# XOR and Non-XOR Differential Probabilities

Philip Hawkes[1] and Luke O'Connor[2]

[1] Qualcomm International,
Suite 410, Birkenhead Point,
Drummoyne, NSW, 2047, Australia
phawkes@qualcomm.com
[2] IBM Research Division
Zurich Research Laboratory
Säumerstrasse 4, Rüschlikon
CH-8803, Switzerland
oco@zurich.ibm.com

**Abstract.** Differential cryptanalysis is a well-known attack on iterated ciphers whose success is determined by the probability of predicting sequences of differences from one round of the cipher to the next. The notion of difference is typically defined with respect to the group operation(s) used to combine the subkey in the round function $F$. For a given round operation $\pi$ of $F$, such as an $S$-box, let $DP_{\otimes}(\pi)$ denote the probability of the most likely non-trivial difference for $\pi$ when differences are defined with respect to $\otimes$. In this paper we investigate how the distribution of $DP_{\otimes}(\pi)$ varies as the group operation $\otimes$ is varied when $\pi$ is a uniformly selected permutation. We prove that $DP_{\otimes}(\pi)$ is maximised with high probability when differences are defined with respect to XOR.

## 1 Introduction

Differential cryptanalysis (DC) is a well-known chosen-plaintext attack based on predicting how certain changes or differences in the plaintext propagate through a cipher. DC was well publicized by Biham and Shamir [3] as a tool for the cryptanalysis of DES-like ciphers. Biham and Shamir defined the *difference* $\Delta(X, X^*)$ between two $n$-bit blocks $X, X^*$ by $\Delta(X, X^*) = X \oplus X^*$ where $\oplus$ denotes the bit-wise exclusive-OR (XOR) operation. To extend the application of DC to other ciphers Lai, Massey and Murphy [14] adapted the definition of differences to $\Delta(X, X^*) = X \otimes (X^*)^{-1}$, where $\otimes$ is an Abelian group operation and $(X^*)^{-1}$ is the group inverse of $X^*$. The choice of difference used to analyse a cipher is usually selected so that the subkey $Z$ is cancelled by the difference operator:

$$\Delta(X, X^*) = \Delta((X \otimes Z), (X^* \otimes Z)) \;=\; X \otimes (X^*)^{-1}. \tag{1}$$

Consequently, the choice of operation used to define differences is typically defined by the group operations(s) used to combine the key into the cipher. Commonly used group operations include XOR $(\mathbb{Z}_2^n, \oplus)$, modular addition $(\mathbb{Z}_{2^n}, \boxplus)$, and modular multiplication $(\mathbb{Z}^*_{2^n+1}, \odot)$, where $(2^n + 1)$ is prime. In general the

J. Stern (Ed.): EUROCRYPT'99, LNCS 1592, pp. 272–285, 1999.

| $n$ | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|
| av. max. $\boxplus$ | 0.2771 | 0.1617 | 0.0919 | 0.0515 | 0.0284 |
| av. max. $\odot$ | 0.2764 | - | - | - | 0.0283 |
| av. max. $\oplus$ | 0.4186 | 0.2487 | 0.1426 | 0.0806 | 0.0443 |

**Table 1.** The average maximum probability for differential approximations to randomly selected bijections $\pi : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2^n$ defined for the operations $\otimes \in \{\boxplus, \oplus, \odot\}$, where $4 \leq n \leq 8$. Note that differences for $\odot$ are only defined for $n$ when $2^n + 1$ is prime.

inputs $x_1, x_2$ to $\odot$ in a cipher will be elements of $\mathbb{Z}_2^n$ rather than $\mathbb{Z}_{2^n+1}^*$, and when evaluating $x_1 \odot x_2$ we first map $x_i$ to $2^n$ if $x_i$ is zero; also $x_1 \odot x_2$ is mapped to zero if it is equal to $2^n$.

The purpose of this paper is to examine how the probability of differential approximations for permutations $\pi$ vary as the group operation $\otimes$ used to define differences is varied. The study of permutations can be justified on two grounds. First, many blocks ciphers make use of permutations: in some cases these permutations are 'small', often referred to as $S$-boxes if implemented as tables, such as in SAFER K-64 [17], TWOFISH [21], CRYPTON [16], E2 [12] and Rijndael [5] (all use 8-bit permutations), while other ciphers use larger permutations such as IDEA [13] (subkey multiplication is equivalent to a 16-bit permutation look-up) and DFC [7] (64-bit permutation). The second reason to study permutations is that a block cipher implements a permutation $\pi$ for any fixed key, and the cipher itself then represents a family of permutations. By studying the properties of permutations we can examine how, for example, permutations generated by an iterative block structure differ from truly random permutations.

Our research was initially motivated by the results presented in Table 1, which shows the average maximum differential approximation to several thousand $n$-bit permutations, $4 \leq n \leq 8$, with respect to the group operations $\oplus$, $\boxplus$ and $\odot$. In all cases $DP_{\otimes}(\pi)$ was maximised for XOR differences. For example, the column for $n = 8$ indicates that the best approximation for 8-bit mappings with respect to $\otimes \in \{\odot, \boxplus\}$ will have a probability between 7/256 and 8/256, while the corresponding probability for XOR differences was between 10/256 and 12/256. Experiments also showed that XOR differences yielded higher probability differentials for the $S$-boxes of DES than differences with respect to $\boxplus$. While this phenomenon is quite likely to be known by some researchers[1], this is the first paper which analyses it mathematically.

We first present our main results and then discuss their implications. We will consider all abelian groups of order $2^n$, and to this end, let $(\mathbb{Z}_2^n, \otimes)$ be an abelian group of order $2^n$ with identity element $I$. For $\alpha, \beta \in \mathbb{Z}_2^n \setminus \{I\}$ and an $n$-bit permutation $\pi$ we define

[1] For example, this observation was stated by M. Dichtl during a seminar presented at Isaac Newton Institute, 1996.

| n | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1024 |
|---|---|---|---|---|---|---|---|---|
| $B_n$ | 4.6 | 7.2 | 11.7 | 20.8 | 34.3 | 60.4 | 108.1 | 195.6 |
| $2 \cdot \lceil B_n \rceil$ | 10 | 16 | 24 | 42 | 70 | 122 | 218 | 392 |

**Table 2.** The values of $B_n = \ln N^2 / \ln \ln N^2$, $N = (2^n - 1)$ for several $n$.

$$DP_{\otimes}(\pi, \alpha, \beta) = \frac{1}{2^n} \cdot \sum_{\substack{X, X^* \\ \Delta(X, X^*) = \alpha}} [\, \Delta(\pi(X), \pi(X^*)) = \beta \}\, ]$$

where $[\cdot]$ is a predicate that evaluates to 0 or 1. Thus $DP_{\otimes}(\pi, \alpha, \beta)$ is the probability that an input difference of $\alpha$ leads to an output difference of $\beta$ in $\pi$ when differences defined with respect to $\otimes$. Further, we define $DP_{\otimes}(\pi) = \max_{\alpha, \beta \neq I} DP_{\otimes}(\pi, \alpha, \beta)$ to be the highest probability difference in $\pi$ with respect to $\otimes$. One of the main results of this paper is to prove that asymptotically, for uniformly selected $\pi$,

$$\Pr\left( \frac{n \ln 2}{2^{n-1} \ln n} \leq DP_{\otimes}(\pi) < \frac{n}{2^{n-1}} \right) \sim 1, \qquad \otimes = \oplus, \tag{2}$$

$$\Pr\left( DP_{\otimes}(\pi) < \frac{n \ln 2}{2^{n-1} \ln n} \right) \sim 1, \qquad \otimes \in \{ \boxplus, \odot \}. \tag{3}$$

Equivalently, the fraction of $n$-bit permutations that do not satisfy the bounds of (2) and (3) tends to 0 and $n$ increases. Our results concentrate on a comparison between $\otimes = \oplus$ and $\otimes \in \{ \boxplus, \odot \}$, since the latter two group operations are the most pertinent to cryptography. The $(n \ln 2)/(2^{n-1} \ln n)$ term in (2) and (3) is derived as an asymptotic estimate of $2B_n/2^n$ where $B_n = \ln N^2 / \ln(\ln N^2))$ and $N = 2^n - 1$ is the number of non-trivial differences. For smaller $n$, $2B_n$ can be used in (2) and (3), and some relevant values of $B_n$ are given in Table 2. For 8-bit permutations the critical value is $B_8 = 10$, meaning that XOR approximations are likely[2] to occur with probability at least $10/256$ while approximations based on $\otimes \in \{ \boxplus, \odot \}$ with probability less than $10/256$. The general conclusion is that it is very likely that selecting a permutation $\pi$ at random will yield higher probability XOR difference approximations than differences defined with respect to the groups $\boxplus$ and $\odot$. [3]

The bounds of (2) and (3) indicate that with high probability the best DC XOR approximation to a 64-bit permutation lies in the interval $[2^{-58.6}, 2^{-57}]$,

[2] We note that the authors of TWOFISH were able to find 8-bit permutations with best XOR difference approximation of at most $10/256$ in 'a few tens of hours' [21, p.24]. These permutations were composed to form the basis of the $S$-boxes for TWOFISH, where for a majority of the keys the best XOR approximation has probability $12/256$.

[3] We note that it is always possible to pick a 'cooked' permutation $\pi$ for which XOR differences have lower probability than $\otimes \in \{ \boxplus, \odot \}$ differences, such as $\pi(x) = x \oplus c$ for some group element $c$. We simply assert that this event is unlikely to happen if $\pi$ is selected randomly or in some unbiased manner.

while for a 128-bit permutation the interval is $[2^{-121.9}, 2^{-120}]$. Thus if we assume that 48-round DES acts as random 64-bit permutation, the best XOR approximation will occur with much higher probability than suggested by extending the 2-round iterative characteristic used for the DC of 16-round DES. While we acknowledge that it may be computationally infeasible to find such a high probability characteristic, the bounds of (2) and (3) strongly suggest that far more probable DC approximations are available than indicated by the round-by-round approximation approach based on characteristics.

We are hesitant to apply our results in general to existing ciphers, say by changing $\oplus$ operations to $\boxplus$ operations and claiming improved security against DC. This is certainly not the case for DES [4]. We believe that to fully take advantage of our results would require the design of a new cipher, and this is not the subject of this paper. We hope that our present results will form the basis for further research into the most appropriate group operation(s) to be used in the design and analysis of block ciphers against DC. We note however that Adams [1] has already used our results to suggest the security of the CAST-256 algorithm. We also note that in general the designer cannot force the cryptanalyst to use differences defined with respect to a given group operation $\otimes$. A case in point is a DC of RC5 [11] where the natural choice of difference was based on $\boxplus$, but differences with respect to $\oplus$ were used regardless. On the other hand, XOR differences give high probability approximations to the two $S$-boxes used in SAFER K-64, but the use of other non-XOR operations such as the Pseudo Hadamard Transform appears to have successfully thwarted on DC based on XOR differences alone.

It remains now to prove (2) and (3). As a first step we determine that the distribution of $DP_{\otimes}(\pi, \alpha, \beta)$, asymptotically follows the Poisson distribution $\Pr(X = t) = e^{-\mu} \cdot \mu^t / t!$, for $t \geq 0$. When both group elements $\alpha, \beta$ have order 2, the Poisson parameter is $\mu = 1/2$, while it is $\mu = 1$ for any other pair of elements with orders both distinct from 2. Note that all elements of $(\mathbb{Z}_2^n - \{0\}, \oplus)$ have order 2, while almost all elements of $(\mathbb{Z}_2^n - \{I\}, \otimes)$, $\otimes \neq \oplus$, have order greater than 2, which will be shown to cause the higher XOR approximations. Also similar comments apply if $\alpha$ is a difference with respect to $\otimes_1$, and $\beta$ is a difference with respect to $\otimes_2$, Such differences have been called a quasi-differentials [18], and naturally arise in the DC of SAFER [17] which uses both $\oplus$ and $\boxplus$ to mask the inputs and outputs to its $S$-boxes.

The upper bound in (2) is from [19], while given $Y_k = \sum_{\alpha,\beta} \Pr(DP_{\otimes}(\pi, \alpha, \beta) = k)$, the upper bound in (3) can be proven directly from $\Pr(DP_{\otimes}(\pi) \leq t) \leq (1 - \sum_{k \geq t} \mathbf{E}[Y_k])$ when $Y_k = o(2^{2n})$. The lower bound in (2) is harder to prove. Note that $Y_k$ defined above is the expected number of entries in the difference table of size $k$. Our approach is to find a value of $k$ for which $\mathbf{E}[Y_k] \geq n$ and $\mathbf{Var}[Y_k] \sim n$, from which it follows via Chebychev's inequality that an entry of size $k$ exists with probability tending to 1 with $n$. As it turns out, $k = B_n$ satisfies these conditions. Even though we work with expectations we note that that bounds in (2) and (3) are not expectations or for the average case.

The paper is set out as follows. In §2 we introduce notation and reduce the problem of enumerating $2^n \cdot DP_{\otimes}(\pi, \alpha, \beta)$ to a counting problem on graphs. This counting problem is combined with the inclusion-exclusion principle to obtain the distribution of probabilities for a differential approximation. The distribution of values for individual entries are shown to be asymptotically Poisson in §2. In §3 the bound given in (2) and (3) are proven. Our conclusions are presented in §4 and several proofs are delegated to the appendix in §5.

## 2 An Equivalent Graph Theory Problem

We let $\Pi^{(n)}$ denote the set of $n$-bit permutations, and write $\pi \in_R \Pi^{(n)}$ to denote a uniformly selected $n$-bit permutation. The problem of determining the distribution of $DP_{\otimes}(\pi)$ can be considered as an enumeration problem: count the number of edge-preserving mappings between two appropriately defined directed graphs, given below. Recall that the set of $n$-bit blocks is denoted $\mathbb{Z}_2^n$ and can represented by the set $\{0, 1, \ldots, 2^n - 1\}$.

**Definition 1.** For a group $(\mathbb{Z}_2^n, \otimes)$ of order $2^n$ and a non-trivial (non-identity) difference $\alpha \in \mathbb{Z}_2^n$ there is an associated directed graph $D_\alpha = (V, E_\alpha)$, $|V| = 2^n$, where each vertex $v \in V$ has a unique label $l(v) \in \mathbb{Z}_2^n$. Then any group element $X$ is uniquely associated with the vertex $u \in V$ such that $l(u) = X$. The directed edge set of $D_\alpha$ is defined as $E_\alpha = \{(u, v) \mid l(u) \otimes (l(v))^{-1} = \alpha\}$, meaning that $(u, v)$ is an edge when $X = l(u)$ and $v = l(X^*)$ and $\Delta(X, X^*) = \alpha$. We call $D_\alpha$ the *difference graph* of $\alpha$ with respect to $\otimes$. □

As a result of the group property, every vertex of $D_\alpha$ and $D_\beta$ has indegree and outdegree one. Consequently, the arcs of $D_\alpha$ and $D_\beta$ form cycles. Further, $D_\alpha$ consists of $\frac{2^n}{\text{ord } \alpha}$ labeled disjoint cycles of length ord $\alpha$, which follows from Lagrange's Theorem since the cycles correspond to cosets. Let $D_\alpha = (V, E_\alpha)$ and $D_\beta = (V, E_\beta)$ be the difference directed graphs representing any two differences $\alpha, \beta \in \mathbb{Z}_2^n$. For a permutation $\pi \in \Pi^{(n)}$ we define

$$d_{\otimes,\pi}(D_\alpha, D_\beta) = \#\left\{(u, v) \in E_\alpha \mid (u^*, v^*) \in E_\beta,\ l(u^*) = \pi(l(u)), l(v^*) = \pi(l(v))\right\}.$$

Thus $2^n \cdot DP_{\otimes}(\pi, \alpha, \beta) = d_{\otimes,\pi}(D_\alpha, D_\beta)$, and this value depends on the number of edges mapped between the two distance graphs.

*Example 2. Consider* $(\mathbb{Z}_2^3, \boxplus)$*, the group of addition modulo 8. The directed graphs* $D_1, D_2$ *representing the differences* $\Delta(X, X^*) = 1$ *and* $\Delta(X, X^*) = 2$ *are shown in Figure 1. Notice that the arcs of* $D_1$ *and* $D_2$ *form cycles of length 8 and 4 respectively, as* ord $1 = 8$ *and* ord $2 = 4$ *with respect to* $\boxplus$*. Let* $\pi \in \Pi^{(3)}$ *be the permutation* $(3, 0, 7, 1, 2, 5, 4, 6)$ *where* $\pi(0) = 3, \pi(1) = 0, \pi(2) = 7$ *and so on. Then the only arcs of* $D_1$ *mapped by* $\pi$ *to arcs of* $D_2$ *are the arcs labeled by* $(3, 2)$ *and* $(7, 6)$ *of* $D_1$ *which are mapped to the arcs labeled by* $(1, 7)$ *and* $(6, 4)$ *respectively of* $D_2$*. Consequently* $DP_{\boxplus}(\pi, 1, 2) = d_{\boxplus,\pi}(D_1, D_2) = 2$. □

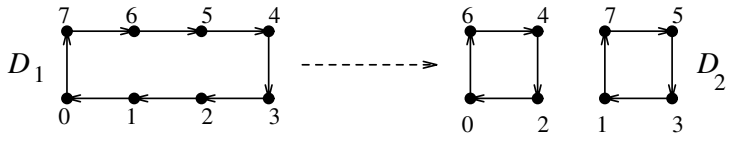


**Fig. 1.** The directed graphs $D_1$ and $D_2$ representing the two differences $\Delta(X, X^*) = 1$ and $\Delta(X, X^*) = 2$ using the 3-bit $\boxplus$ operation to define the differences.

**Theorem 3.** For any Abelian group $G$ and elements $\alpha, \beta, \in G$, the probability $\Pr(2^n \cdot DP_{\otimes}(\pi, \alpha, \beta) = t)$ only depends on $t$, ord $G = \#G$, $a = $ ord $\alpha$ and $b =$ ord $\beta$. For $a = 2^r$, $b = 2^s$, $1 \le r \le n$, $1 \le s \le n$, and $0 \le t \le 2^n$, define

$$p_t(\#G, a, b) \stackrel{\text{def}}{=} \Pr\left(2^n \cdot DP_{\otimes}(\pi, \alpha, \beta) = t \mid \pi \in_R \Pi^{(n)}\right). \tag{4}$$

Our main goal is to show that $p_t(\#G, a, b)$ asymptotically follows the Poisson distribution. To show this we need to consider the the distribution of (element) orders in $(\mathbb{Z}_2^n, \otimes)$. In the group $(\mathbb{Z}_2^n, \oplus)$, all the nonzero elements have order 2, and the resulting directed graphs $D_\alpha$ consist of $2^{n-1}$ cycles of length 2. However, in the group $(\mathbb{Z}_2^n, \boxplus)$ there are $2^{a-1}$ elements of order $2^a$, $1 \le a \le n$, and the identity (0) has order one. For $2^n + 1$ prime, the groups $(\mathbb{Z}_2^n, \odot)$ and $(\mathbb{Z}_2^n, \boxplus)$ are isomorphic, and thus have the same distribution of orders.

**Corollary 4.** Let $\otimes \in \{\boxplus, \odot\}$. Then there are $2^{2a-2}$ pairs of group elements $\alpha, \beta$ for which ord $\alpha =$ ord $\beta = 2^a$, $1 \le a \le n$, and $2^{a+b-2}$ pairs for which $\{\text{ord } \alpha, \text{ord } \beta\} = \{2^a, 2^b\}$, $1 \le a < b \le n$.

To bound the value of $DP_{\otimes}(\pi)$, we need only determine $p_t(2^n, a, b)$ for $a = 2^r$, $b = 2^s$, $1 \le r \le n$, $1 \le s \le n$, and $0 \le t \le 2^n$, and apply Corollary 4. We now cast determining $p_t(2^n, a, b)$ to an enumeration problem in terms of the *inclusion-exclusion principle (IEP)* (see for example Hall [9]).

Let $\alpha$ and $\beta$ be elements of $(\mathbb{Z}_2^n, \otimes)$, and let $D_\alpha = (V, E_\alpha)$ and $D_\beta = (V, E_\beta)$ be their respective (difference) graphs. For each edge $uv \in E_\alpha$ define $A_{uv}$ as $A_{uv} = \left\{\pi \in \Pi^{(n)} \mid (\pi(u), \pi(v)) \in E_\beta\right\}$, which is the set of permutations $\pi$ that preserve the edge $uv$ of $D_\alpha$ in $D_\beta$. Then, by the inclusion-exclusion principle, the number of permutations $\pi$ that preserve exactly $t$ edges from $D_\alpha$ in $D_\beta$ is

$$P_t = \sum_{i=0}^{j-t} (-1)^i \binom{t+i}{i} S_{t+i}, \qquad S_k = \sum_{\substack{\mathcal{Y} \subseteq E_\alpha \\ |\mathcal{Y}| = k}} \left| \bigcap_{uv \in \mathcal{Y}} A_{uv} \right|, \tag{5}$$

and it follows that $p_t(2^n, a, b) = P_t/(2^n!)$. In the case of XOR differences ($\otimes = \oplus$) it is known [19] that

$$P_{2t} \sim \binom{2^{n-1}}{t}^2 \cdot 2^t \cdot t! \cdot \frac{(2^{n-1} - t)!}{e^{1/2}}. \tag{6}$$

In this case we can immediately prove that $p_{2t}(2^n, 2, 2)$ is asymptotically Poisson distributed.

**Lemma 5.** If $t \in o(2^{n/2})$ as $n \to \infty$, then

$$p_{2t}(2^n, 2, 2) \;=\; \frac{e^{-\frac{1}{2}}}{2^t \cdot t!} \cdot (1 + O((t+1)^2/2^n)).$$

*Proof.* From [20] we have that

$$p_{2t}(2^n, 2, 2) = \frac{1}{2^n!} \cdot \binom{2^{n-1}}{t}^2 \cdot 2^t \cdot t! \cdot \Phi(2^{n-1} - t) \tag{7}$$

where $\Phi(2^{n-1} - t) = (2^n - 2t)! \cdot e^{-1/2} \cdot (1 + O(1/(2^n - 2t)))$. If $t \in o(\sqrt{2^n})$ then it can be shown that

$$\binom{2^{n-1}}{t} = \frac{(2^{n-1})^t}{t!} \cdot (1 + O(t^2/2^{n-1})) \;=\; \frac{(2^n)^t}{2^t \cdot t!} \cdot (1 + O(t^2/2^n)),$$
$$\Rightarrow \binom{2^{n-1}}{t}^2 = \left(\frac{(2^n)^t}{2^t \cdot t!}\right)^2 \cdot (1 + O(t^2/2^{n-1})),$$

as $(1 + O(t^2/2^n))^2 = 1 + 2 \cdot O(t^2/2^n) + O(t^4/2^{2n}) = 1 + O(t^2/2^n)$. Substituting these approximations into (7) yields the theorem. □

In this case determining an exact expression for $P_t$ is assisted by the fact that ord $\alpha =$ ord $\beta = 2$, and the sets $A_{uv}$ are 'independent' in the sense that $uv$ is the only edge incident on $u$ and $v$. For a general group operation $\otimes \neq \oplus$, most groups elements $\alpha$ will have ord $\alpha > 2$, and hence induce a difference graph for which there exist sets $A_{u_1v_1}, A_{u_2,v_2}$ and $v_1 = u_2$. Dependence between the $A_{uv}$ sets considerably complicates the expressions for $P_t$. The following expression for $p_t(2^n, 2, 4)$ taken from [10], which also gives an involved formula for $p_t(2^n, 4, 4)$.

**Lemma 6.** For $n \geq 2$, and $0 \leq t \leq 2^{n-1}$,

$$p_t(2^n, 2, 4) = \frac{1}{2^n} \cdot \sum_{i=0}^{2^{n-1}-t} (-1)^k \binom{t+i}{i} S_{t+i},$$

where for $0 \leq k \leq 2^{n-1}$,

$$S_k = \left( \sum_{i=\lceil k/2 \rceil}^{\min(k, 2^{n-2})} \binom{2^{n-2}}{i} \cdot \binom{i}{k-i} \cdot 2^{3i} \right) \cdot \binom{2^{n-1}}{k} \cdot k! \cdot (2^n - 2k)!. \tag{8}$$

For general $a, b > 4$ the expression for $S_k = S_k^{(n)}(a, b)$ becomes increasing difficult to determine exactly, and we therefore consider an asymptotic approximation. We denote $\pi(\mathcal{Y}) = \{(u^*, v^*) \mid l(u^*) = \pi(l(u)), l(v^*) = \pi(l(v)), (u, v) \in \mathcal{Y}\}$, so that we can represent $\cap_{uv \in \mathcal{Y}} A_{uv} = \{\pi \mid \pi(\mathcal{Y}) \subseteq E_\beta\}$. Observe that $S_k$ is

defined in terms of *preserved edges* , but it may be further decomposed into terms of *preserved vertices* . Observe that a set of $k$ edges is incident on at least $k$ vertices (a cycle) and at most $2k$ vertices (disjoint edges). Let $p(\mathcal{Y})$ be the number of vertices which are incident to the edges of $\mathcal{Y}$, where $k \leq p(\mathcal{Y}) \leq 2k$. For $k \leq j \leq 2k$, define

$$\phi(k,j) \;=\; \sum_{\substack{\mathcal{Y} \subseteq E_\alpha, \\ |\mathcal{Y}|=k,\ p(\mathcal{Y})=j}} \left|\{\pi \mid \pi(\mathcal{Y}) \subseteq E_\beta\}\right|,$$

such that $S_k$ can be expressed as $S_k = \sum_{j=k}^{2k} \phi(k,j)$. As it turns out, $S_k \sim \phi(k,2k)$, meaning that $S_k$ is dominated by the term mapping disjoint edges $D_\alpha$ to edges to disjoint edges in $D_\beta$. In [10] it was proven that for $k = o(2^{n/2})$, $\phi(k,2k) = \frac{N!}{k!} \cdot (1 + o(1))$, which leads to the next theorem.

**Theorem 7.** Suppose that $n \geq 0$, $a = 2^r$, $b = 2^s$, $1 \leq r \leq n$ and $2 \leq s \leq n$. Then $S_k = \frac{2^n!}{k!} \cdot \left(1 + O\left(k^2/2^n\right)\right)$ for $k \in o(2^{n/2})$ as $n \to \infty$.

The proof of Theorem 7 is involved and lengthy, and the reader is referred to [10] for details. It still remains to derive an expression for general $p_t(2^n, a, b)$ from $P_t$ and $S_t$. Our results are based on the following adaptation of a theorem by Bender [2].

**Theorem 8.** Suppose there is a function $A(n)$ and a value $\lambda \geq 0$, such that

$$S_k = A(n) \cdot \frac{\lambda^k}{k!} \cdot (1 + O(f(k)/g(n))),$$

and $f(k) \in o(g(n))$ for $0 \leq k \leq l(n)$, where $l(n)$ goes to infinity with $n$. Let $j = l(n) - t$ and define $f^*(t) = \sum_{i=0}^{j-1} f(t+i) \cdot \lambda^i / i!$. If $m(n)$ is a function such that $l(n) - m(n)$ goes to infinity with $n$, then for each $t$, $0 \leq t \leq m(n)$,

$$P_t = A(n) \cdot e^{-\lambda} \cdot \frac{\lambda^t}{t!} \cdot (1 + O(f^*(t)/g(n))). \tag{9}$$

By applying this theorem we are able to show that $p_t(2^n, a, b)$ is asymptotically Poisson.

**Corollary 9.** Provided $a > 2$ or $b > 2$, and $t \in o(2^{n/2}/2)$,

$$p_t(2^n, a, b) = \frac{e^{-1}}{t!} \cdot (1 + O((t+1)^2/2n)).$$

*Proof.* Theorem 7 proves that $S_k = 2^n!/k! \cdot (1 + O(k^2/2^n))$ for $k = o(2^{n/2})$. Theorem 8 can now be applied with $A(n) = 2^n!$, $\lambda = 1$, $l(n) = o(2^{n/2})$, $f(k) = k^2$, $g(n) = 2^n$, $f^*(t) = O((t+1)^2)$ and $m(n) = o(2^{n/2})$. □

The main result of this section can now be stated, which we call the *asymptotic Poisson approximation (PA)* to $DP_\otimes(\pi, \alpha, \beta)$.

**Theorem 10.** Let $(\mathbb{Z}_2^n, \otimes)$ be an Abelian group of order $2^n$ and $\alpha, \beta \in \mathbb{Z}_2^n$ be non-trivial differences. If $\pi \in_R \Pi^{(n)}$ and $t = o(2^{n/2})$

$$\begin{array}{rl} \Pr\left(DP_{\otimes}(\pi,\alpha,\beta) = \frac{t}{2^{n-1}}\right) \sim e^{-\frac{1}{2}} \cdot \left(\frac{1}{2}\right)^t / t! & \text{if ord } \alpha = \text{ord } \beta = 2, \\ \Pr\left(DP_{\otimes}(\pi,\alpha,\beta) = \frac{t}{2^n}\right) \sim e^{-1}/t! & \text{otherwise.} \end{array}$$

Let $\mathbf{E}[X]$, $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ and $\sigma[X] = \sqrt{\mathbf{Var}[X]}$ denote the expectation, variance and standard deviation of the random variable $X$. It is known that if the distribution of values for $X$ is Poisson, then $\mathbf{Var}[X] = \mathbf{E}[X] = \mu$. Then, for example, a little algebraic manipulation reveals that the distribution of values for $DP_{\otimes}(\pi,\alpha,\beta)$ has $\mathbf{E}[DP_{\otimes}(\pi,\alpha,\beta)] \sim 1/2^n$ and $\sigma[DP_{\otimes}(\pi,\alpha,\beta)] \sim \eta/2^n$, where $\eta = \sqrt{2}$ if ord $\alpha$ = ord $\beta = 2$ and $\eta = 1$ otherwise. This indicates that the probabilities for a differential approximation $\Delta X = \alpha \to \Delta\pi(X) = \beta$ where ord $\alpha$ = ord $\beta = 2$ are distributed $\sqrt{2}$ times as far from $1/2^n$ as the probabilities for other differential approximations. Consequently, differential approximations for which ord $\alpha$ = ord $\beta = 2$ are more likely to have higher probabilities.

## 3 Bounding the Maximum Difference Table Entry

In this section we use the PA to obtain bounds on $DP_{\otimes}(\pi)$ that hold asymptotically with probability one. The distribution of differences with respect to $\oplus$ is approximated using a Poisson distribution with $\mu = \frac{1}{2}$, as all non-trivial elements have order two. The distribution of differences for $\otimes \in \{\boxplus, \odot\}$ is approximated using a Poisson distribution with $\mu = 1$, as there is only one pair $(\alpha, \beta)$ with ord $\alpha$ = ord $\beta = 2$. We determine the expectation and variance of $\theta_t(\otimes, \pi)$, defined to be

$$\theta_t(\otimes,\pi) = \frac{1}{(2^n-1)^2} \sum_{\alpha \neq I} \sum_{\beta \neq I} [2^n \cdot DP_{\otimes}(\alpha,\beta) = t] \tag{10}$$

which is the fraction input/output differences that map exactly $t$ pairs, $0 \leq t \leq 2^n$.

**Corollary 11.** For $\pi \in_R \Pi_{2^n}$, $\mathbf{E}[\theta_{2t}(\oplus,\pi)] \sim e^{-\frac{1}{2}} \cdot \left(\frac{1}{2}\right)^t / t!$ and $\mathbf{E}[\theta_t(\otimes,\pi)] \sim e^{-1}/t!$ uniformly for $t = o(2^{n/2})$ where $\otimes \in \{\boxplus, \odot\}$.

This information is sufficient for obtaining upper bounds on $DP_{\otimes}(\pi)$ for $\otimes \in \{\oplus, \boxplus, \odot\}$. However, to obtain our lower bound on the maximum entry in differences tables with respect to $\oplus$, the variance of $\theta_{2t}(\oplus,\pi)$ is required. We have not attempted to determine the variance in $\theta_t(\otimes,\pi)$ for $\otimes \in \{\boxplus, \odot\}$ as the counting problem is very complex, and this variance is not required for the results of this paper. See [10] for a proof of the next lemma.

**Lemma 12.** For $\pi \in_R \Pi_{2^n}$ and $t = o(2^{n/2})$

$$\mathbf{Var}[\theta_{2t}(\oplus,\pi)] \sim \frac{1}{(2^n-1)^2} \cdot e^{-\frac{1}{2}} \cdot \left(\frac{1}{2}\right)^t / t! \cdot \left(1 - e^{-\frac{1}{2}} \cdot \left(\frac{1}{2}\right)^t / t!\right).$$

For nontrivial $\alpha, \beta$ define $\Psi^{(t)}_{\alpha,\beta}$, $0 \leq t \leq 2^{n-1}$, where $\Psi^{(t)}_{\alpha,\beta} = 1$ if $2^n \cdot DP_{\oplus}(\pi, \alpha, \beta) = 2t$ and $\Psi^{(t)}_{\alpha,\beta} = 0$ otherwise. It follows that $\Psi^{(t)} = \sum_{\alpha,\beta \neq I} \Psi^{(t)}_{\alpha,\beta} = (2^n - 1)^2 \cdot \theta_{2t}(\oplus, \pi)$. Note that $\mathbf{E}[\Psi^{(t)}] = (2^n - 1)^2 \cdot \mathbf{E}\left[\theta_{2t}(\oplus, \pi)\right] \sim (2^n - 1)^2 \cdot e^{-\frac{1}{2}} \cdot \frac{1}{2}^t / t!$ for $t = o(2^{n/2})$. Similarly,

$$\mathbf{Var}\left[\Psi^{(t)}\right] = (2^n - 1)^4 \cdot \mathbf{Var}[\theta_{2t}(T_{\oplus,\pi})] \tag{11}$$

$$\sim (2^n - 1)^2 \cdot \frac{e^{-\frac{1}{2}}}{2^t \cdot t!} \cdot \left(1 - \frac{e^{-\frac{1}{2}}}{2^t \cdot t!}\right) \tag{12}$$

$$\sim \mathbf{E}[\Psi^{(t)}] \cdot \left(1 - \frac{e^{-\frac{1}{2}}}{2^t \cdot t!}\right) \tag{13}$$

drawing on the result of Lemma 12. Define $B_n$ as $B_n = \ln N^2 / \ln \ln N^2$, where $N = (2^n - 1)$, and observe that the Poisson approximation (Corollary 10) holds for $0 \leq t \leq 2B_n$ since $2B_n = o(2^{n/2})$. The next two lemmas are proved using the previous variance calculations in the Appendix.

**Lemma 13.** If $\pi \in_R \Pi^{(n)}$, then $\Pr\left(B_n/2^{n-1} \leq DP_{\oplus}(\pi) < n/2^{n-1}\right) \sim 1$.

**Lemma 14.** If $\pi \in_R \Pi^{(n)}$, then $\Pr\left(DP_{\otimes}(\pi) < B_n/2^{n-1}\right) \sim 1$, where $\otimes \in \{\boxplus, \odot\}$.

Asymptotically $B_n$ tends to $(2n \ln 2)/\ln n$, which when applied to the previous two lemmas, determines the bounds given in (2) and (3). Statements concerning the best differential approximation of a randomly selected permutation can now be made. For example, the probability of the best approximation with respect to XOR differences is in the range $[2^{-58.6}, 2^{-57}]$ for a random 64-bit permutation and in the range $[2^{-121.9}, 2^{-120}]$ for a random 128-bit approximation. The values $2^{-58.6}$ and $2^{-121.9}$ are also upper bounds on the probability of approximations with respect to $\otimes \neq \oplus$ for random 64-bit and 128-bit permutations respectively. Further bounds on the maximum entry can be obtained for difference tables with respect to other group operations, and these bounds will rely primarily on the fraction of entries in the difference table for which both elements have order 2.

Finally, Lemma 13 and Lemma 14 combine to confirm our initial observation that in general XOR differences yield higher probability approximations than differences with respect to modular addition and modular multiplication.

**Corollary 15.** If $\pi \in_R \Pi^{(n)}$, then $\Pr\left(DP_{\oplus}(\pi) > DP_{\otimes}(\pi)\right) \sim 1$, for $\otimes \in \{\boxplus, \odot\}$.

## 4 Conclusion

We have shown that with high probability, XOR differences yield better differential approximations than differences with respect $\otimes \in \{\boxplus, \odot\}$. Furthermore, we determined asymptotic approximations to the difference distribution of three

group operations $\otimes \in \{\oplus, \boxplus, \odot\}$, and bound the probability of the most likely difference. Further bounds on the maximum entry can be obtained for difference tables with respect to other group operations, and these bounds will rely primarily on the fraction of entries in the difference table for which both elements have order 2. The Poisson approximation (Corollary 10) can also be applied to quasi-differentials and the maximum probability can be similarly bounded.

We have concentrated on the three groups defined by $\oplus$, $\boxplus$ and $\odot$, but the other groups can be considered using the same analysis. The Poisson approximation of $DP_{\otimes}(\alpha, \beta)$ holds for all group elements with order at least 2. On the other hand, the bounds on $DP_{\otimes}(\pi)$ depend on the distribution of group elements. Bounding $DP_{\otimes}(\pi)$ for a given group $(G, \otimes)$ requires knowledge of how element orders are distributed within $G$. Our results in this paper are based on all non-identity elements of $(\mathbb{Z}_2^n, \oplus)$ having order 2, and the element orders of $(\mathbb{Z}_{2^n}, \boxplus)$ and $(\mathbb{Z}_{2^n+1}^*, \odot)$ being determined as in Corollary 4.

The distribution of entries in difference tables has previously been predicted using a "balls-in-bins" model [15], summarized as follows. In modeling differences tables with respect to XOR, we let the "balls" represent the *unordered* pairs of difference $\alpha$ and let the "bins" represent the possible non-trivial output differences. If the $2^{n-1}$ input pairs of input difference $\alpha$ (the "balls") can be allocated randomly and independently to any of the $(2^n - 1)$ "bins", then the resulting distribution approaches a Poisson distribution with parameter $\mu = \frac{2^{n-1}}{2^n-1} \sim \frac{1}{2}$. In modeling differences tables with respect to $\otimes \neq \oplus$, we let the "balls" represent the *ordered* pairs of difference $\alpha$ and let the "bins" represent the possible non-trivial output differences. If the input pairs of input difference $\alpha$ (the "balls") can be allocated randomly and independently to any of the $(2^n - 1)$ "bins", then the resulting distribution approaches a Poisson distribution with parameter $\mu = \frac{2^n}{2^n-1} \sim 1$. Our results add validity to the "balls-in-bins" approach for predicting $DP_{\otimes}(\pi)$.

Acknowledgments

We would like to thank the diligent referees for their comments on this work, especially Eli Biham.

## 5 Appendix

**Lemma 13** If $\pi \in_R \Pi^{(n)}$, then $\Pr\left(B_n/2^{n-1} DP_{\oplus}(\pi) < n/2^{n-1}\right) \sim 1$, where $\otimes \in \{\boxplus, \odot\}$.

*Proof.* O'Connor [19] proved that $\Pr\left(DP_{\oplus}(\pi) \geq n/2^{n-1}\right) = o(1)$. Denote $\Psi = \Psi^{(B_n)}$, and observe that $\mathbf{Var}[\Psi] \sim \mathbf{E}[\Psi]$ as $B_n$ increases with $n$. Chebychev's inequality (see for example [6]) is applied to show that

$$\Pr(DP_{\oplus}(\pi) < 2B_n) \leq \Pr(\Psi = 0) \leq \Pr\left(|\Psi - \mathbf{E}[\Psi]| \geq \mathbf{E}[\Psi]\right) \leq \frac{\mathbf{Var}[\Psi]}{(\mathbf{E}[\Psi])^2}$$

which is asymptotic to $\frac{1}{\mathbf{E}[\Psi]}$. The expected number of entries $2B_n$ in the differences tables with respect to $\oplus$ is equal to

$$\mathbf{E}[\Psi] \;=\; (2^n-1)^2 \cdot \mathbf{E}[\theta_{2B_n}(\oplus,\pi)] \;\sim\; N^2 \cdot \frac{e^{-\frac{1}{2}}}{2^{B_n} \cdot B_n!}$$

By applying Stirling's formula for $n!$ (see, for example [8, page 213]),

$$B_n! \sim \left(\frac{B_n}{e}\right)^{B_n} \cdot \sqrt{2\pi B_n} \;=\; \frac{(\ln N^2)^{\ln N^2/\ln\ln N^2}}{(e\cdot\ln\ln N^2)^{B_n}} \cdot \sqrt{2\pi B_n},$$

where $(\ln N^2)^{\ln N^2/\ln\ln N^2} = (e^{\ln\ln N^2})^{\ln N^2/\ln\ln N^2} = e^{\ln N^2} = N^2$. Consequently,

$$\begin{aligned}
\Pr(DP_{\oplus}(\pi) < 2B_n) &\le \frac{1}{\mathbf{E}[\Psi]} && (14)\\
&\sim \frac{1}{N^2} \cdot \frac{2^{B_n}}{e^{-\frac{1}{2}}} \cdot \frac{N^2 \cdot \sqrt{2\pi B_n}}{(e\cdot\ln\ln N^2)^{B_n}} && (15)\\
&= e^{\frac{1}{2}} \cdot \frac{\sqrt{2\pi B_n}}{((e/2)\cdot\ln\ln N^2)^{B_n}} \;=\; o(1), && (16)
\end{aligned}$$

as $\sqrt{2\pi B_n} = o\left(((e/2)\cdot\ln\ln N^2)^{B_n}\right)$. Therefore, the probability that the maximum entry is either less than $2B_n$ or greater than or equal to $2n$ is $o(1)$, and the lemma is proved. □

**Lemma 14** If $\pi \in_R \Pi^{(n)}$, then $\Pr\left(DP_{\otimes}(\pi) < B_n/2^{n-1}\right) \sim 1$, where $\otimes \in \{\boxplus, \odot\}$.

*Proof.* Assume $\otimes \neq \oplus$. Let $\Omega^{(t)} = (2^n-1)^2 \cdot \theta_t(\otimes\pi)$ denote the number of entries $t$ in the differences table with respect to $\otimes$, and in particular denote $\Omega = \Omega^{(2B_n)}$. Recall that $\mathbf{E}[\Omega] = (2^n-1)^2 \cdot \mathbf{E}[\theta_{2B_n}(\otimes,\pi)] \sim N^2 \cdot e^{-1}/(2B_n)!$. By applying Stirling's formula for $n!$,

$$\begin{aligned}
(2B_n)! &\sim \left(\frac{2B_n}{e}\right)^{2B_n} \cdot \sqrt{2\pi\cdot(2B_n)} \;=\; \left(\frac{2\ln N^2}{e\cdot\ln\ln N^2}\right)^{2B_n} \cdot 2\sqrt{\pi B_n} && (17)\\
&= \frac{(\ln N^2)^{2\ln N^2/\ln\ln N^2}}{((e/2)\cdot\ln\ln N^2)^{2B_n}} \cdot 2\sqrt{\pi B_n}, && (18)
\end{aligned}$$

where $(\ln N^2)^{2\ln N^2/\ln\ln N^2} = (e^{\ln\ln N^2})^{2\ln N^2/\ln\ln N^2} = e^{2\ln N^2} = N^2$. Thus

$$\begin{aligned}\mathbf{E}[\Omega] &\sim N^2 \cdot \frac{e^{-1}}{N^4} \cdot \left((e/2)\cdot\ln\ln N^2\right)^{2\ln N^2/\ln\ln N^2} \cdot \frac{1}{2\sqrt{\pi B_n}} \\ &= \frac{e^{-1}}{2\sqrt{\pi B_n}} \cdot \left(\frac{\left((e/2)\cdot\ln\ln N^2\right)^{2/\ln\ln N^2}}{(N^2)^{1/\ln N^2}}\right)^{\ln N^2} \\ &= \frac{e^{-1}}{2\sqrt{\pi B_n}} \cdot \left(\underbrace{\frac{\left((e/2)\cdot\ln\ln N^2\right)^{2/\ln\ln N^2}}{e}}_{y(N)}\right)^{\ln N^2},\end{aligned}$$

and we can show that $y(N) \le 1$. Therefore,

$$\mathbf{E}[\Omega] \sim \frac{e^{-1}}{2\sqrt{\pi B_n}} \cdot y(N)^{\ln N^2} = o(1)$$

as $B_n$ increases with $n$. Now, for $2B_n = o(2^{n/2})$, $\mathbf{E}[\Omega^{(t)}] \le \mathbf{E}[\Omega]/(2B_n)^{t-2B_n}$. (The value of $\mathbf{E}[\Omega^{(t)}]$ is insignificant for $t \ne o(2^{n/2})$). Therefore, the expected number of entries greater than or equal to $2B_n$ in a difference table with respect to $\otimes$ is

$$\begin{aligned}\sum_{t\ge 2B_n} \mathbf{E}[\Omega^{(t)}] &\le \sum_{t\ge 2B_n} \frac{1}{(2B_n)^{t-2B_n}} \cdot \mathbf{E}[\Omega] && (19)\\ &= \mathbf{E}[\Omega] \cdot \sum_{i\ge 0} \frac{1}{(2B_n)^i} && (20)\\ &= \frac{\mathbf{E}[\Omega]}{1 - 1/(2B_n)} && (21)\\ &\sim \mathbf{E}[\Omega] = o(1). && (22)\end{aligned}$$

Note that the probability that $DP_{\otimes}(\pi) \ge B_n/2^{n-1}$ is less than the expected number of entries of size $t \ge 2B_n$. Therefore, $\Pr(DP_{\otimes}(\pi) \ge B_n/2^{n-1}) = o(1)$ as $n \to \infty$. □

## References

1. C. M. Adams. The CAST-256 Encryption Algorithm. NIST Advanced Encryption Standard (AES) submission, description available at http://www.entrust.com/resources/pdf/cast.pdf.
2. E. A. Bender. Asymptotic methods in enumeration. *SIAM Review*, 16(4):485–515, 1974.
3. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.
4. E. Biham and A. Shamir. *Differential cryptanalysis of Data Encryption Standard*. Springer–Verlag, 1993.

5. J. Daemen and V. Rijmen. AES proposal: Rijndael. NIST Advanced Encryption Standard submission, description available at http://www.esat.kuleuven.ac.be/∼rijmen/rijndael.
6. W. Feller. *An Introduction to Probability Theory and its Applications*. New York: Wiley, 3rd edition, Volume 1, 1968.
7. H. Gilbert, M. Girault, P. Hoogvorst, F. Noilhan, T. Pornin, G. Poupard, J. Stern, and S. Vaudenay. Decorrelated Fast Cipher. NIST Advanced Encryption Standard (AES) submission, description available http://www.dmi.ens.fr/∼vaudenay/dfc.html.
8. R. P. Grimaldi. *Discrete and Combinatorial Mathematcis: An Applied Introduction*. Addison Wesley Publishing Company, 1989.
9. M. Hall. *Combinatorial Theory*. Blaisdell Publishing Company, 1967.
10. P. Hawkes and L. J. O'Connor. Aymptotic bounds on differential probabilities. Technical Report RZ 3018, IBM Research Report, May, 1998. Available from http://www.research.ibm.com.
11. B. S Kaliski and L. Y. Yiqun. On differential and linear cryptanalysis of the RC5 algorithm. *Advances in Cryptology, CRYPTO 95, Lecture Notes in Computer Science, vol. 963, D. Coppersmith eds., Springer-Verlag*, pages 171–184, 1995.
12. M. Kanda, S. Moriai, A. Kazumaro, H. Ueda, M. Ohkubo, Y. Takashima, K. Ohta, and T. Matsumoto. Specification of E2 - a 128-bit block cipher. NIST Advanced Encryption Standard submission, description available at http://titan.isl.ntt.co.jp/e2.
13. X. Lai. *On the design and security of block ciphers*. ETH Series in Information Processing, editor J. Massey, Hartung-Gorre Verlag Konstanz, 1992.
14. X. Lai and J. L. Massey. A proposal for a new block encryption standard. In *Advances in Cryptology, EUROCRYPT 90, Lecture Notes in Computer Science, vol. 473, I. B. Damgård ed., Springer-Verlag*, pages 389–404, 1991.
15. J. Lee, H. M. Heys, and S. E. Tavares. Resistance of a CAST-like encryption algorithm to linear and differential cryptanalysis. *Designs, Codes and Cryptography*, 12(3):267–282, 1997.
16. C. H. Lim. Specification and analysis of CRYPTON version 1.0. NIST Adavanced Encryption Standard (AES) submission, description available at http://crypt.future.co.kr/∼chlim/crypton.html.
17. J. L. Massey. SAFER: a byte-oriented ciphering algorithm. *Fast Software Encryption, Lecture Notes in Computer Science, vol. 809, R. Anderson ed., Springer-Verlag*, pages 1–17, 1993.
18. J. L. Massey. SAFER K-64: one year later. *Fast Software Encryption, Lecture Notes in Computer Science, vol. 1008, B. Preneel ed., Springer-Verlag*, pages 212–241, 1994.
19. L. J. O'Connor. On the distribution of characteristics in bijective mappings. *Advances in Cryptology, EUROCRYPT 93, Lecture Notes in Computer Science, vol. 765, T. Helleseth ed., Springer-Verlag*, pages 360–370, 1994.
20. L. J. O'Connor. On the distribution of characteristics in bijective mappings. *Journal of Cryptology*, 8(2):67–86, 1995.
21. B. Schneier, J. Kelsey, D. Whiting, D. Wagner, C. Hall, and N. Ferguson. Twofish: a 128-bit block cipher. NIST Advanced Encryption Standard (AES) submission, description available http://www.counterpane.com/twofish.html.