Pairwise independence of AES-like block ciphers

Tim Beyne[1], Gregor Leander[2], and Immo Schütt[2]

[1] COSIC, KU Leuven, Leuven, Belgium
firstname.lastname@esat.kuleuven.be
[2] Faculty of Computer Science, Ruhr University Bochum, Bochum, Germany
firstname.lastname@rub.de

Abstract. We prove that $4r+4$ rounds of an AES variant with independent and uniform random round keys are ε-close to pairwise independent with $\varepsilon = 2^{14}\,2^{-39r}$. This result follows from a near-optimal bound for a two-norm version of pairwise independence for the SHARK construction, depending on the third singular value of the difference-distribution table of the S-boxes. Our analysis combines insights from cryptanalysis — in particular, truncated differentials — and linear algebra over the reals.

Keywords: Pairwise independence · AES · SHARK · Truncated differentials

1 Introduction

The design and analysis of symmetric primitives has always been driven by the goal of achieving efficient constructions that no one knows how to break. This strategy has proven successful: the primitives we have today are efficient and provide excellent security in practice. As a consequence, symmetric cryptography in general — and the AES in particular — is used virtually everywhere whenever sensitive data is transmitted, stored, or processed.

Although focusing on efficiency and adapting to new attacks in order to avoid them may seem like a natural strategy, the approach in other areas of cryptography is different. There, the search for constructions whose security can be argued more formally has attracted more attention.

Even if it might seem unlikely that the AES will be broken any time soon, we lack rigorous security arguments to prove this. Moreover, although the likelihood of a practical attack on the AES seems small, the impact of such an attack would be catastrophic. Based on the costs of smaller, local security incidents [13], even a conservative estimate easily adds up to billions of euros in economic losses. Thus, we are well-advised to strive for a better understanding of the security of the AES. Beyond economic motivations, this is even more compelling if the arguments are technically interesting, as they could open up new directions for the design and analysis of symmetric primitives.

Previous work. The main attack vectors studied in the literature are differential and linear cryptanalysis [9,25], along with their variants. The initial attacks

were published in the early 1990s, and since then many ideas have been developed to argue that they do not apply to a given design. For the AES, the wide-trail strategy [14,15] convincingly rules out linear trails with high absolute correlations and differential characteristics with high key-averaged probabilities (for independent round keys).

However, even bounding *key-averaged* squared correlations of linear approximations or probabilities of differentials is already difficult. Such bounds were first proposed by Nyberg and Knudsen [27]. For four rounds of the AES with independent round keys, a line of work has led to tight bounds [12, 17, 19]. A few other constructions such as BISON [11] also offer bounds of this type. All of these results are limited to individual linear and differential properties.

Several information-theoretical security notions have been proposed that would allow ruling out broader classes of attacks. An important example is Vaudenay's decorrelation theory [32,33]. Although conceptually interesting, such designs never became mainstream — either because they were shown to be insecure, or because of efficiency concerns. More recently, the related notion of pairwise independence has been studied in depth [21–23]. A cipher is pairwise independent if, for any two distinct fixed plaintexts and a random key, the corresponding ciphertext pair is uniformly distributed over all pairs of distinct ciphertexts.

At Crypto 2021, Liu, Tessaro and Vaikuntanathan [23] demonstrated that pairwise independence can actually be applied to an AES variant with independent round keys. These were the first meaningful pairwise independence results on a real-world, widely deployed construction. However, while technically difficult, the bounds were rather weak: more than 9,000 rounds of the AES would be required to achieve a statistical distance below 2^{-128}. Practical expectations suggest that 10 rounds suffice, leaving a huge gap between what we can currently prove and what we would like to show.

Our contribution. In this paper, we narrow this gap by showing that 20 rounds of the AES with independent round keys are 2^{-128}-close to pairwise independent in statistical distance.

Technically, we deviate quite drastically from previous approaches: instead of bounding the statistical distance between distributions directly, we rely on linear algebra. In a nutshell, we prove pairwise independence by bounding the operator norm distance between the difference distribution table (DDT) of the cipher and that of a uniform random permutation.

After introducing the necessary background in Section 2, we show in Section 3 how pairwise independence is related to operator norms and to the third singular value of the DDT. These observations already yield — rather straightforwardly — a natural proof of the amplification lemma [26], and strong pairwise independence bounds for the cipher MiMC [2]. In a similar way, we prove bounds for the key-alternating cipher with round function $x \mapsto 1/x$ (patched at $x = 0$). The latter is an excellent example of a block cipher that is pairwise independent but easily broken using quadruples of plaintexts, as already mentioned in [23]. Here, we also clarify the cryptanalytic importance of pairwise independence by using

the geometric approach to cryptanalysis [4, 5, 7] to formalize the class of techniques it can rule out. This interpretation shows that the significance of pairwise independence does not depend on the assumption of round-key independence.

Section 4 states our main results and gives a high-level overview of the proof. Our first result (Theorem 3) is an explicit, closed-form pairwise independence bound for arbitrary SHARK-like ciphers. This bound depends on the MDS property of the linear layer, and on the third singular value of the DDT of the S-boxes. Our second result (Theorem 5) is a concrete bound for four rounds of the AES. Corollary 2 shows that $4r+4$ rounds of an AES variant with independent round keys are $2^{14}\,2^{-39r}$-close to pairwise independent.

The main body of our proof consists of two parts. The first part is concerned with the analysis of truncated differentials (Section 5), and depends only on the properties of the MDS matrix. The second part is less classical and additionally depends on the third singular value of the DDT of the S-boxes.

Finally, based on our insights, we argue in Section 7 that rather than studying only the linearity and differential uniformity (i.e., the maximum entries of the LAT and DDT) of S-boxes, the third singular value of the DDT deserves attention. We make initial observations for small S-boxes, and propose a conjecture for the inverse function. This new criterion has the potential to open up many interesting directions in the area of Boolean functions.

Parallel and independently of our work, Dinur [16] obtained significant improvements over the state of the art, which are complementary to our work. Whereas our focus is mainly on truncated differentials and pairwise independence — leading to the strongest known bounds in these cases — Dinur focuses on concrete differentials, yielding the strongest known bounds there. We expect that combining both approaches would lead to improvements to the bounds, and to further insights for the AES and symmetric cryptography in general.

2 Preliminaries

The following three sections discuss the main preliminaries required to understand our proofs: the structure of AES-like primitives, differential cryptanalysis, and linear algebra.

2.1 AES-like ciphers

The results in this paper apply to block ciphers and cryptographic permutations that follow the SHARK-structure [29], as illustrated in Figure 1. The round function consists of a substitution layer, a linear mixing layer and a key addition. The linear layer is MDS, and the S-box layer applies n identical S-boxes $S\colon \mathbb{F}_q \to \mathbb{F}_q$ in parallel. Throughout this paper, $\text{SHARK}[n, q, S, L]$ denotes the composition of the S-box layer, key-addition, the linear layer, and another S-box layer.

The most widely used block cipher based on the SHARK structure is the *Advanced Encryption Standard* (AES). The standard description of the AES follows the structure in Figure 1 with $n = 16$ S-boxes on $\mathbb{F}_q$ for $q = 2^8$. The

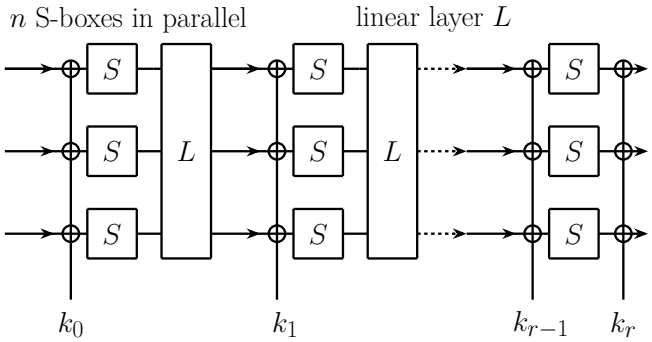

Fig. 1. An r-round block cipher based on the SHARK structure.

number of rounds r depends on the key length, and is either 10, 12, or 14. However, in this representation, the linear layer L is not an MDS matrix. In this paper, we instead use the superbox representation (see Figure 2) of the AES [1]. This allows describing the AES as a SHARK construction with four key-dependent 32-bit S-boxes and an MDS linear layer. For more details, we refer the reader to Section B.1.

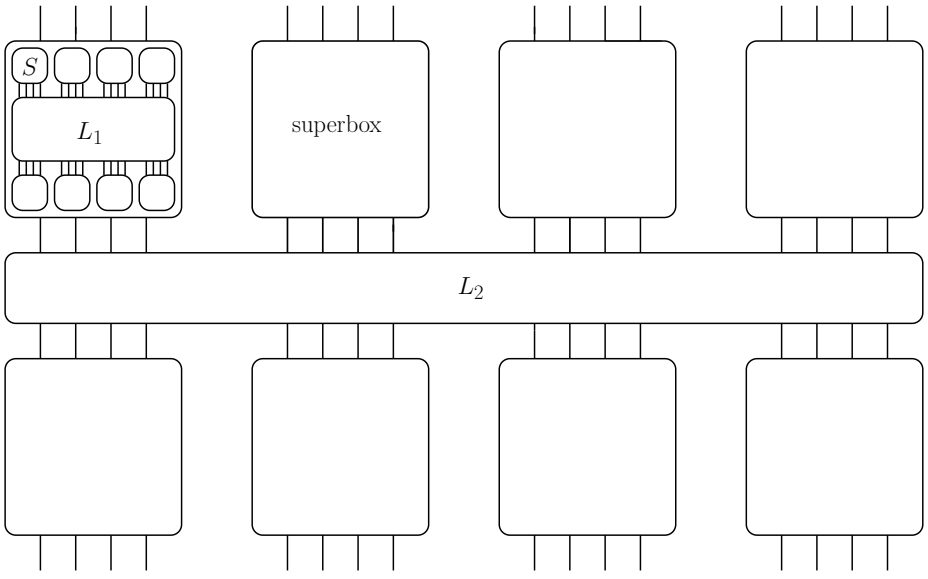

Fig. 2. The superbox representation of the AES.

Recall that an additive function $L\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ is MDS if $\operatorname{wt} x + \operatorname{wt} L(x) \geq n+1$ for all nonzero x, with $\operatorname{wt} x$ the Hamming weight of x. Throughout this paper, we frequently use the following property of MDS maps.

Lemma 1 (cf. [24, §11.4, Theorem 8]). *If $L\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ is MDS, then all square subblocks of L are invertible. Here, a block of L is obtained by setting some of the input coordinates to zero and deleting some of the output coordinates.*

2.2 Differential cryptanalysis

The difference-distribution matrix of a (possibly random) function F on $\mathbb{F}_q^n$ is a matrix D^F with coordinates labeled by elements of $\mathbb{F}_q^n$ and

$$D^F_{b,a} := \Pr_{F,x}\big[F(x+a) = F(x) + b\big]\,, \tag{1}$$

where the probability is with respect to F and a uniform random input x. The values a and b are called differences, and the pair (a, b) is a *differential*. If F is a fixed function, then D^F is (up to scaling and transposition) the difference-distribution table introduced by Biham and Shamir in [9]. If F is an AES-like cipher with independent and uniform random round keys, then D^F contains the key-averaged probabilities of differentials. Finally, if F is a uniform random permutation, then $D^F_{b,a} = 1/(q^n - 1)$ for all nonzero a and b, $D^F_{0,0} = 1$, and zero elsewhere. We denote this matrix by U.

Finally, note that the notation for the matrix D^F should not be confused with the quasidifferential transition matrix of F [7]. However, for fixed F, the matrix D^F is a submatrix of the quasidifferential transition matrix of F. In particular, unlike for quasidifferential transition matrices, it is *not* true in general that

$$D^{F_2 \circ F_1} = D^{F_2} D^{F_1}\,. \tag{2}$$

However, if the output of F_1 or the input of F_2 is whitened by the addition of a uniform random key (independent of F_1 and F_2, if they are random), then Equation (2) does hold. In this case, the matrices D^{F_i} are transition matrices of a Markov chain, as explained by Lai, Massey and Murphy [20]. Section 3.3 provides a different interpretation of the matrices D^{F_i} in terms of the geometric approach to cryptanalysis.

2.3 Linear algebra over $\mathbb{R}$

For a finite set X, let $\mathbb{R}[X]$ be the free vector space on X. This is the space of all formal $\mathbb{R}$-linear combinations of the elements of X. To avoid confusion, we denote the standard basis vector of $\mathbb{R}[X]$ that corresponds to an element x of X by δ_x. Free vector spaces play an important role in the geometric approach to cryptanalysis [4,5], see [8, §2.3] and [6] in particular. As explained in Section 3.3, our overall proof strategy is an application of this framework.

Norms on $\mathbb{R}[X]$. For $p \geq 1$, the p-norm of a vector v in $\mathbb{R}[X]$ with coordinates v_x in the standard basis is $\|v\|_p = \left(\sum_x |v_x|^p\right)^{1/p}$. The maximum norm $\|v\|_\infty$ is the limiting case $p \to \infty$. The operator norm of a linear map $L\colon \mathbb{R}[X] \to \mathbb{R}[X]$ with respect to norms $\|\cdot\|_p$ and $\|\cdot\|_q$ is defined by

$$\|L\|_{q,p} := \max_{\|v\|_p=1} \|L(v)\|_q \,.$$

In particular, we write $\|L\|_{p,p} = \|L\|_p$. These norms are submultiplicative: $\|L_2 L_1\|_{r,p} \leq \|L_2\|_{r,q} \, \|L_1\|_{q,p}$ for all p, q and $r \geq 1$. The norm of a matrix M is defined as the norm of the map $v \mapsto Mv$. In Section 4, we need the following lemma. A proof can be found in Section B.2.

Lemma 2. *For all matrices A, B, C and D, we have*

$$\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_2 \leq \left\| \begin{bmatrix} \|A\|_2 & \|B\|_2 \\ \|C\|_2 & \|D\|_2 \end{bmatrix} \right\|_2 .$$

Tensor product. The tensor product of $\mathbb{R}[X]$ and $\mathbb{R}[Y]$ is $\mathbb{R}[X \times Y]$. In particular, $\otimes$ denotes the bilinear map $\otimes\colon \mathbb{R}[X] \times \mathbb{R}[Y] \to \mathbb{R}[X \times Y]$ defined by $\delta_x \otimes \delta_y = \delta_{(x,y)}$ on basis vectors. If vectors u and v have coordinates u_x and v_y respectively, then $u \otimes v$ has coordinates $(u \otimes v)_{(x,y)} = u_x v_y$.

The tensor product of linear maps $A\colon \mathbb{R}[X] \to \mathbb{R}[X]$ and $B\colon \mathbb{R}[Y] \to \mathbb{R}[Y]$ is denoted by $A \otimes B\colon \mathbb{R}[X \times Y] \to \mathbb{R}[X \times Y]$. The coordinates of the matrix representation of $A \otimes B$ with respect to the standard basis are given by

$$(A \otimes B)_{(c,d),(a,b)} = A_{c,a} B_{d,b} \,.$$

Here, we used the same notation for the linear maps A, B and $A \otimes B$ as for their matrix representation with respect to the standard basis. For all A, B, C and D, we have $(A \otimes B)(C \otimes D) = AC \otimes BD$ if this is well-defined. The m-fold tensor product of A is denoted by $A^{\otimes m} := A \otimes A \otimes \cdots \otimes A$.

Inclusion and orthogonal projection. Let V be a subspace of $\mathbb{R}[X]$, which is equipped with the inner product $\left\langle \sum_{x \in X} u_x \delta_x, \sum_{x \in X} v_x \delta_x \right\rangle = \sum_{x \in X} u_x v_x$. There is an *inclusion* map $i_V\colon V \to \mathbb{R}[X]$, defined by $i_V(v) = v$. If $V^\perp$ denotes the orthogonal complement of V with respect to the inner product on $\mathbb{R}[X]$, then $\mathbb{R}[X] = V \oplus V^\perp$. The unique linear map $\pi_V\colon \mathbb{R}[X] \to V$ with $\pi_V(u) = u$ for u in V and kernel $V^\perp$ is called the *orthogonal projection* onto V. It is the adjoint of i_V because $\langle u, i_V(v) \rangle = \langle \pi_V(u), v \rangle$ for all u in $\mathbb{R}[X]$ and v in V. Indeed, this follows from $u - \pi_V(u) \in V^\perp$, so that $\langle u, i_V(v) \rangle = \langle \pi_V(u), v \rangle$.

Lemma 3. *Let V be a subspace of $\mathbb{R}[X]$, and let B be a matrix whose columns contain the coordinates of the basis vectors for V relative to the standard basis for $\mathbb{R}[X]$. The matrix representation of π_V relative to the standard basis for $\mathbb{R}[X]$ and the basis consisting of the columns of B, is $(B^\intercal B)^{-1} B^\intercal$.*

Proof. The proof can be found in Section B.3. □

Some care should be taken when using Lemma 3 to compute the orthogonal projection $\pi_V(x)$. The coordinates of $\pi_V(x)$ relative to the standard basis of $\mathbb{R}[X]$ are given by $B(B^\intercal B)^{-1}B^\intercal x$ rather than $(B^\intercal B)^{-1}B^\intercal x$. In particular, the norm of $\pi_V(x)$ is equal to $\|B(B^\intercal B)^{-1}B^\intercal x\|_2$.

3 Pairwise independence in terms of linear algebra

A random permutation[3], such as a block cipher E_k with a random key k, is called t-wise independent if the probability distribution of t-tuples of ciphertexts for distinct plaintexts is uniform random. This can be relaxed to the notion of almost t-wise independence, which stipulates that the statistical distance to the uniform distribution is less than some parameter ε [23]. In this paper, we exclusively focus on almost pairwise independence.

Definition 1 (Almost pairwise independence [23]). *A random permutation F on $\mathbb{F}_q^n$ is ε-pairwise independent if for all pairs $x_1 \neq x_2$ in $\mathbb{F}_q^n$,*

$$\frac{1}{2}\sum_{y_1 \neq y_2} \left| \Pr\left[F(x_1) = y_1 \text{ and } F(x_2) = y_2\right] - \frac{1}{q^n} \times \frac{1}{q^n - 1} \right| \leq \varepsilon\,,$$

where the probability is with respect to F and the sum is over y_1 and y_2 in $\mathbb{F}_q^n$.

In Section 3.1, we rephrase Definition 1 in terms of the one norm of the difference distribution matrix of F. Section 3.2 contains some initial applications to MiMC and the key-alternating cipher with the inverse map as the round function. The cryptanalytic interpretation of pairwise independence is discussed in Section 3.3.

3.1 Pairwise-independence in terms of norms

If F is a random permutation with independent and uniform random whitening key additions at the input and output, then we have the following result. An equivalent result already appeared in the work of Liu et al. [23, Lemma 2.7], but here we rephrase it in terms of matrix norms.

Lemma 4. *Let G be a random permutation on $\mathbb{F}_q^n$ with difference-distribution matrix D. If F is the random permutation defined by $F(x) \coloneqq G(x+k_1)+k_2$ with k_1 and k_2 uniform random on $\mathbb{F}_q^n$ and independent of G, then F is ε-pairwise independent if and only if $\|D - U\|_1 \leq 2\varepsilon$.*

Proof. Let $x_1 \neq x_2$ be distinct values in $\mathbb{F}_q^n$. By definition, for all $y_1 \neq y_2$, $\Pr\left[F(x_1) = y_1 \wedge F(x_2) = y_2\right] = \Pr\left[G(x_1+k_1) = y_1-k_2 \wedge G(x_2+k_1) = y_2-k_2\right]$. The random variables $x \coloneqq x_1 + k_1$ and $y_1 - k_2$ are uniform random on $\mathbb{F}_q^n$, so

$$\Pr\left[F(x_1) = y_1 \wedge F(x_2) = y_2\right] = \frac{1}{q^n} \times \Pr\left[G(x + x_2 - x_1) = G(x) + y_2 - y_1\right]\,.$$

[3] A random variable (usually not uniform) on the set of permutations on $\mathbb{F}_q^n$.

Let $a := x_2 - x_1$ and $b := y_2 - y_1$, then the right-hand side is equal to $D_{b,a}/q^n$. Hence, the left-hand side of Definition 1 is equal to

$$\begin{aligned}
&\frac{1}{2} \sum_{y_1 \neq y_2} \left| \Pr\left[F(x_1) = y_1 \wedge F(x_2) = y_2\right] - \frac{1}{q^n} \times \frac{1}{q^n - 1} \right| \\
&= \frac{1}{2} \sum_{y_1 \in \mathbb{F}_q^n} \frac{1}{q^n} \sum_{b \neq 0} \left| D_{b,a} - \frac{1}{q^n - 1} \right| \\
&= \frac{1}{2} \sum_{b \neq 0} \left| D_{b,a} - \frac{1}{q^n - 1} \right| .
\end{aligned}$$

Since $U_{b,a} = 1/(q^n - 1)$ for $a, b \neq 0$, the result now follows from the following expression for the one norm:

$$\|D - U\|_1 = \max_{a \in \mathbb{F}_q^n} \sum_{b \in \mathbb{F}_q^n} |D_{b,a} - U_{b,a}| .$$

In particular, $\|D - U\|_1 \leq 2\varepsilon$ if and only if F is ε-pairwise independent. □

As a first application of Lemma 4, the amplification lemma for pairwise independence [23, 26] is implied by the submultiplicativity of the one norm.

Lemma 5 (Amplification lemma [26]). *Let F_1 and F_2 be independent random permutations with input and output whitening key additions. If F_1 and F_2 are ε_1- and ε_2-pairwise independent, then $F_2 \circ F_1$ is $2\varepsilon_1\varepsilon_2$-pairwise independent.*

Proof. Let D^{F_1} and D^{F_2} be the difference-distribution matrices of F_1 and F_2 respectively. As discussed in Section 2.2, the difference-distribution matrix of $F_2 \circ F_1$ is equal to $D^{F_2} D^{F_1}$. Since $D^{F_i} U = U = U D^{F_i}$ and $U^2 = U$, we have $D^{F_2} D^{F_1} - U = (D^{F_2} - U)(D^{F_1} - U)$. By submultiplicativity of matrix norms,

$$\|D^{F_2} D^{F_1} - U\|_1 \leq \|D^{F_2} - U\|_1 \|D^{F_1} - U\|_1 \leq 4\varepsilon_1\varepsilon_2 .$$

By Lemma 4, this implies that $F_2 \circ F_1$ is $2\varepsilon_1\varepsilon_2$-pairwise independent. □

Lemma 5 generalizes as follows:

$$\|D^{F_2} D^{F_1} - U\|_{r,p} \leq \|D^{F_2} - U\|_{r,q} \|D^{F_1} - U\|_{q,p} , \tag{3}$$

for all p, q and $r \geq 1$. This generalization also allows us to convert a bound from one norm to another. This is important in our analysis, since our proofs rely on the two norm rather than the one norm.

Lemma 6 (Conversion). *Let F_1 and F_2 be independent random permutations on $\mathbb{F}_q^n$ with difference-distribution matrices D^{F_1} and D^{F_2} respectively. If $D^{F_1}_{b,a} \leq p$ for all differences $a \neq 0$ and b, then*

$$\|D^{F_2} D^{F_1} - U\|_1 \leq q^n p \, \|D^{F_2} - U\|_2 .$$

Proof. By inequality (3),

$$\|D^{F_2} D^{F_1} - U\|_1 \leq \|D^{F_2} - U\|_{1,2} \|D^{F_1} - U\|_{2,1} \leq q^{n/2} \|D^{F_2} - U\|_2 \|D^{F_1} - U\|_{2,1} \,.$$

By definition of the operator norm $\|\cdot\|_{2,1}$, we have

$$\|D^{F_1} - U\|_{2,1} = \max_{a \in \mathbb{F}_q^n} \sqrt{\sum_{b \in \mathbb{F}_q^n} \left(D^{F_1}_{b,a} - U_{b,a}\right)^2} \leq \max_{a \neq 0} \sqrt{\sum_{b \neq 0} \left(D^{F_1}_{b,a}\right)^2} \leq q^{n/2}\, p \,.$$

The result follows by combining both inequalities. □

3.2 Singular value decomposition

Lemma 4 shows that almost pairwise independence is measured by the one norm of the difference-distribution matrix. In practice, our proof bounds the two norm and deduces a bound on the one norm using Lemma 6. Our analysis of the two norm relies on the singular value decomposition of the difference-distribution matrix of the S-box.

For every real matrix A, the matrix $A^\intercal A$ is symmetric. Hence, it is orthogonally diagonalizable. Let $v_1, \ldots, v_d$ be an orthogonal basis of eigenvectors of $A^\intercal A$. The corresponding eigenvalues are necessarily non-negative, since $\langle Au_i, Au_i \rangle \geq 0$. The singular values of A are the square roots $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_d$ of the eigenvalues of $A^\intercal A$. The orthogonal eigenvectors $v_1, \ldots, v_d$ are called right singular vectors of A. The corresponding left singular vectors are defined by $u_i \coloneqq Av_i/\sigma_i$. The singular value decomposition of A is the matrix factorization $A = U \Sigma V^\intercal = \sum_{i=1}^d \sigma_i\, u_i v_i^\intercal$. Geometrically, the singular values express the extent of stretching (or rather compressing in our setting) under a linear map with respect to the two norm. In particular, the largest singular value of A is equal to $\sigma_1 = \|A\|_2$.

Lemma 7. *Let D be the difference-distribution matrix of a (possibly random) permutation on $\mathbb{F}_q^n$. All singular values of D are less than or equal to one, and there are at least two singular values equal to one with corresponding singular vectors δ_0 and $\delta_{\mathbb{F}_q^n \setminus \{0\}}/\sqrt{q^n - 1}$. These are both left and right singular vectors. Furthermore, the third singular value of D is equal to $\|D - U\|_2$.*

Proof. The matrix D satisfies $\sum_a D_{b,a} = 1$ for every b, whence $\|D\|_\infty = 1$. Likewise, $\|D\|_1 = 1$, so that $\|D\|_2 \leq 1$. Hence, the singular values of D are at most one. In addition,

$$D\,\delta_0 = D^\intercal \delta_0 = \delta_0 \text{ and } D\,\delta_{\mathbb{F}_q^n \setminus \{0\}} = D^\intercal \delta_{\mathbb{F}_q^n \setminus \{0\}} = \delta_{\mathbb{F}_q^n \setminus \{0\}} \,.$$

It follows that δ_0 and $\delta_{\mathbb{F}_q^n \setminus \{0\}}/\sqrt{q^n - 1}$ are (both left and right) singular vectors. To prove the claim about the third singular value of D, note that

$$U = \delta_0\, \delta_0^\intercal + \frac{1}{q^n - 1}\, \delta_{\mathbb{F}_q^n \setminus \{0\}}\, \delta_{\mathbb{F}_q^n \setminus \{0\}}^\intercal \,.$$

The two terms above correspond to the first two terms in a singular value decomposition of D. The remaining terms are given by the singular value decomposition of $D-U$. Hence, the third singular value of D is the first singular value of $D-U$, which is equal to $\|D-U\|_2$. □

Intuitively, the two singular vectors in Lemma 7 correspond to probability-one (truncated) differentials $0 \to 0$ and $\mathbb{F}_q^n \setminus \{0\} \to \mathbb{F}_q^n \setminus \{0\}$. Additional details about this interpretation are given in Section 3.3.

These ingredients already allow us to obtain strong and sometimes even optimal pairwise independence bounds, if we are able to compute the third singular value of the entire round function. This is indeed possible for ciphers with a power function $x \mapsto x^d$ as their round function, like in Figure 3.

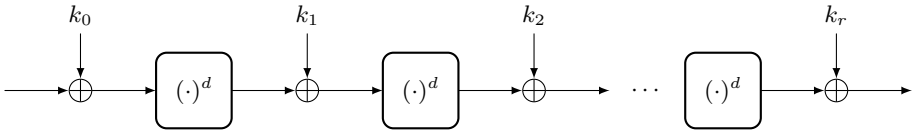

Fig. 3. An r-round MiMC-type key-alternating ciphers with round function $x \mapsto x^d$.

The most prominent example of such a cipher is MiMC [2], for which $q = 2^m$ and exponent $d = 3$. As shown in Lemma 14 of Section A, the third singular value of the difference-distribution matrix of the cube function $x \mapsto x^3$ is $2^{-m/2}$. Here, m is assumed odd, as otherwise cubing is not a permutation.

Theorem 1. *The random permutation on $\mathbb{F}_{2^m}$ with odd m, defined by $r+1$-round MiMC with uniform random round keys, is $2^{-rm/2}$-pairwise independent.*

Proof. To apply Lemma 6, we split $r+1$-round MiMC into two parts: the first round and the last r rounds. The maximum probability of a nontrivial differential for the cube function is $p = 2/2^m$, because the quadratic equation $(x+a)^3 = x^3 + b$ in x has at most two solutions. If D is the difference-distribution matrix of r-round MiMC with independent and uniform random round keys, then Lemma 5 and Lemma 6 imply

$$\|D-U\|_1 \leq 2^m p \cdot \left(2^{-m/2}\right)^r = 2 \cdot 2^{-rm/2}\,.$$

The result now follows from Lemma 4. □

Theorem 1 shows that three rounds of MiMC are already almost 2^{-m}-close to pairwise independence. Another interesting example, already studied in Theorem A.1 from [23], is the key-alternating cipher with round function $x \mapsto 1/x$ (and output zero for $x = 0$). In this case, Lemma 15 of Appendix A shows that the third singular value of the difference-distribution matrix of the inverse function (for odd m, as in this case it is an APN permutation) is at most $2^{-m/2} + 2^{-m+1}$. The same arguments as in the proof of Theorem 1 yield the following result.

Theorem 2. *For odd m, the $r+1$-round key-alternating cipher on $\mathbb{F}_{2^m}$ with the inverse round function and uniform random round keys is $(2^{-m/2}+2^{-m+1})^r$-pairwise independent.*

The bound in Theorem 2 improves over the bound in Corollary A.2 from [23], although the authors of [23] examined the case with even m. In the above, we focused on odd m mainly because it yields a slightly better bound. However, the same approach yields an improvement over the bound from [23] for even m.

Finally, it is worth recalling the remark made in [23]: this cipher achieves pairwise independence for a small number of rounds, but remains insecure for any reasonable number of rounds. This is due to the rational interpolation attack of Jakobsen and Knudsen [18].

3.3 Cryptanalytic interpretation in terms of the geometric approach

The following may be helpful for readers familiar with the geometric approach, but is not necessary to understand our proofs. Sections 3.1 and 3.2 discussed how the properties of the difference-distribution matrix lead to pairwise independence bounds. More generally, the difference-distribution matrix also determines many other cryptanalytic properties. The scope of our arguments is best understood using the geometric approach to cryptanalysis [4,5]. The paragraphs below provide a short overview of the connection to cryptanalysis. The main elements necessary to understand our proof strategy are introduced in Section 4.

Pairwise cryptanalysis. The results we prove in this paper can be interpreted as security arguments against a class of *pairwise* cryptanalytic techniques. Examples of pairwise cryptanalysis include differential cryptanalysis and its variants such as multiple and truncated differential cryptanalysis. Linear cryptanalysis is another example. In this case the use of pairs is redundant, since individual values also suffice to describe linear cryptanalysis.

In the geometric approach to cryptanalysis, pairwise properties of a primitive $F\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ are described in terms of linear algebra over the free vector space on pairs $k[\mathbb{F}_q^n \times \mathbb{F}_q^n]$, over a field k. In this paper, we assume $k = \mathbb{R}$ or $k = \mathbb{C}$. From the point of view of the geometric approach, the analysis of the primitive F is equivalent to the analysis of a linear map $\mathbb{C}[\mathbb{F}_q^n \times \mathbb{F}_q^n] \to \mathbb{C}[\mathbb{F}_q^n \times \mathbb{F}_q^n]$. Different matrix representations of this map lead to different pairwise cryptanalytic techniques. One example is the quasidifferential transition matrix [7].

Key-independent pairwise cryptanalysis. The security arguments we develop are applicable to pairwise methods that are independent of the round keys, i.e. all one-round properties used in the analysis are invariant under the addition of keys (and, in general, constants). This can be formalized in terms of *trails*: in the geometric approach, a trail for a composition of functions $F = F_r \circ \cdots \circ F_2 \circ F_1$ is a sequence $(V_1, V_2, \ldots, V_r)$ of subspaces of $\mathbb{R}[\mathbb{F}_q^n \times \mathbb{F}_q^n]$. These subspaces result from decompositions of $\mathbb{R}[\mathbb{F}_q^n \times \mathbb{F}_q^n]$ as direct sums. A trail $(V_1, V_2, \ldots, V_r)$ is

invariant under key addition if and only if

$$V_i \subseteq \mathrm{Span}\left\{\sum_{x\in\mathbb{F}_q^n} \delta_{(x,x+a)} \mid a \in \mathbb{F}_q^n\right\}, \tag{4}$$

for every subspace V_i in the trail. Hence, in this case, all of the vector spaces V_i may be considered to be subspaces of the free vector space $\mathbb{R}[\mathbb{F}_q^n]$ on the set of differences. Every trail in the geometric approach comes with an *approximation map* [4, §5], which generalizes correlations of one-dimensional trails (such as linear trails, or quasidifferential trails). The approximation map of the largest key-independent trail, i.e. so that (4) is an equality, is $D^{F_r} \cdots D^{F_2} D^{F_1}$.

Bounds for $\|D^{F_r} \cdots D^{F_2} D^{F_1} - U\|_2$ allow ruling out all methods based on key-independent trails. In the geometric approach, the singular values of the approximation map of a trail are called *principal correlations*. These determine the data-complexity of known plaintext attacks [4,5]. This provides a more direct interpretation of our two-norm bounds.

On the one hand, the assumption of key-independence is a strong limitation. For example, for differential cryptanalysis, it implies that the analysis cannot take into account the effect of quasidifferential trails with nonzero masks. On the other hand, it does provide a strong argument against a broad class of attacks that includes differential and linear cryptanalysis in their key-independent form. This is valuable in itself, because most analysis in the literature is (often implicitly) based on these key-independent methods. This also applies to cryptographic permutations and other primitives without a key.

Truncated differentials. As we explain in Section 4, our proof strategy uses the geometric approach to reduce an important part of the analysis to an argument about truncated differentials defined by activity patterns. Hence, below, we provide a brief overview of truncated differential cryptanalysis as a key-independent pairwise cryptanalysis method. This also serves as a concrete example of the interpretation of our results in terms of the geometric approach.

A truncated differential is a pair of subsets (A, B) of $\mathbb{F}_q^n$. In our case, A and B will be defined by *activity patterns*. An activity pattern specifies which coordinates of the differences in A or B are zero, and which are arbitrary. The probability of (A, B) is (with x and y uniform random on $\mathbb{F}_q^n$)

$$\Pr[A \xrightarrow{F} B] \coloneqq \Pr_{x,y}\left[F(y) \in F(x) + B \mid y \in x + A\right].$$

In the geometric approach, a truncated differential (A, B) is described as a pair of one-dimensional vector spaces spanned by the indicator vectors $\delta_A = \sum_{x\in A} \delta_x$ and $\delta_B = \sum_{x\in B} \delta_x$ in $\mathbb{R}[\mathbb{F}_q^n]$ respectively. The probability of (A, B) is then

$$\Pr[A \xrightarrow{F} B] = \sum_{a\in A} \frac{1}{|A|} \sum_{b\in B} D_{b,a} = \left\langle \delta_B, D\,\delta_A \right\rangle / |A|.$$

Since $\|\delta_A/|A|\|_1 = 1$ and $\|\delta_B\|_\infty = 1$, the probability of truncated differentials is at most $\|D\|_1$. The bias of every truncated differential is less than $\|D - U\|_1$.

4 Main results and proof outline

In this section, we state our main results about the pairwise independence of the SHARK construction and the AES in particular. As explained in Section 4.1, our proofs are based on the concepts introduced in Section 3. The main technical challenge is bounding the two norm of the difference-distribution matrix.

4.1 High-level proof strategy

The main idea of our proof strategy is to split the analysis into two cases: (1) bounding the bias of truncated differentials defined by activity patterns over the linear layer, and (2) bounding all other properties in terms of the third singular value of the difference-distribution matrix of the S-boxes. To formalize this, we define a subspace V of $\mathbb{R}[\mathbb{F}_q^n]$ that captures truncated differential properties defined by an activity pattern.

Activity patterns. For a bitvector z in $\{0,1\}^n$, we denote $[z] \coloneqq [z_1] \times [z_2] \times \cdots \times [z_n]$ with $[0] = \{0\}$ and $[1] = \mathbb{F}_q$. That is, every bitvector or 'activity pattern' z defines a set of differences $[z] \subseteq \mathbb{F}_q^n$. The subspace V is defined as follows:

$$V = \operatorname{Span}\left\{ \delta_{[z]} \;\middle|\; z \in \{0,1\}^n \right\}.$$

This is a subspace of $\mathbb{R}[\mathbb{F}_q^n]$ of dimension 2^n. If $S\colon \mathbb{F}_q \to \mathbb{F}_q$ is a bijection with difference-distribution matrix D^S, then $(D^S)^{\otimes n}\, v = v$ for all vectors v in V.

Case distinction. To split the analysis into two cases, we use the decomposition $\mathbb{R}[\mathbb{F}_q^n] = V \oplus V^\perp$, where $V^\perp$ is the orthogonal complement of V. As in Section 2.3, let $\pi_V\colon \mathbb{R}[\mathbb{F}_q^n] \to V$ be orthogonal projection and $i_V\colon V \to \mathbb{R}[\mathbb{F}_q^n]$ the inclusion map. Likewise, there are maps $\pi_{V^\perp}$ and $i_{V^\perp}$. Since $\mathbb{R}[\mathbb{F}_q^n] = V \oplus V^\perp$, there exists an orthogonal basis for $\mathbb{R}[\mathbb{F}_q^n]$ so that $D - U$ has the following block matrix form:

$$\begin{bmatrix} \pi_V\ (D-U) i_V & \pi_V\ (D-U) i_{V^\perp} \\ \pi_{V^\perp} (D-U) i_V & \pi_{V^\perp} (D-U) i_{V^\perp} \end{bmatrix}.$$

Here, with some abuse of notation, each block is a matrix representation of the corresponding linear map with respect to an arbitrary orthogonal basis. In the geometric approach, the pairs of subspaces such as $(V, V^\perp)$ are called approximations and the maps such as $\pi_{V^\perp} D i_V$ are called approximation maps [4, Definition 4.1]. For the two norm,

$$\begin{aligned} \|D-U\|_2 &= \left\| \begin{bmatrix} \pi_V\ (D-U) i_V & \pi_V\ (D-U) i_{V^\perp} \\ \pi_{V^\perp} (D-U) i_V & \pi_{V^\perp} (D-U) i_{V^\perp} \end{bmatrix} \right\|_2 \\ &\leq \left\| \begin{bmatrix} \|\pi_V\ (D-U) i_V\|_2 & \|\pi_V\ (D-U) i_{V^\perp}\|_2 \\ \|\pi_{V^\perp} (D-U) i_V\|_2 & \|\pi_{V^\perp} (D-U) i_{V^\perp}\|_2 \end{bmatrix} \right\|_2. \end{aligned} \tag{5}$$

The inequality follows from Lemma 2 in Section 2.3. The term $\|\pi_V(D-U)i_V\|_2$ is controlled by truncated differentials, and is analyzed in Section 5. The other three terms are analyzed in Section 6. We will refer to them as *spectral terms*, because for the SHARK construction their bound depends on the third singular value of the difference-distribution matrix of the S-boxes.

4.2 Results for the SHARK construction

Let D be the difference-distribution matrix of $\text{SHARK}[n, q, S, L]$, i.e. the SHARK construction with n S-boxes $S\colon \mathbb{F}_q \to \mathbb{F}_q$ and an MDS linear layer L. The whitening keys and round keys are assumed to be independent and uniform random. Hence,

$$D = \left(D^S\right)^{\otimes n} D^L \left(D^S\right)^{\otimes n},$$

with D^S and D^L the difference-distribution matrices of the S-box and linear layer, respectively. Based on the results from Sections 5 and 6, we first prove a general closed-form bound for $\|D-U\|_2$.

Theorem 3. *Let D be the difference-distribution matrix of $\text{SHARK}[n, q, S, L]$ with $n \geq 2$ and $q \geq 4$. If the third singular value of the difference-distribution matrix of S is at most σ, then for every choice of $1 \leq l \leq n-1$,*

$$\|D-U\|_2^2 \leq \left(\frac{2}{\sqrt{q}-1}\right)^{2n} + \sum_{k=2}^{l} C(n,k)\,\frac{\sigma^{2k}}{(\sqrt{q}-1)^{2n-2k}} + (l+2)\sigma^{2(l+1)}\,,$$

where $C(n,k)$ depends only weakly on q:

$$C(n,k) = \left(1+\frac{1}{\sqrt{q}-1}\right)^{k} \binom{2n}{k} \sum_{i=1}^{k-1} \binom{2n-k}{n-i} 2^i\,.$$

Proof. The result is based on the following loose consequence of inequality (5):

$$\begin{aligned}\|D-U\|_2^2 \leq \|\pi_V(D-U)i_V\ \|_2^2 + \|\pi_{V^\perp}(D-U)i_V\ \|_2^2 \\ +\|\pi_V(D-U)i_{V^\perp}\|_2^2 + \|\pi_{V^\perp}(D-U)i_{V^\perp}\|_2^2\,.\end{aligned}$$

Theorem 6 in Section 5 shows that the first term is bounded by $(2/(\sqrt{q}-1))^{2n}$. The bound on the three other terms is given in Theorem 7 in Section 6. □

Theorem 3 provides a complicated but concrete bound. This will be important in Section 4.3 to derive our pairwise independence bound for the AES. From a theoretical point of view, the importance of this result is that the SHARK construction achieves an asymptotically optimal two norm bound of $\mathcal{O}(q^{-n/2})$ if the S-box is well chosen. In particular, we have the following corollary.

Corollary 1. *Let D be the difference-distribution matrix of $\text{SHARK}[n, q, S, L]$ with $n \geq 2$. If the third singular value of the difference-distribution matrix of S is at most $\mathcal{O}(1/\sqrt{q})$ as $q \to \infty$, then $\|D-U\|_2 = \mathcal{O}(q^{-n/2})$ for constant n.*

Proof. Take $l = n-1$ in Theorem 3. The terms in the upper bound on $\|D-U\|_2^2$ are of the order $\mathcal{O}(\sigma^{2k}/q^{n-k}) = \mathcal{O}(1/q^n)$. Indeed, $C(n,k) = \mathcal{O}(1)$. □

Theorem 3 implies a pairwise independence bound. In particular, if the maximum nontrivial differential probability of S is p, then the maximum nontrivial differential probability of $\textsc{Shark}[n, q, S, L]$ is equal to p^n. Hence, Lemmas 5 and 6 show that $2r+2$ rounds with independent round keys are ε-pairwise independent for $\varepsilon = \frac{1}{2}q^n p^n \|D-U\|_2^r$. In particular, for large q, we have $\varepsilon = \mathcal{O}(q^{-nr/2})$ if σ^2 and p are both bounded by $\mathcal{O}(1/q)$.

For fixed values of n and q, and if n is not too large, the closed-form bound in Theorem 3 admits numerical improvements. For example, for the AES superbox, we have the following result.

Theorem 4 (AES superbox). *For the difference-distribution matrix D of the AES superbox, $\|D-U\|_2 \leq 2^{-10.146}$.*

Proof. In Section 5, we show that $\|\pi_V(D-U)i_V\|_2 \leq 2^{-13.351}$ for $n=4$ and $q=2^8$. For the same parameters and specifically for the AES S-box, Section 6 shows that $\|\pi_V(D-U)i_{V^\perp}\|_2$ and $\|\pi_{V^\perp}(D-U)i_V\|_2$ are at most $2^{-11.178}$, and $\|\pi_{V^\perp}(D-U)i_{V^\perp}\|_2 \leq 2^{-10.598}$. Hence, by inequality (5),

$$\|D-U\|_2 \leq \left\| \begin{bmatrix} 2^{-13.351} & 2^{-11.178} \\ 2^{-11.178} & 2^{-10.598} \end{bmatrix} \right\|_2 \leq 2^{-10.146}.$$

This is the desired result. □

4.3 Results for four rounds of the AES

For the AES, we use the four-round superbox structure that is explained in Section B.1. The four round result can then be amplified to more rounds using Lemma 5. Theorem 5 is obtained by plugging the result of Theorem 4 into Theorem 3.

Theorem 5 (AES, four rounds). *For the difference-distribution matrix D of four rounds of the AES, $\|D-U\|_2 \leq 2^{-39}$.*

Proof. The superbox layer of four round AES consists of four 32-bit S-boxes. Theorem 4 shows that $\sigma \leq 2^{-10.146}$ for each superbox. Plugging this into Theorem 3 with $q=2^{32}$, $n=4$ and $l=3$ yields $\|D-U\|_2 \leq 2^{-39}$. □

From the work of Keliher and Sui [19], we know that the key-averaged probability of nontrivial differentials over four AES rounds with independent round keys is bounded by $(53 \cdot 2^{-34})^4 \leq 2^{-113}$. The result now follows from Lemma 6.

Corollary 2. *The construction consisting of $4r+4$ rounds of the AES with independent and uniform random round keys is $2^{14} \cdot 2^{-39r}$-pairwise independent.*

Proof. Let D be the difference-distribution matrix of four rounds of the AES. From Lemma 6 with $p = 2^{-113}$ and $q = 2^{128}$, we get

$$\|D^{r+1} - U\|_1 \leq 2^{15} \|D^r - U\|_2 .$$

Finally, since $D^r - U = (D - U)^r$, we have $\|D^r - U\|_2 \leq \|D - U\|_2^r \leq 2^{-39r}$. The result now follows from Lemma 4. □

According to Corollary 2, to achieve 2^{-128}-pairwise independence, we need $20 = 4 + 4 \cdot 4$ rounds of the AES. This is a large improvement compared to the bound from Crypto 2021 [23, Theorem 3.14], where it was shown that $6r$-round AES is $(0.944)^r/2$-close to pairwise independence. Hence, over 9000 rounds would be required to achieve 2^{-128}-pairwise independence.

5 Truncated differentials

The goal of this section is to bound $\|\pi_V(D - U)i_V\|_2$ from above. In fact, we will show how to compute this quantity exactly, so the results in this section also yield a lower bound on $\|D - U\|_2$.

The starting point is the observation that all vectors v in V are eigenvectors of $(D^S)^{\otimes n}$ with eigenvalue one. This is due to Lemma 7. Intuitively, truncated differentials defined by an activity pattern are preserved by the S-box layer. Hence, if D^L is the difference-distribution matrix of L, then

$$\|\pi_V(D - U)i_V\|_2 = \|\pi_V(D^L - U)i_V\|_2 .$$

Indeed, $D = (D^S)^{\otimes n} D^L (D^S)^{\otimes n}$ as explained in Section 4.2.

5.1 Activity patterns and matrix coordinates

To compute or bound $\|\pi_V(D^L - U)i_V\|_2$, we first construct a matrix representation of the linear map $\pi_V(D - U)i_V$. Recall from Section 4.1 that V is the span of all indicators $\delta_{[z]}$ of difference sets $[z] \subseteq \mathbb{F}_q^n$ defined by activity patterns. In particular, $\dim V = 2^n$. Here we compute the matrix representation of $\pi_V(D^L - U)i_V$ relative to the following basis of V:

$$\left\{ \delta_{[z]} / \sqrt{q^{\operatorname{wt} z}} \;\middle|\; z \in \{0,1\}^n \right\},$$

This basis is nearly, but not quite, orthogonal. Let B be the $q^n \times 2^n$ matrix with columns equal to these basis vectors, and coordinates indexed by elements of $\mathbb{F}_q^n$ and $\{0,1\}^n$. Explicitly, we have $B_{x,z} = 1/\sqrt{q^{\operatorname{wt} z}}$ if $x \in [z]$ and zero otherwise. By Lemma 3, the matrix representation of $\pi_V(D^L - U)i_V$ is given by

$$(B^\intercal B)^{-1} B^\intercal (D^L - U) B . \tag{6}$$

The properties of the *Gram matrix* $B^\intercal B$ are important in our proof.

Gram matrix. The coordinates of the matrix $B^\intercal B$ are equal to

$$\left(B^\intercal B\right)_{y,x} = \left\langle \delta_{[y]}/\sqrt{q^{\operatorname{wt} y}},\, \delta_{[x]}/\sqrt{q^{\operatorname{wt} x}} \right\rangle = 1/\sqrt{q^{\operatorname{wt} x \oplus y}}\,,$$

for all x and y in $\{0,1\}^n$. Here, wt denotes the Hamming weight of a bitvector, and $x \oplus y$ is the exclusive or of x and y. Indeed, we have $\langle \delta_{[x]}, \delta_{[y]} \rangle = |[x] \cap [y]|$ and $2 \operatorname{wt} x \wedge y - \operatorname{wt} x - \operatorname{wt} y = -\operatorname{wt} x \oplus y$. Equivalently,

$$B^\intercal B = \begin{bmatrix} 1 & \frac{1}{\sqrt{q}} \\ \frac{1}{\sqrt{q}} & 1 \end{bmatrix}^{\otimes n} . \tag{7}$$

The first column corresponds to [0] and the second to [1]. In particular, we have the following lemma that plays a role in our proofs.

Lemma 8. *The smallest singular value of the matrix B is $(1 - 1/\sqrt{q})^{n/2}$.*

Proof. The singular values of B are the square roots of the eigenvalues of $B^\intercal B$. The eigenvalues of $B^\intercal B$ are n-fold products of eigenvalues of the 2×2 matrix in Equation (7). A trite calculation shows that these are $1 \pm 1/\sqrt{q}$. □

Matrix representation of $\boldsymbol{\pi_V(D^L - U)i_V}$. By Equation (6) and Lemma 8, the remaining problem is to compute the coordinates of $B^\intercal D^L B$ and $B^\intercal U B$. The latter case is straightforward, as shown by the following lemma.

Lemma 9. *Let x and y be activity patterns in $\{0,1\}^n$. The coordinates of $B^\intercal U B$ satisfy the equation*

$$\sqrt{q^{\operatorname{wt} x + \operatorname{wt} y}} \left(B^\intercal U B\right)_{y,x} = 1 + \frac{(q^{\operatorname{wt} x} - 1)(q^{\operatorname{wt} y} - 1)}{q^n - 1}\,.$$

Proof. By the definition of B, we have $\sqrt{q^{\operatorname{wt} x + \operatorname{wt} y}} \left(B^\intercal U B\right)_{y,x} = \langle \delta_{[y]}, U \delta_{[x]} \rangle$. As pointed out in the proof of Lemma 7, the matrix U is given by

$$U = \delta_0\, \delta_0^\intercal + \frac{1}{q^n - 1}\, \delta_{\mathbb{F}_q^n \setminus \{0\}}\, \delta_{\mathbb{F}_q^n \setminus \{0\}}^\intercal\,.$$

Since $\delta_{\mathbb{F}_q^n \setminus \{0\}}^\intercal \delta_{[x]} = q^{\operatorname{wt} x} - 1$, it follows that the inner product $\langle \delta_{[y]}, U \delta_{[x]} \rangle$ is equal to the desired result. □

To compute the coordinates of the matrix $B^\intercal D^L B$, we rely on the properties of MDS matrices and Lemma 1 in particular.

Lemma 10. *Let x and y be activity patterns in $\{0,1\}^n$. If $L \colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ is a linear function that is MDS with difference distribution matrix D^L, then the coordinates of $B^\intercal D^L B$ satisfy*

$$\sqrt{q^{\operatorname{wt} x + \operatorname{wt} y}} \left(B^\intercal D^L B\right)_{y,x} = \begin{cases} q^{\operatorname{wt} x + \operatorname{wt} y - n} & \text{if } \operatorname{wt} x + \operatorname{wt} y \geq n + 1\,, \\ 1 & \text{else}\,. \end{cases}$$

Proof. Since L is a linear function, D^L is a permutation matrix. In particular, $D^L_{b,a} = 1$ if $b = L(a)$ and zero elsewhere. Hence,

$$\sqrt{q^{\operatorname{wt} x + \operatorname{wt} y}} \left(B^\intercal D^L B\right)_{y,x} = \left\langle \delta_{[y]},\, D^L \delta_{[x]} \right\rangle = \left| [y] \cap L([x]) \right| .$$

Since the MDS property in Lemma 1 holds for every block, we can assume that $x = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$ and $y = (0, 0, \ldots, 0, 1, 1, \ldots, 1)$ without loss of generality. The number of elements in $[y] \cap L([x])$ is equal to number of solutions of the following system of linear equations, where coordinates indicated by '*' are arbitrary values in $\mathbb{F}_q$:

$$\underbrace{\left[\begin{array}{c|c} L_1 & L_2 \\ \hline L_3 & L_4 \end{array}\right]}_{L} \left[\begin{array}{c} * \\ \vdots \\ * \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline * \\ \vdots \\ * \end{array}\right] .$$

By Lemma 1, the block L_1 has full rank equal to $\min\{\operatorname{wt} x, n - \operatorname{wt} y\}$. If $n - \operatorname{wt} y \geq \operatorname{wt} x$, then the system only has the trivial solution. This corresponds to the case $\operatorname{wt} x + \operatorname{wt} y \leq n$ in the theorem. Otherwise, the dimension of the solution space is $\operatorname{wt} x - \operatorname{rank} L_1 = \operatorname{wt} x + \operatorname{wt} y - n$. That is, $|[y] \cap L([x])| = q^{\operatorname{wt} x + \operatorname{wt} y - n}$. □

5.2 Numerical bounds

For small n such as $n = 4$, the $2^n \times 2^n$ matrix $B^\intercal(D^L - U)B$ can be computed explicitly from Lemmas 9 and 10. This yields the exact two norm of $\pi_V(D^L - U)\, i_V$, since by Equation (6) the matrix representation of this map is given by $M = (B^\intercal B)^{-1} B^\intercal (D^L - U)B$. As mentioned in the paragraph following Lemma 3, some care should be taken: the norm of $\pi_V(D^L - U)i_V$ is equal to $\|R^{-1}MR\|_2$ rather than $\|M\|$, because the columns of B are not orthonormal. Here, R is an invertible $2^n \times 2^n$ matrix so that $RR^\intercal = (B^\intercal B)^{-1}$. Such a matrix exists because the Gram matrix $B^\intercal B$ is positive definite. Since BR is then a matrix with orthonormal columns,

$$\left\|\pi_V(D^L - U)i_V\right\|_2 = \left\|R^{-1}MR\right\|_2 = \left\|R^\intercal B^\intercal (D^L - U)BR\right\|_2 .$$

Table 1. Numerical values of $\|\pi_V(D^L - U)i_V\|_2$ for $n = 4$.

q	$\|\pi_V(D^L - U)i_V\|_2$	$\|\pi_V(D^L - U)i_V\|_{\mathrm{fr}}$
2^8	$2^{-13.351}$	$2^{-12.928}$
2^{32}	$2^{-61.415}$	$2^{-60.956}$

Some numerical results are given in Table 1. These results were computed using arbitrary precision arithmetic, although the reported numbers have been rounded up. Note that the Gram matrix $B^\intercal B$ is given in Equation (7).

Table 1 also lists the Frobenius norm of $\pi_V(D^L - U)i_V$, i.e., the square root of the sum of the squared matrix coordinates with respect to an orthonormal basis. Like for the two norm, it is enough to compute the Frobenius norm of the matrix $R^\intercal B^\intercal (D^L - U)BR$. In general, $\|\pi_V(D^L - U)i_V\|_2 \leq \|\pi_V(D^L - U)i_V\|_{\mathsf{fr}}$. The reason for including the Frobenius norm in Table 1 is that we use it to obtain the closed-form bounds in Section 5.3. The results in Table 1 show that for $n = 4$, less than a factor of two is lost by making this simplification.

5.3 Closed-form bounds

To obtain a closed-form bound on the two norm, we bound the Frobenius norm. This results in the following theorem, which is used in the proof of Theorem 3.

Theorem 6. *Let $L\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ be a linear MDS map with difference-distribution matrix D^L. If $q \geq 4$ and $n \geq 2$, then $\|\pi_V(D^L - U)i_V\|_2 \leq \big(2/(\sqrt{q} - 1)\big)^n$.*

Proof. By Lemma 8, we have $\|(B^\intercal B)^{-1}\|_2 \leq 1/(1 - 1/\sqrt{q})^n$. Indeed, the norm of the inverse of a matrix is the reciprocal of its smallest singular value. Hence,

$$\begin{aligned}\|\pi_V(D^L - U)i_V\|_2 &= \|R^\intercal B^\intercal (D^L - U)BR\|_2 \\ &\leq \|(B^\intercal B)^{-1}\|_2 \, \|B^\intercal (D^L - U)B\|_2 \\ &\leq \|B^\intercal (D^L - U)B\|_{\mathsf{fr}}/(1 - 1/\sqrt{q})^n \,.\end{aligned}$$

In the remainder of the proof, we bound $\|B^\intercal (D^L - U)B\|_{\mathsf{fr}}$. For notational convenience, let $\Delta \coloneqq B^\intercal (D^L - U)B$. For all bitvectors x and y with $\operatorname{wt} x + \operatorname{wt} y \leq n$, Lemmas 9 and 10 show that

$$\Delta_{y,x}^2 = \frac{1}{q^{\operatorname{wt} x + \operatorname{wt} y}} \left(\frac{(q^{\operatorname{wt} x} - 1)(q^{\operatorname{wt} y} - 1)}{q^n - 1} \right)^2 \leq \frac{q^{\operatorname{wt} x + \operatorname{wt} y}}{(q^n - 1)^2} \leq \frac{q^n}{(q^n - 1)^2} \,.$$

For the case $\operatorname{wt} x + \operatorname{wt} y \geq n + 1$, first observe that $\Delta_{y,x} = 0$ if $\operatorname{wt} x = \operatorname{wt} y = n$. Indeed, in this case $(B^\intercal D^L B)_{y,x} = 1 = (B^\intercal U B)_{y,x}$. Assuming $\operatorname{wt} x + \operatorname{wt} y \leq 2n - 1$, we have $\Delta_{y,x} \leq 0$ and

$$\begin{aligned}\sqrt{q^{\operatorname{wt} x + \operatorname{wt} y}} \, |\Delta_{y,x}| &= 1 + \frac{(q^{\operatorname{wt} x} - 1)(q^{\operatorname{wt} y} - 1)}{q^n - 1} - q^{\operatorname{wt} x + \operatorname{wt} y - n} \\ &= \frac{(q^n - 1) + (q^{\operatorname{wt} x} - 1)(q^{\operatorname{wt} y} - 1) - (q^n - 1)q^{\operatorname{wt} x + \operatorname{wt} y - n}}{q^n - 1} \\ &= \frac{q^n - q^{\operatorname{wt} x} - q^{\operatorname{wt} y} + q^{\operatorname{wt} x + \operatorname{wt} y - n}}{q^n - 1} \\ &\leq \frac{q^n + q^{\operatorname{wt} x + \operatorname{wt} y - n}}{q^n - 1} \,.\end{aligned}$$

After squaring, we obtain (using $n+1 \le \operatorname{wt} x + \operatorname{wt} y \le 2n-1$)

$$\Delta_{y,x}^2 \le \frac{1}{q^{\operatorname{wt} x+\operatorname{wt} y}} \left[\frac{q^n + q^{\operatorname{wt} x+\operatorname{wt} y-n}}{q^n-1}\right]^2 \le \frac{q^{2n}+2q^{2n-1}+q^{2n-2}}{q^{n+1}(q^n-1)^2} \le \frac{q^n}{2(q^n-1)^2}\,,$$

where the second inequality requires $q \ge 4$. The square of the Frobenius norm of Δ is bounded by $\|\Delta\|_{\mathsf{fr}}^2 = \sum_{x,y} \Delta_{y,x}^2 \le 2^{2n}\, q^n/(q^n-1)^2$, because the total number of activity patterns is 2^{2n}. This can still be improved by a factor $7/8 = 1/2{\cdot}(1{+}3/4)$ because less than $3/4$ of all activity patterns satisfy $\operatorname{wt} x{+}\operatorname{wt} y \le n$. This yields the overall bound

$$\|\Delta\|_{\mathsf{fr}}^2 \le 2^{2n}\,\frac{7}{8}\,\frac{q^n}{(q^n-1)^2} = \left(\frac{4}{q}\right)^n \cdot \frac{7}{8}\left(\frac{q^n}{q^n-1}\right)^2\,.$$

For $q \ge 4$ and $n \ge 2$, we have $q^n/(q^n-1) \le \sqrt{8/7}$. Hence, the overall bound is

$$\|\pi_V(D-U)i_V\|_2 \le \frac{1}{(1-1/\sqrt{q})^n} \cdot \left(\frac{2}{\sqrt{q}}\right)^n = \left(\frac{2}{\sqrt{q}-1}\right)^n\,.$$

This concludes the proof. □

6 Spectral terms

The goal of this section is to bound the three terms $\|\pi_V(D-U)i_{V^\perp}\|_2$, $\|\pi_{V^\perp}(D-U)i_V\|_2$ and $\|\pi_{V^\perp}(D-U)i_{V^\perp}\|_2$ in Equation (5). The bound depends on the third singular value of the difference-distribution matrix D^S of the S-box.

A first observation is that $\pi_V(D-U)i_{V^\perp} = \pi_V D\, i_{V^\perp}$, $\pi_{V^\perp}(D-U)i_V = \pi_{V^\perp} D\, i_V$ and $\pi_{V^\perp}(D-U)i_{V^\perp} = \pi_{V^\perp} D\, i_{V^\perp}$. This is due to the fact that $Uv = 0$ for all v in $V^\perp$, i.e. $V^\perp \subseteq \ker U$. The idea of our proof is to decompose $V^\perp$ as a sum of subspaces, each one essentially corresponding to a different number of S-boxes with an input difference state that is neither zero nor saturated.

6.1 Decomposition of $V^\perp$

Lemma 7 shows that the largest two singular values of D^S are equal to one, with corresponding singular vectors to an orthogonal basis for $A := \operatorname{Span}\{\delta_0, \delta_{\mathbb{F}_q}\}$. Since the subspace $V \subseteq \mathbb{R}[\mathbb{F}_q^n]$ is by definition the span of all vectors $w_1 \otimes w_2 \otimes \cdots \otimes w_n$ with w_i in A, the subspace $V^\perp$ is the span of all vectors $w_1 \otimes w_2 \otimes \cdots \otimes w_n$ with at least one of the vectors w_i is orthogonal to δ_0 and $\delta_{\mathbb{F}_q}$. Explicitly,

$$V^\perp = \operatorname{Span}\Big\{ w_1 \otimes \cdots \otimes w_n \;\Big|\; \exists\, i\colon w_i \in A^\perp \Big\}\,.$$

Orthogonal decomposition of $V^\perp$. In our proof, we make a case distinction based on the number of vectors w_i that are orthogonal to A. To enable this, we define vector spaces W_l as follows:

$$W_l = \mathrm{Span}\left\{ w_1 \otimes \cdots \otimes w_n \;\middle|\; \exists I \subseteq \{1,\ldots,n\}, |I| = l \colon \begin{cases} w_i \in A^\perp & \text{for } i \in I \\ w_i \in A & \text{for } i \notin I \end{cases} \right\}.$$

In particular, $W_0 = V$. It is not difficult to see that $W_l \perp W_k$ whenever $l \neq k$. Indeed, suppose $l < k$. For every $w_1 \otimes \cdots \otimes w_n$ in W_l and $w'_1 \otimes \cdots \otimes w'_n$ in W_k, there is at least one i so that $w_i \in A$ and $w'_i \in A^\perp$. Hence, $w_1 \otimes \cdots \otimes w_n$ and $w'_1 \otimes \cdots \otimes w'_n$ are orthogonal.

It follows from the above, and because every $w_1 \otimes \cdots \otimes w_n$ contains between one and n factors orthogonal to A, that $V^\perp = W_1 \oplus W_2 \oplus \cdots \oplus W_n$. The significance of this decomposition is that if $w_1 \otimes \cdots \otimes w_n \in W_l$, then

$$\left\| \left(D^S\right)^{\otimes n} w_1 \otimes \cdots \otimes w_n \right\|_2 = \left\| D^S w_1 \otimes \cdots \otimes D^S w_n \right\|_2 \leq \sigma^l \left\| w_1 \otimes \cdots \otimes w_n \right\|_2, \quad (8)$$

with σ the third singular value of D^S. Furthermore, Equation (8) also holds for every vector in W_l. In fact, since $D = \left(D^S\right)^{\otimes n} D^L \left(D^S\right)^{\otimes n}$, this inequality implies $\|\pi_V D i_{V^\perp}\|_2 \leq \sigma$, $\|\pi_{V^\perp} D i_V\|_2 \leq \sigma$ and $\|\pi_{V^\perp} D i_{V^\perp}\|_2 \leq \sigma^2$. For the last two inequalities, observe that inequality (8) also holds for the transpose of D^S. However, it turns out that these bounds can be improved significantly by analyzing the norm on the spaces $W_1, \ldots, W_l$ separately.

Basis for W_l. In Section 6.3, we will use a particular basis for the subspaces W_l. Like in Section 5, our analysis requires the singular values of the Gram matrix for this basis. Like in Section 5.1, the bases we choose will be almost but not quite orthogonal. Specifically, consider the basis consisting of vectors $w_1 \otimes \cdots \otimes w_n$ so that precisely l vectors w_i are elements of an orthonormal basis for $\mathrm{Span}\{\delta_0, \delta_{\mathbb{F}_q}\}^\perp$, and the other vectors w_i are either δ_0 or $\delta_{\mathbb{F}_q}/\sqrt{q}$. Let B_i be the matrix whose columns are equal to these basis vectors for W_i respectively.

Lemma 11. *The smallest singular value of the matrix B_l is $(1 - 1/\sqrt{q})^{(n-l)/2}$.*

Proof. This result generalizes Lemma 8, which corresponds to the case $l = 0$. For every column $w_1 \otimes \cdots \otimes w_n$ of B_l, there exists an index set $I \subset \{1,\ldots,n\}$ of size l so that w_i is orthogonal to δ_0 and $\delta_{\mathbb{F}_q}$ for all i in I. Columns with different index sets are orthogonal.

Hence, up to a permutation of rows and columns, the Gram matrix $B_l^\intercal B_l$ is a block diagonal matrix. Every diagonal block corresponds to a different choice of the index set. Up to a permutation of factors, the diagonal blocks are

$$I^{\otimes l} \otimes \begin{bmatrix} 1 & \frac{1}{\sqrt{q}} \\ \frac{1}{\sqrt{q}} & 1 \end{bmatrix}^{\otimes (n-l)}$$

Like in the proof of Lemma 8, the result follows from the fact that the eigenvalues of the 2×2 matrix are equal to $1 \pm 1/\sqrt{q}$. ∎

6.2 Vanishing results on W_1

As a first improvement to the bounds $\|\pi_V D i_{V^\perp}\|_2 \le \sigma$ and $\|\pi_{V^\perp} D i_V\|_2 \le \sigma$ from Section 6.1, we prove that $\pi_V D i_{W_1} = 0$ and $\pi_{W_1} D i_V = 0$. As shown below, this immediately improves these bounds to σ^2. We prove this result in the following form.

Lemma 12. *Let $L\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ be linear with difference-distribution matrix D^L. Let $w_1, \dots, w_{2n}$ be vectors in $\mathbb{R}[\mathbb{F}_q]$. Assume there are k vectors w_i with $w_i = \delta_0$ and $l \ge 1$ vectors $w_i \perp \{\delta_0, \delta_{\mathbb{F}_q}\}$. If L is MDS, and $k \ge n$ or $k + l \le n$, then*

$$\Big\langle w_1 \otimes \cdots \otimes w_n,\, D^L(w_{n+1} \otimes \cdots \otimes w_{2n}) \Big\rangle = 0\,.$$

Proof. The order of $w_1, \dots, w_{2n}$ does not play a role in the proof, so we can assume without loss of generality that $w_1, \dots, w_l$ are orthogonal to both δ_0 and $\delta_{\mathbb{F}_q}$. We can assume that $w_{l+1}, \dots, w_{2n}$ are either equal to δ_0 or $\delta_{\mathbb{F}_q}$.

Let $C \subseteq \mathbb{F}_q^{2n}$ be the vector space consisting of all $x = y \| L(y)$ with y in $\mathbb{F}_q^n$ so that $x_i = 0$ whenever $w_i = \delta_0$. If $w_i = \sum_{x \in \mathbb{F}_q} w_{i,x}\, \delta_x$, then

$$\Big\langle w_1 \otimes \cdots \otimes w_n,\, D^L(w_{n+1} \otimes \cdots \otimes w_{2n}) \Big\rangle = \sum_{x \in C} w_{1,x_1} w_{2,x_2} \cdots w_{l,x_l}\,.$$

Indeed, the nonzero coordinates of $w_{l+1}, \dots, w_{2n}$ are all one. If $k \ge n$, then $\dim C = 0$ and the sum is zero because $w_{i,0} = 0$ for $i \le l$. Otherwise, $\dim C = n - k$ by the MDS property. If $n - k \ge l$, then the MDS property ensures that the coordinates $(x_1, \dots, x_l)$ take all possible values in $\mathbb{F}_q^l$ an equal number of times. This implies the sum is zero, because (for all $i \le l$)

$$\sum_{x_i \in \mathbb{F}_q} w_{1,x_i} = \big\langle w_i, \delta_{\mathbb{F}_q} \big\rangle = 0\,.$$

Hence, the inner product vanishes. □

Lemma 12 also applies to the matrix D rather than D^L, because $(D^S)^{\otimes n}$ preserves the structure of the vectors $w_1 \otimes \cdots \otimes w_n$ and $w_{n+1} \otimes \cdots \otimes w_{2n}$. For $l = 1$, at least one of the conditions $k \ge n$ and $k + l \le n$ is always true. Hence, $\pi_V D\, i_{W_1} = 0$ and $\pi_{W_1} D\, i_V = 0$. If $W_{\ge 2} = W_2 \oplus W_3 \oplus \cdots \oplus W_n$, then we have

$$\|\pi_V D\, i_{V^\perp}\|_2 \le \|\pi_V D\, i_{W_1}\|_2 + \|\pi_V D\, i_{W_{\ge 2}}\|_2 = \|\pi_V D\, i_{W_{\ge 2}}\|_2 \le \sigma^2\,.$$

Likewise,

$$\|\pi_{V^\perp} D\, i_V\|_2 \le \|\pi_{W_1} D\, i_V\|_2 + \|\pi_{W_{\ge 2}} D\, i_V\|_2 = \|\pi_{W_{\ge 2}} D\, i_V\|_2 \le \sigma^2\,.$$

This yields the overall bound $\|D - U\|_2^2 \le \|\pi_V (D^L - U)\, i_V\|_2^2 + 3\,\sigma^4$. Section 6.3 shows how these bounds can be substantially improved by analyzing $W_2, W_3 \dots, W_n$ in more detail.

6.3 Norm bounds on W_l with $l \geq 2$

To bound the two norm of linear maps such as $\pi_{W_l} D\, i_V$ for $l \geq 2$, we rely on the Frobenius norm. This is easier, because it allows us to focus on bounding individual coordinates.

Lemma 13. *Let l_1 and l_2 be non-negative integers so that $l_1+l_2 = l$. If the linear map $L\colon \mathbb{F}_q^n \to \mathbb{F}_q^n$ is MDS, then its difference-distribution matrix D^L satisfies*

$$\left\|\pi_{W_{l_1}} D^L\, i_{W_{l_2}}\right\|_2^2 \leq C(n,l_1,l_2)\, \left(\frac{1}{\sqrt{q}-1}\right)^{2n-2l},$$

where $C(n,l_1,l_2)$ depends only weakly on q:

$$C(n,l_1,l_2) = \left(1+\frac{1}{\sqrt{q}-1}\right)^{l} \binom{n}{l_1}\binom{n}{l_2} \sum_{i=1}^{l-1} \binom{2n-l}{n-i} 2^i\,.$$

Proof. To bound the Frobenius norm, we use the basis for the spaces W_{l_1} and W_{l_2} that was described in Section 6.1. Let B_{l_1} and B_{l_2} be matrices whose columns are equal to these basis vectors for W_{l_1} and W_{l_2} respectively. Similar to the argument in the proof of Theorem 6, we have

$$\begin{aligned}\left\|\pi_{W_{l_1}} D^L i_{W_{l_2}}\right\|_2 &= \left\|B_{l_1}(B_{l_1}^\intercal B_{l_1})^{-1} B_{l_1}^\intercal D^L B_{l_2} (B_{l_2}^\intercal B_{l_2})^{-1} B_{l_2}^\intercal\right\|_2 \\ &\leq \|B_{l_1}(B_{l_1}^\intercal B_{l_1})^{-1}\|_2\, \|(B_{l_2}^\intercal B_{l_2})^{-1} B_{l_2}^\intercal\|_2\, \left\|B_{l_1}^\intercal D^L B_{l_2}\right\|_{\mathsf{fr}}.\end{aligned}$$

The singular values of $B_{l_1}(B_{l_1}^\intercal B_{l_1})^{-1}$ and $(B_{l_2}^\intercal B_{l_2})^{-1}B_{l_2}$ are the reciprocals of those of B_{l_1} and B_{l_2}, respectively. By Lemma 11, the singular values of B_{l_i} are at least $(1-1/\sqrt{q})^{(n-l_i)/2}$. Hence, $\|\pi_{W_{l_1}} D^L i_{W_{l_2}}\|_2^2 \leq \|B_{l_1}^\intercal D^L B_{l_2}\|_{\mathsf{fr}}^2/(1-1/\sqrt{q})^{2n-l}$. Let $w_1\otimes\cdots\otimes w_n$ and $w_{n+1}\otimes\cdots\otimes w_{2n}$ be basis vectors of W_{l_1} and W_{l_2} respectively. To bound the Frobenius norm, we bound the squared inner products

$$\left|\left\langle w_1\otimes\cdots\otimes w_n, D^L\left(w_{n+1}\otimes\cdots\otimes w_{2n}\right)\right\rangle\right|^2.$$

Without loss of generality, assume $w_1,\ldots,w_l$ are orthogonal to δ_0 and $\delta_{\mathbb{F}_q}$ (the bound does not depend on the order of $w_1,\ldots,w_{2n}$). Let k be the number of vectors w_i so that $w_i = \delta_0$. By Lemma 12, we can assume $k \leq n-1$ and $k+l \geq n+1$. This leaves $2n-k-l \leq n$ vectors w_i equal to $\delta_{\mathbb{F}_q}/\sqrt{q}$. As in the proof of Lemma 12, let $C \subseteq \mathbb{F}_q^{2n}$ be the subspace consisting of all vectors $x = y\|L(y)$ for y in $\mathbb{F}_q^n$ so that $x_i = 0$ when $w_i = \delta_0$. The inner product equals

$$\left|\left\langle \bigotimes_{i=1}^n w_i, D^L \bigotimes_{i=n+1}^{2n} w_i\right\rangle\right|^2 = \frac{1}{q^{2n-k-l}}\left(\sum_{x\in C} w_{1,x_1} w_{2,x_2}\cdots w_{l,x_l}\right)^2.$$

The dimension of C is equal to $\dim C = n-k$. The Frobenius norm is the sum of these inner products over all choices of the basis vectors. First consider the

choice of the vectors $w_1, \ldots, w_l$:

$$\sum_{w_1,\ldots,w_l} \left|\langle \bigotimes_{i=1}^{n} w_i, D^L \bigotimes_{i=n+1}^{2n} w_i \rangle\right|^2$$
$$\leq \frac{1}{q^{2n-k-l}} \sum_{x,y\in C} \left(\sum_{w_1} w_{1,x_1} w_{1,y_1}\right)\left(\sum_{w_2} w_{2,x_2} w_{2,y_2}\right)\cdots\left(\sum_{x_l} w_{l,x_l} w_{l,y_l}\right).$$

The basis vectors w_i are an orthonormal basis for $\mathrm{Span}\{\delta_0, \delta_{\mathbb{F}_q}\}^\perp$. Each of the factors above is a sum over the rows of a matrix whose columns are the basis vectors w_i. If $x_i \neq 0$ and $y_i \neq 0$, then

$$\sum_{w_i} w_{i,x_i} w_{i,y_i} = -\frac{1}{q-1} + \begin{cases} 1 & \text{if } x_i = y_i\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

The term $-1/(q-1)$ is due to the fact that we sum over a complete orthonormal basis for $\mathbb{R}[\mathbb{F}_q]$ *except* δ_0 and $(\delta_{\mathbb{F}_q} - \delta_0)/\sqrt{q-1}$. Hence, if $\delta_{x_i,y_i} = 1$ for $x_i = y_i$ and zero otherwise, then

$$\sum_{w_1,\ldots,w_l} \left|\langle \bigotimes_{i=1}^{n} w_i, D^L \bigotimes_{i=n+1}^{2n} w_i \rangle\right|^2 \leq \frac{1}{q^{2n-k-l}} \sum_{\substack{x,y\in C \\ x_1,\ldots,x_l\neq 0 \\ y_1,\ldots,y_l\neq 0}} \prod_{i=1}^{l} \Big(\delta_{x_i,y_i} - \frac{1}{q-1}\Big)\,.$$

Since $\dim C = n-k$ and due to the MDS property, the sum is over all possible values for the first $n-k$ coordinates of x and y. If t of the first $n-k$ coordinates satisfy $x_i = y_i$, then $\prod_{i=1}^{l} |\delta_{x_i,y_i} - 1/(q-1)| \leq 1/(q-1)^{n-k-t}$. For a fixed choice of t indices i, there are $(q-1)^{2(n-k)-t}$ choices for x and y. Hence,

$$\sum_{\substack{x,y\in C \\ x_1,\ldots,x_l\neq 0 \\ y_1,\ldots,y_l\neq 0}} \prod_{i=1}^{l} \Big(\delta_{x_i,y_i} - \frac{1}{q-1}\Big) \leq \sum_{t=0}^{n-k} \binom{n-k}{t} (q-1)^{n-k} = 2^{n-k}\,(q-1)^{n-k}\,. \quad (9)$$

The overall bound for the squared Frobenius norm is obtained by summing over all $\binom{2n-l}{k}$ choices for the k vectors $w_i = \delta_0$, the choice of $n+1-l \leq k \leq n-1$ and the choice of the l vectors orthogonal to both δ_0 and $\delta_{\mathbb{F}_q}$:

$$\left\|B_{l_1}^{\intercal} D^L B_{l_2}\right\|_{\mathsf{fr}}^2 \leq \binom{n}{l_1}\binom{n}{l_2} \sum_{k=n+1-l}^{n-1} \frac{(q-1)^{n-k}}{q^{2n-k-l}}\, 2^{n-k} \binom{2n-l}{k}\,.$$

Since $(q-1)^{n-k}/q^{2n-k-l} \leq q^{l-n}$, we have

$$\left\|\pi_{W_{l_1}} D^L i_{W_{l_2}}\right\|_2^2 \leq \frac{q^{l-n}}{(1-1/\sqrt{q})^{2n-l}} \binom{n}{l_1}\binom{n}{l_2} \sum_{k=n+1-l}^{n-1} 2^{n-k} \binom{2n-l}{k}\,.$$

The result follows from $q^{l-n}/(1-1/\sqrt{q})^{2n-l} = (1+1/(\sqrt{q}-1))^l/(\sqrt{q}-1)^{2n-2l}$ and the substitution $i = n-k$. □

The result of Lemma 13 implies the following bound on $\|\pi_{W_{l_1}} D i_{W_{l_2}}\|_2$:

$$\left\|\pi_{W_{l_1}} D\, i_{W_{l_2}}\right\|_2^2 \leq C(n,l_1,l_2)\,\sigma^{2l}\left(\frac{1}{\sqrt{q}-1}\right)^{2n-2l}.$$

Asymptotically for large q, $C(n,l_1,l_2) = \mathcal{O}(1)$, so the norm is bounded by $\mathcal{O}(\sigma^l/\sqrt{q}^{n-l})$. As shown in Corollary 1, this yields an optimal result for $\sigma = \mathcal{O}(1/\sqrt{q})$. Concretely, the factor $C(n,l_1,l_2)$ is quite important and Lemma 13 admits some improvement in this regard. The following remark illustrates this for $l = 2$, the smallest meaningful choice of l. This result will also be useful for the AES superbox.

Remark 1. For $l = 2$, Lemma 13 yields the following bound:

$$\left\|\pi_{W_{l_1}} D^L\, i_{W_{l_2}}\right\|_2^2 \leq \left(\frac{1}{\sqrt{q}-1}\right)^{2n-4}\left(1+\frac{1}{\sqrt{q}-1}\right)^2 2\binom{n}{l_1}\binom{n}{l_2}\binom{2n-2}{n-1}.$$

There is only one choice for k in the proof of Lemma 13, namely $k = n-1$. This means that $\dim C = 1$ and inequality (9) can be improved to the equality

$$\sum_{\substack{x,y\in C\\ x_1,\dots,x_l\neq 0\\ y_1,\dots,y_l\neq 0}}\prod_{i=1}^{l}\left(\delta_{x_i,y_i}-\frac{1}{q-1}\right) = \sum_{x_1,y_1\in\mathbb{F}_q^\times}\left(\delta_{x_1,y_1}-\frac{1}{q-1}\right)\left(\delta_{c\,x_1,c\,y_1}-\frac{1}{q-1}\right),$$

for some nonzero c. Explicit computation yields (split into $x_1 = y_1$ and $x_1 \neq y_1$)

$$(q-1)\left(1-\frac{1}{q-1}\right)^2+\frac{(q-1)(q-2)}{(q-1)^2} = \frac{q-2}{q-1}(q-2+1) = q-2\,.$$

Hence, following the proof of Lemma 13, we have (for $l_1+l_2=2$)

$$\left\|\pi_{W_{l_1}} D^L\, i_{W_{l_2}}\right\|_2^2 \leq \frac{q-2}{(\sqrt{q}-1)^{2n-2}}\binom{n}{l_1}\binom{n}{l_2}\binom{2n-2}{n-1}.$$

For large q, this improves the bound by roughly a factor of two. ◁

6.4 Overall bound

The following theorem gives a closed-form bound on $\|\pi_V D i_{V^\perp}\|_2$, $\|\pi_{V^\perp} D i_V\|_2$ and $\|\pi_{V^\perp} D i_{V^\perp}\|_2$. This result is used in the proof of Theorem 3. Finally, we discuss numerical improvements to this bound for the AES superbox.

Theorem 7. *Let D be the difference-distribution matrix of* Shark$[n,q,S,L]$*. If the third singular value of the difference-distribution matrix of S is at most σ, then for all $1 \leq l \leq n-1$,*

$$\begin{aligned}&\|\pi_V D i_{V^\perp}\|_2^2 + \|\pi_{V^\perp} D i_V\|_2^2 + \|\pi_{V^\perp} D i_{V^\perp}\|_2^2\\ &\leq (l+2)\sigma^{2(l+1)} + \sum_{k=2}^{l} C(n,k)\,\frac{\sigma^{2k}}{(\sqrt{q}-1)^{2n-2k}}\,,\end{aligned}$$

where $C(n,k)$ depends only weakly on q:

$$C(n,k) = \left(1 + \frac{1}{\sqrt{q}-1}\right)^k \binom{2n}{k} \sum_{i=1}^{k-1} \binom{2n-k}{n-i} 2^i \,.$$

Proof. For the first two terms, use the decomposition $V^\perp = W_1 \oplus \cdots \oplus W_l \oplus W_{>l}$. For the third term, let $V^\perp = W_1 \oplus \cdots \oplus W_i \oplus W_{>i}$, and then for each choice of $i \leq l-1$ let $V^\perp = W_1 \oplus \cdots \oplus W_{l-i} \oplus W_{>l-i}$. This yields the following bound:

$$\sum_{i=2}^{l} \|\pi_{W_i} Di_V\|_2^2 + \sum_{i=2}^{l} \|\pi_V Di_{W_i}\|_2^2 + \sum_{i=1}^{l-1} \sum_{j=1}^{l-i} \|\pi_{W_i} Di_{W_j}\|_2^2 + (l+2)\sigma^{2(l+1)} \,.$$

The term $(l+2)\sigma^{2(l+1)}$ is an upper bound on the sum of $\|\pi_{W_{>l}} Di_V\|_2^2$, $\|\pi_V Di_{W_{>l}}\|_2^2$ and $\sum_{i=1}^{l} \|\pi_{W_{\geq i}} Di_{W_{>l-i}}\|_2^2$. This already proves the result for $l = 1$. By Lemma 13, the first two terms above are bounded by $\sum_{k=2}^{l} C(n,k,0)\,\sigma^{2k}/(\sqrt{q}-1)^{2n-2k}$. The third term is bounded by

$$\sum_{i=1}^{l-1} \sum_{j=1}^{l-i} C(n,i,j) \frac{\sigma^{2(i+j)}}{(\sqrt{q}-1)^{2n-2i-2j}} = \sum_{k=2}^{l} \frac{\sigma^{2k}}{(\sqrt{q}-1)^{2n-2k}} \sum_{i=1}^{k-1} C(n,i,k-i) \,.$$

This can be simplified using the identity $\sum_{i=1}^{k-1} \binom{n}{i}\binom{n}{k-i} = \binom{2n}{k} - 2\binom{n}{k}$. In particular, the inner sum simplifies to

$$\sum_{i=1}^{k-1} C(n,i,k-i) = \left(1 + \frac{1}{\sqrt{q}-1}\right)^k \binom{2n}{k} \sum_{i=1}^{k-1} \binom{2n-k}{n-i} 2^i - 2C(n,k,0)$$

Denote the first term by $C(n,k)$. With this notation, the overall bound becomes

$$(l+2)\sigma^{2(l+1)} + \sum_{k=2}^{l} C(n,k) \frac{\sigma^{2k}}{(\sqrt{q}-1)^{2n-2k}} \,.$$

This is the desired result. □

For the AES superbox, $n = 4$, $q = 2^8$ and $\sigma = 9/128$. For these parameters, Theorem 7 with $l = 2$ yields the bound $(2^{-9.897})^2$. This result can be improved based on Remark 1. In particular, for $l = 2$, we have the following bounds:

$$\begin{aligned} \|\pi_{V^\perp} Di_V\|_2^2, \|\pi_V Di_{V^\perp}\|_2^2 &\leq \sigma^6 + \frac{(q-2)\,\sigma^4}{(\sqrt{q}-1)^{2n-2}} \binom{n}{2}\binom{2n-2}{n-1} && \leq (2^{-11.178})^2 \,, \\ \|\pi_{V^\perp} Di_{V^\perp}\|_2^2 &\leq 2\sigma^6 + \frac{(q-2)\,\sigma^4}{(\sqrt{q}-1)^{2n-2}}\, n^2 \binom{2n-2}{n-1} && \leq (2^{-10.598})^2 \,. \end{aligned}$$

For large q, such as for four rounds of the AES with 32-bit superboxes, Theorem 7 can be used directly with $l = 3$. As shown in the proof of Theorem 5, this yields the bound $2^{-39.29}$.

7 Future work: a new S-box criterion

All of our pairwise independence bounds, including Theorem 3 and corollary 2, depend on the third singular value of the difference-distribution matrix of the S-box. For a permutation F, let $\sigma(F)$ denote the third largest singular value of its DDT. We call this important parameter the *spectral uniformity* of F.

Small S-boxes. Since composing a function F with affine bijections only permutes the rows and columns of the DDT, $\sigma(F)$ is invariant under affine equivalence. This allows us to compute the spectral uniformity for all 4-bit S-boxes. As listed in [10], up to affine equivalence, there are 302 permutations on $\mathbb{F}_2^4$. Computing the spectral uniformity for all 4-bit permutations is therefore feasible.

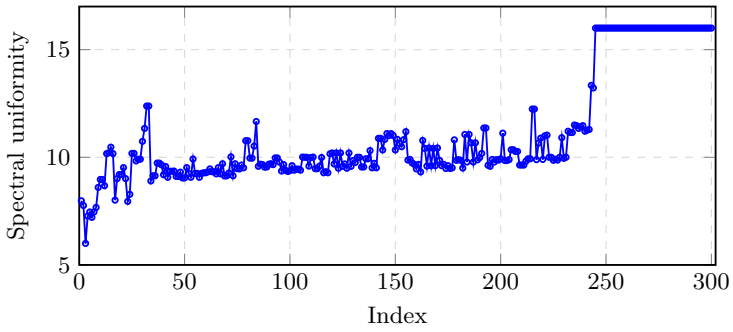

Fig. 4. Spectral uniformity of all 4-bit S-boxes.

In Figure 4, the 302 equivalence classes of S-boxes are listed in the same order as in [10]. That is, they are sorted according to their DDT and LAT spectra. The magnitude of DDT and LAT entries is correlated with the spectral uniformity. A more interesting observation is that there is exactly one function that achieves the minimum spectral uniformity of 6, namely the inverse function.

For $n \geq 5$, it is infeasible to list all permutations up to affine equivalence. We computed the spectral uniformity for some cryptographically interesting functions, in particular the only known class of APN permutations and the inverse function. Among those, the inverse again had the smallest spectral uniformity.

Inverse function. For odd n, we proved a bound on the spectral uniformity in Section A, which we used in Section 3.2. This proof does not cover the case of even n, for which the inverse function is not APN. Based on numerical experiments, we conjecture that the spectral uniformity of the inverse is $2^{n/2} + 2$.

An interesting open problem is to find a function with spectral uniformity smaller than that of the inverse, or to prove that no such function exists. The latter problem appears to be difficult at first sight; a more approachable special case would be to answer this question for power permutations.

Acknowledgements

Our basic methodology was introduced in 2020 by the first author [3] to analyze truncated differential attacks on the AES, proposed by Sondre Rønjom in 2019 [3,30]. We thank him for his questions. These results were presented during the workshop 'New Directions in Cryptanalysis of AES', at Bar-Ilan university in September 2022. The application to pairwise independence appeared in the master's thesis of Immo Schütt [31]. We thank Itai Dinur for pointing out an error in the numerical calculations in an earlier version of Section 5.2. We also thank Zhongfeng Niu, Kai Hu, Thomas Peyrin, Haolin Liu, and Meiqin Wang for pointing out a missing term in the previous version of Theorem 7.

Tim Beyne is supported by a junior postdoctoral fellowship from the Research Foundation – Flanders (FWO) with reference number 1274724N. This work was in part supported by the European Research Council (ERC) project 101097056 (SYMTRUST).

References

1. Advanced Encryption Standard (AES). National Institute of Standards and Technology, NIST FIPS PUB 197, U.S. Department of Commerce (Nov 2001)
2. Albrecht, M.R., Grassi, L., Rechberger, C., Roy, A., Tiessen, T.: MiMC: Efficient encryption and cryptographic hashing with minimal multiplicative complexity. In: Cheon, J.H., Takagi, T. (eds.) ASIACRYPT 2016, Part I. LNCS, vol. 10031, pp. 191–219. Springer, Berlin, Heidelberg (Dec 2016). https://doi.org/10.1007/978-3-662-53887-6_7
3. Beyne, T.: Note on the probability of truncated differentials. Personal communication with Sondre Rønjom and Vincent Rijmen (July 2020)
4. Beyne, T.: A geometric approach to linear cryptanalysis. In: Tibouchi, M., Wang, H. (eds.) ASIACRYPT 2021, Part I. LNCS, vol. 13090, pp. 36–66. Springer, Cham (Dec 2021). https://doi.org/10.1007/978-3-030-92062-3_2
5. Beyne, T.: A geometric approach to symmetric-key cryptanalysis. Ph.D. thesis, KU Leuven (June 2023)
6. Beyne, T.: Lecture 1: Introduction to the geometric approach. Lecture series at Shandong University (September 2025), https://www.bilibili.com/video/BV1eEaozJEDh/
7. Beyne, T., Rijmen, V.: Differential cryptanalysis in the fixed-key model. In: Dodis, Y., Shrimpton, T. (eds.) CRYPTO 2022, Part III. LNCS, vol. 13509, pp. 687–716. Springer, Cham (Aug 2022). https://doi.org/10.1007/978-3-031-15982-4_23
8. Beyne, T., Verbauwhede, M.: Integral cryptanalysis in characteristic p. Cryptology ePrint Archive, Paper 2025/932 (2025), https://eprint.iacr.org/2025/932
9. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. In: Menezes, A.J., Vanstone, S.A. (eds.) CRYPTO'90. LNCS, vol. 537, pp. 2–21. Springer, Berlin, Heidelberg (Aug 1991). https://doi.org/10.1007/3-540-38424-3_1
10. Cannière, C.D.: Analysis and Design of Symmetric Encryption Algorithms. Ph.D. thesis, KU Leuven (May 2007)
11. Canteaut, A., Lallemand, V., Leander, G., Neumann, P., Wiemer, F.: bison instantiating the whitened swap-or-not construction. In: Ishai, Y., Rijmen, V. (eds.) EUROCRYPT 2019, Part III. LNCS, vol. 11478, pp. 585–616. Springer, Cham (May 2019). https://doi.org/10.1007/978-3-030-17659-4_20

12. Canteaut, A., Roué, J.: On the behaviors of affine equivalent sboxes regarding differential and linear attacks. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015, Part I. LNCS, vol. 9056, pp. 45–74. Springer, Berlin, Heidelberg (Apr 2015). https://doi.org/10.1007/978-3-662-46800-5_3
13. Cybersecurity and Infrastructure Security Agency (CISA): Cost of a Cyber Incident: Systematic Review and Cross-Validation. Tech. rep., U.S. Department of Homeland Security (2020), https://www.cisa.gov/sites/default/files/2024-10/CISA-OCECostofCyberIncidentsStudy_508.pdf, accessed: 2025-09-30
14. Daemen, J.: Cipher and hash function design, strategies based on linear and differential cryptanalysis. Ph.D. thesis, KU Leuven (1995)
15. Daemen, J., Rijmen, V.: The wide trail design strategy. In: Honary, B. (ed.) 8th IMA International Conference on Cryptography and Coding. LNCS, vol. 2260, pp. 222–238. Springer, Berlin, Heidelberg (Dec 2001). https://doi.org/10.1007/3-540-45325-3_20
16. Dinur, I.: New Techniques for Analyzing Differentials with Application to AES. Cryptology ePrint Archive, Paper 2025/1326 (2025), https://eprint.iacr.org/2025/1326
17. Hong, S., Lee, S., Lim, J., Sung, J., Cheon, D.H., Cho, I.: Provable security against differential and linear cryptanalysis for the SPN structure. In: Schneier, B. (ed.) FSE 2000. LNCS, vol. 1978, pp. 273–283. Springer, Berlin, Heidelberg (Apr 2001). https://doi.org/10.1007/3-540-44706-7_19
18. Jakobsen, T., Knudsen, L.R.: The interpolation attack on block ciphers. In: Biham, E. (ed.) FSE'97. LNCS, vol. 1267, pp. 28–40. Springer, Berlin, Heidelberg (Jan 1997). https://doi.org/10.1007/BFb0052332
19. Keliher, L., Sui, J.: Exact maximum expected differential and linear probability for two-round Advanced Encryption Standard. Information Security, IET **1**, 53–57 (07 2007). https://doi.org/10.1049/iet-ifs:20060161
20. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Davies, D.W. (ed.) EUROCRYPT'91. LNCS, vol. 547, pp. 17–38. Springer, Berlin, Heidelberg (Apr 1991). https://doi.org/10.1007/3-540-46416-6_2
21. Liu, T., Pelecanos, A., Tessaro, S., Vaikuntanathan, V.: Layout graphs, random walks and the t-wise independence of SPN block ciphers. In: Handschuh, H., Lysyanskaya, A. (eds.) CRYPTO 2023, Part III. LNCS, vol. 14083, pp. 694–726. Springer, Cham (Aug 2023). https://doi.org/10.1007/978-3-031-38548-3_23
22. Liu, T., Pelecanos, A., Tessaro, S., Vaikuntanathan, V.: How fast does the inverse walk approximate a random permutation? Cryptology ePrint Archive, Report 2024/1795 (2024), https://eprint.iacr.org/2024/1795
23. Liu, T., Tessaro, S., Vaikuntanathan, V.: The t-wise independence of substitution-permutation networks. In: Malkin, T., Peikert, C. (eds.) CRYPTO 2021, Part IV. LNCS, vol. 12828, pp. 454–483. Springer, Cham, Virtual Event (Aug 2021). https://doi.org/10.1007/978-3-030-84259-8_16
24. MacWilliams, F.J., Sloane, N.J.A.: The theory of error-correcting codes, vol. 16. Elsevier (1977)
25. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Helleseth, T. (ed.) EUROCRYPT'93. LNCS, vol. 765, pp. 386–397. Springer, Berlin, Heidelberg (May 1994). https://doi.org/10.1007/3-540-48285-7_33
26. Maurer, U.M., Tessaro, S.: Computational indistinguishability amplification: Tight product theorems for system composition. In: Halevi, S. (ed.) CRYPTO 2009. LNCS, vol. 5677, pp. 355–373. Springer, Berlin, Heidelberg (Aug 2009). https://doi.org/10.1007/978-3-642-03356-8_21

27. Nyberg, K., Knudsen, L.R.: Provable security against a differential attack. Journal of Cryptology **8**(1), 27–37 (Dec 1995). https://doi.org/10.1007/BF00204800
28. OpenAI: ChatGPT: GPT-5 Language Model (2025), accessed: 2024-08-28
29. Rijmen, V., Daemen, J., Preneel, B., Bossalaers, A., De Win, E.: The cipher SHARK. In: Gollmann, D. (ed.) FSE'96. LNCS, vol. 1039, pp. 99–111. Springer, Berlin, Heidelberg (Feb 1996). https://doi.org/10.1007/3-540-60865-6_47
30. Rønjom, S.: A short note on a weight probability distribution related to SPNs. Cryptology ePrint Archive, Report 2019/750 (2019), https://eprint.iacr.org/2019/750
31. Schütt, I.: Upper bounding key-averaged differential probabilities. Master's thesis, Ruhr University Bochum (March 2025)
32. Vaudenay, S.: Provable Security for Block Ciphers by Decorrelation. In: Morvan, M., Meinel, C., Krob, D. (eds.) STACS 98, 15th Annual Symposium on Theoretical Aspects of Computer Science, Paris, France, February 25-27, 1998, Proceedings. Lecture Notes in Computer Science, vol. 1373, pp. 249–275. Springer (1998). https://doi.org/10.1007/BFB0028566
33. Vaudenay, S.: Decorrelation: A theory for block cipher security. Journal of Cryptology **16**(4), 249–286 (Sep 2003). https://doi.org/10.1007/s00145-003-0220-6

A Singular values for specific S-boxes

In this appendix, we bound the third singular value of the difference-distribution matrix of the cube and inverse functions.

Lemma 14. *Let $n \geq 3$ be an odd integer. The difference-distribution matrix of the cube function $x \mapsto x^3$ on $\mathbb{F}_{2^n}$ has singular values one with multiplicity two and singular values $2^{-n/2}$ with multiplicity $2^n - 2$.*

Proof. Let D be the difference distribution matrix of $x \mapsto x^3$. By Lemma 7, it suffices to compute the singular values of the matrix

$$X = D - 2^{-n}\, \delta_{\mathbb{F}_{2^n}} \delta_{\mathbb{F}_{2^n}}^{\mathsf{T}},$$

Since $x \mapsto x^3$ is differentially 2-uniform, it holds that $X_{b,a} = \pm 2^{-n}$ if either a or b is nonzero. It will now be shown that the columns of X are almost orthogonal. Consider an input difference a and output difference b. The right pairs satisfy $(x+a)^3 + x^3 = b$, or equivalently

$$a\, x^2 + a^2\, x + a^3 + b = 0.$$

For $a \neq 0$, this equation has either two distinct solutions or no solutions. Specifically, the equation has two solutions iff $\mathrm{Tr}(1 + b/a^3) = 0$ where Tr denotes the absolute trace function on $\mathbb{F}_{2^n}$. Consequently, $X_{b,a} = (-1)^{\mathrm{Tr}(1+b/a^3)}\, 2^{-n}$ for nonzero a and b. Accordingly, by the orthogonality of additive group characters,

$$\sum_{c \in \mathbb{F}_{2^n}} X_{c,b} X_{c,a} = 2^{-2n} \sum_{c \in \mathbb{F}_{2^n}} (-1)^{\mathrm{Tr}(c/b^3 + c/a^3)} = \begin{cases} 2^{-n} & \text{if } a = b\,, \\ 0 & \text{else}\,. \end{cases}$$

For $a = 0$ and $b \neq 0$, one obtains

$$\sum_{c \in \mathbb{F}_{2^n}} X_{c,b} X_{c,0} = -2^{-n} \sum_{c \in \mathbb{F}_{2^n}^{\times}} X_{c,b} - 2^{-n}(1 - 2^{-n}) = -2^{-n},$$

and likewise for $b = 0$ and $a \neq 0$. For $a = b = 0$, one obtains

$$\sum_{c \in \mathbb{F}_{2^n}} X_{c,b} X_{c,0} = 2^{-2n}(2^n - 1) + (1 - 2^{-n})^2 = 1 - 2^{-n}.$$

In matrix form, these equalities state $X^{\mathsf{T}} X = 2^{-n} I + \delta_0 \delta_0^{\mathsf{T}} - 2^{-n}(\delta_{\mathbb{F}_{2^n}} \delta_0^{\mathsf{T}} + \delta_0 \delta_{\mathbb{F}_{2^n}}^{\mathsf{T}})$. Hence, using the fact that $D \delta_{\mathbb{F}_{2^n}} = \delta_{\mathbb{F}_{2^n}}$ and $\delta_{\mathbb{F}_{2^n}}^{\mathsf{T}} D = \delta_{\mathbb{F}_{2^n}}^{\mathsf{T}}$ (see Lemma 7),

$$D^{\mathsf{T}} D = X^{\mathsf{T}} X + 2^{-n} \delta_{\mathbb{F}_{2^n}} \delta_{\mathbb{F}_{2^n}}^{\mathsf{T}} = 2^{-n} I + (1 - 2^{-n}) \delta_0 \delta_0^{\mathsf{T}} + 2^{-n} \delta_{\mathbb{F}_{2^n}^{\times}} \delta_{\mathbb{F}_{2^n}^{\times}}^{\mathsf{T}}\,.$$

Since $U = \delta_0 \delta_0^{\mathsf{T}} + \delta_{\mathbb{F}_{2^n}^{\times}} \delta_{\mathbb{F}_{2^n}^{\times}}^{\mathsf{T}} / (2^n - 1)$, this can be rewritten as follows:

$$D^{\mathsf{T}} D = 2^{-n}(I - U) + \delta_0 \delta_0^{\mathsf{T}} + \left(\frac{1}{2^n (2^n - 1)} + \frac{1}{2^n} \right) \delta_{\mathbb{F}_{2^n}^{\times}} \delta_{\mathbb{F}_{2^n}^{\times}}^{\mathsf{T}} = 2^{-n}(I - U) + U\,.$$

Hence, the eigenvalues of $(D^S)^{\mathsf{T}} D^S$ are one (with multiplicity two) and 2^{-n} (with multiplicity $2^n - 2$). □

The following lemma was generated with the help of generative AI. In particular, ChatGPT 5 [28] was used to generate a first proof sketch, which was verified manually and (heavily) edited.

Lemma 15. *Let $n \geq 2$ be an odd integer and $S\colon \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ the permutation defined by $S(x) = x^{-1}$ for $x \neq 0$ and $S(0) = 0$. The third singular value of the difference-distribution matrix of S is at most $2^{-n/2} + 2^{1-n}$.*

Proof. Let D be the difference-distribution matrix of S. Since S is APN, $D_{b,a}$ is either zero or $2/2^n$ for nonzero a. More specifically, the coordinates are determined by the number of solutions of the equation (for nonzero x and $x + a$)

$$S(x + a) = S(x) + b \iff x^2 + ax + \tfrac{a}{b} = 0\,.$$

This quadratic equation in x has a solution if and only if $\mathrm{Tr}(\frac{1}{ab}) = 0$. The values $x = 0$ and, equivalently, $x = a$ are solutions to the difference equation if and only if $b = 1/a$. Hence, if a and b are nonzero, then

$$D_{b,a} = \begin{cases} 2/2^n & \text{if } \mathrm{Tr}(\frac{1}{ab}) = 0 \text{ or } ab = 1\,, \\ 0 & \text{else}\,. \end{cases}$$

Up to a row or column permutation given by $a \mapsto 1/a$, the matrix D is circulant with respect to the multiplicative group $\mathbb{F}_{2^n}^\times$. The eigenvalues of circulant matrices are character sums. In particular, for every character $\chi\colon \mathbb{F}_{2^n}^\times \to \mathbb{C}$ of the multiplicative group $\mathbb{F}_{2^n}^\times$, there is a corresponding eigenvalue

$$\frac{2}{2^n} + \frac{2}{2^n} \sum_{\substack{x \in \mathbb{F}_{2^n}^\times \\ \mathrm{Tr}(x)=0}} \chi(x)\,.$$

The first term $2/2^n$ is due to the fact that $D_{1,1} = 2/2^n$ despite $\mathrm{Tr}(1) = 1$. The singular values of the $(2^n - 1) \times (2^n - 1)$ submatrix of D obtained by deleting the row and column corresponding to differences $a = 0$ and $b = 0$ are equal to

$$\left| \frac{2}{2^n} + \frac{2}{2^n} \sum_{\substack{x \in \mathbb{F}_{2^n}^\times \\ \mathrm{Tr}(x)=0}} \chi(x) \right| \leq \frac{2}{2^n} + \frac{1}{2^n} \left| \sum_{x \in \mathbb{F}_{2^n}^\times} \chi(x)\,(-1)^{\mathrm{Tr}(x)} \right|,$$

for $\chi \neq 1$ and one for $\chi = 1$. The inequality follows from the triangle inequality and the orthogonality of characters: $\sum_{x \in \mathbb{F}_{2^n}^\times} \chi(x) = 0$ for $\chi \neq 1$. Finally, a standard result for Gauss sums shows that

$$\left| \sum_{x \in \mathbb{F}_{2^n}^\times} \chi(x)\,(-1)^{\mathrm{Tr}(x)} \right| = 2^{n/2}\,.$$

Hence, we obtain the overall bound $2^{1-n} + 2^{-n/2}$. □

B Additional preliminaries

The following sections contain additional details about results from Section 2.

B.1 Superbox representation of the AES

To give a detailed description of the superbox representation, recall that the 16 bytes of the AES state are typically arranged in a 4×4 array. The AES round function consists of an S-box layer SB, a linear layer $L = \mathsf{MC} \circ \mathsf{SR}$, and a key addition layer AK. The linear map SR shifts every row of the state: the first row is not shifted, the second row is shifted by one position, the third by two, and the last by three. The MC operation multiplies each column of the state by an MDS matrix in $\mathbb{F}_{2^8}^{4\times 4}$. For additional details, we refer the interested reader to [1]. Hence, four AES rounds (without the last linear layer and key addition) correspond to the following sequence of operations:

$$\mathsf{SB} \circ (\mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SR} \circ \mathsf{SB}) \circ (\mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SR} \circ \mathsf{SB}) \circ (\mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SR} \circ \mathsf{SB}) \,.$$

Observe that these operations can be rearranged as follows (see Figure 2):

$$(\mathsf{SB} \circ \mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SB}) \circ \mathsf{AK} \circ (\mathsf{SR} \circ \mathsf{MC} \circ \mathsf{SR}) \circ (\mathsf{SB} \circ \mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SB}) \circ \mathsf{SR} \,,$$

where the middle round key must be modified by applying SR^{-1} to it. The function $\mathsf{SB} \circ \mathsf{AK} \circ \mathsf{MC} \circ \mathsf{SB}$ applies four key-dependent 32-bit 'super S-boxes' to the four columns of the state. This is illustrated in Figure 2, with $L_1 = \mathsf{MC}$ and $L_2 = \mathsf{SR} \circ \mathsf{MC} \circ \mathsf{SR}$. Note that L_2 is $\mathbb{F}_{2^8}$-linear and MDS with respect to the four 32-bit words, but it is not $\mathbb{F}_{2^{32}}$-linear.

B.2 Proof of Lemma 2

The squared norm of the block matrix is the maximum of the following expression over all x and y with $\|x\|_2^2 + \|y\|_2^2 = 1$:

$$\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2^2 \leq \|Ax + By\|_2^2 + \|Cx + Dy\|_2^2 \leq \left(\left\| \begin{bmatrix} \|A\|_2 & \|B\|_2 \\ \|C\|_2 & \|D\|_2 \end{bmatrix} \right\| \left\| \begin{bmatrix} \|x\|_2 \\ \|y\|_2 \end{bmatrix} \right\|_2 \right)^2 .$$

The result follows by the definition of the two norm. □

B.3 Proof of Lemma 3

Let P be the matrix representation of π_V, so that the coordinates of $\pi_V(x)$ relative to the basis B are given by Px. Hence, for all x in $\mathbb{R}[X]$, the vector $BPx - x$ is orthogonal to V. Equivalently, every column of B is orthogonal to $BPx - x$. That is, $B^\intercal(BPx - x) = 0$. It follows that $P = (B^\intercal B)^{-1} B^\intercal$. □