy. Markov Ciphers and Differentail Cryptanalysis. In Donald W. Davies, editor, *EUROCRYPT*, volume 547 of *LNCS*, pages 17–38. Springer, 1991.
30. Gregor Leander. Small Scale Variants Of The Block Cipher PRESENT. *IACR Cryptology ePrint Archive*, 2010:143, 2010.
31. Florian Mendel, Christian Rechberger, Martin Schläffer, and Søren S. Thomsen. The Rebound Attack: Cryptanalysis of Reduced Whirlpool and Grøstl. In Orr Dunkelman, editor, *FSE*, volume 5665 of *LNCS*, pages 260–276. Springer, 2009.
32. Bart Preneel and Tsuyoshi Takagi, editors. *Cryptographic Hardware and Embedded Systems - CHES 2011 - 13th International Workshop, Nara, Japan, September 28 - October 1, 2011. Proceedings*, volume 6917 of *LNCS*. Springer, 2011.
33. Vincent Rijmen, Deniz Toz, and Kerem Varici. AES Characteristics. In *WCC 2013*, To appear.
34. Serge Vaudenay. Decorrelation: A theory for block cipher security. *J. Cryptology*, 16(4):249–286, 2003.
35. David Wagner. The Boomerang Attack. In Lars R. Knudsen, editor, *FSE*, volume 1636 of *LNCS*, pages 156–170. Springer, 1999.
36. Hongjun Wu. The Hash Function JH, 2011.

A Plateau Characteristics and Our Model

Especially for the AES, the existence of so called *plateau characteristics* (introduced in [21, 22]) is a well studied phenomena. Basically, a given characteristic is plateau if there are only two possible values, 0 and 2^{h-1} for some positive integer h, for the number of pairs following this characteristics. In other words, the number of right pairs following a plateau characteristic with EDP p can be modeled by a random variable Y satisfying

$$Y = \begin{cases} 2^{h-1} & \text{with probability } p2^{n-h} \\ 0 & \text{with probability } (1-p)2^{n-h} \end{cases}.$$

The value of h is called the height of the characteristic[1].

Plateau characteristics for two rounds of AES are well understood. In particular it has been shown that for two rounds plateau characteristics of height up to 5 exist. Recently, in [33] some four-round plateau characteristics with height greater than one have been presented for AES. We are not aware of any results for more than 4 rounds of AES.

Note that a plateau characteristic of height one for a round-reduced variant of the primitive, trivially extends to any number of rounds simply because the number of right pairs never increases as the number of rounds grows. This is not (in general) the case for plateau characteristic of height greater than one.

Clearly, a plateau characteristic does not follow a binomial distribution. Even more, in the case where the height h is greater than one, the binomial distribution clearly underestimates the probability of the characteristic to be fulfilled by 2^{h-1} or more pairs. Thus, in this case our model does not fit as is.

However, *plateau characteristics of height 1* actually do not pose a problem for our model. In this case, the characteristic is never fulfilled by more than one pair. By assuming a binomial distribution in our model, we therefore *overestimate* the probability of having more than one right pair.

Intuitively, plateau characteristics of height greater than one for all rounds of AES (or similar constructions) seem unlikely. However, until now their existence cannot be excluded an, thus, some doubts on the unrestricted applicability of our model remain.

Finally note that our model can tolerate plateau characteristics of height greater than one as long as they are not too frequent. More precisely, assume the existence of A plateau characteristics of height h with EDP below p. In this case the probability that at least one of them is fulfilled by 2^{h-1} pairs is upper bounded by $Ap2^{n-h}$. Thus is A is sufficiently smaller than $\frac{2^{h-n}}{p}$ with good probability all those plateau characteristics are fulfilled by zero pairs and thus do not affect the validity of our bounds.

[1] The "-1" in 2^{h-1} stems from the fact that we consider unordered pairs, i.e. the total number of pairs is $N = 2^{n-1}$ while [22] considers ordered pairs.