Pairwise Independence of Ten-Round AES

Zhongfeng Niu[1] and Kai Hu[2]

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
zhongfeng.niu@ntu.edu.sg

[2]School of Cyber Science and Technology, Shandong University, Qingdao, China
kai.hu@sdu.edu.cn

Abstract

We prove that ten rounds of AES with independent and uniform random round keys are ε_{10}-close to pairwise independent, with

$$\varepsilon_{10} \leq \frac{3017}{2^{141}} < 2^{-129}.$$

For every two distinct plaintexts, the joint ciphertext distribution is therefore within this statistical distance of the uniform distribution on ordered pairs of distinct ciphertexts. The same sampled key tuple is used for both encryptions. The construction retains the AES round transformations, including the omission of MixColumns in the final round, with input and output whitening. Let E_{10} denote this random permutation and Ω the set of nonzero differences. We establish the result through the induced one-norm bound $\left\| D_{\Omega}^{E_{10}} - U \right\|_1 \leq 3017/2^{140}$ on its averaged difference-distribution matrix restricted to Ω, where U projects onto the uniform distribution on Ω. The proof groups the ten rounds into five superbox layers and decomposes the deviation from uniformity into fourteen orthogonal subspaces according to the number of active words and residual factors. Local Gram moments and MDS code constraints control the dependencies between word differences. Bounds for the distributions produced by the first rounds preserve their common source variables, while joint estimates for the last four rounds retain the constraints across two linear layers. The final conversion to statistical distance uses the support size of each component. Exact rational evaluation of the resulting finite recurrence establishes the stated bound.

Keywords: AES, almost pairwise independence, induced one norm, Gram moments, MDS codes, computer-assisted proof.

1 Introduction

Pairwise independence asks whether a block cipher, under a random key, behaves like a uniform random permutation on every pair of distinct plaintexts. More precisely, the corresponding ciphertext pair must be uniform over all ordered pairs of distinct ciphertexts. Almost pairwise independence allows a statistical distance of at most ε from this distribution [4]. This gives a distributional guarantee for two plaintexts, including events that collect many possible output differences.

The wide-trail strategy bounds differential trail probabilities through the number of active S-boxes, as described by Daemen and Rijmen [8, Chapter 9]. A differential probability sums over all trails connecting its endpoints. Pairwise independence requires the deviation accumulated over all output differences, not the maximum over single trails [4]. Keliher and Sui [2] determined the exact maximum expected differential probability for two core AES rounds. Such local bounds enter the proof below; the object bounded here is the distance of the entire output distribution from uniform.

Liu, Tessaro, and Vaikuntanathan [6] proved almost pairwise independence for substitution–permutation networks with concrete S-boxes and suitable linear layers, including AES with independent round keys after sufficiently many rounds. Their analysis relates S-box difference propagation to sampling from subspaces and uses Fourier analysis to control the resulting distributions.

Dinur [7] developed functional-analytic estimates for key-averaged differentials and proved that forty-round AES has statistical distance at most 2^{-135} from pairwise independence.

Beyne, Leander, and Schütt (BLS) [4] expressed almost pairwise independence through norms of difference-distribution matrices. Their analysis separates word activity from variation within activity patterns and combines the MDS property of the linear layer with local singular-value bounds. In the independent-round-key model, their bound reaches statistical distance below 2^{-128} after twenty AES rounds. The remaining challenge is to obtain this guarantee with fewer rounds while retaining the dependencies created by diffusion.

Ren and Yu [5] strengthened the local estimates using second Gram moments on additive MDS codes. They also bounded the distributions arising from fixed input differences more sharply than arbitrary input distributions. Their result gives statistical distance below $2^{-14-61k}$ after $4+4k$ rounds for each integer $k \geq 0$, reaching the 2^{-128} threshold after twelve rounds. We use their local AES estimates and refine the global composition to reach this threshold after ten rounds.

Geometric approach. Beyne [3] developed a geometric approach to linear cryptanalysis using correlation matrices and orthogonal projections. Beyne and Rijmen [10] introduced quasidifferential transition matrices to express fixed-key differential probabilities as sums over quasidifferential trails.

1.1 Our result

Let E_{10} have the ten-round AES structure, with MixColumns omitted in the last round as specified in the standard [1]. The eleven 128-bit round keys $K_0, \ldots, K_{10}$ are independent and uniform, with K_0 and K_{10} providing input and output whitening. The key tuple is sampled once and shared by both encryptions. Let $N = 2^{128}$ be the size of the state space $\mathbb{F}_2^{128}$.

Contributions. Following Beyne, Leander, and Schütt [4, Definition 1], statistical distance is half the sum of absolute probability differences. We prove that ten-round AES in this model is ε_{10}-pairwise independent, where

$$\begin{aligned} \varepsilon_{10} &:= \frac{1}{2} \max_{x \neq x'} \sum_{y \neq y'} \left| \Pr[E_{10}(x) = y, E_{10}(x') = y'] - \frac{1}{N(N-1)} \right| \\ &\leq \frac{3017}{2^{141}} < 2^{-129}. \end{aligned} \tag{1.1}$$

Here x, x', y, y' range over $\mathbb{F}_2^{128}$, and the probability is over the sampled key tuple. The bound holds for every distinct plaintext pair.

The proof bounds the complete output-difference distribution. Let $\Omega = \mathbb{F}_2^{128} \setminus \{0\}$ and $n = |\Omega| = 2^{128} - 1$. For each $a \in \Omega$, let $D_\Omega^{E_{10}} \delta_a$ be the output-difference probability vector for input difference a, where δ_a is the corresponding standard basis vector. Input whitening makes this distribution independent of the first plaintext. Write $\delta_\Omega = \sum_{a \in \Omega} \delta_a$ and $U = n^{-1} \delta_\Omega \delta_\Omega^*$. The subscript Ω restricts the difference-distribution matrix to nonzero differences. We define

$$\Delta_{10} := \left\| D_\Omega^{E_{10}} - U \right\|_1 = \max_{a \in \Omega} \sum_{b \in \Omega} \left| D_\Omega^{E_{10}}(b, a) - \frac{1}{n} \right|. \tag{1.2}$$

Our quantitative bound is

$$\Delta_{10} \leq \frac{3017}{2^{140}} < 2^{-128}. \tag{1.3}$$

With input and output whitening, the norm formulation of Beyne, Leander, and Schütt [4, Lemma 4] gives $\varepsilon_{10} = \Delta_{10}/2$. Thus, (1.3) yields the pairwise-independence guarantee in (1.1).

We use BLS's activity and residual decomposition [4, Sections 5–6] and Ren and Yu's exact-support incidence bounds [5, Eq. (27)]. We retain these blocks throughout the ten-round recurrence and the final one-norm conversion. We sharpen the local Gram-moment and Fourier estimates, retain the actual common-source distributions in the middle segment, and bound the last four rounds jointly.

1.2 Proof overview

The proof bounds the deviation from uniformity after five superbox layers. Daemen and Rijmen [9] used superboxes to study two-round AES differentials. Here each superbox acts on a 32-bit word and retains the full AES S-box in both substitution layers. Write S for the random single-word superbox and D^{S} for its averaged difference-distribution matrix. The ten-round channel in the grouped coordinates is

$$\widetilde{D}_{\Omega}^{E_{10}} = ((D^{\mathsf{S}})^{\otimes 4} D^{L})^{4} (D^{\mathsf{S}})^{\otimes 4},$$

where L is the linear layer between successive superbox layers and D^{L} is its difference-distribution matrix. This repeats BLS's superbox grouping [4, Appendix B.1] and contains ten substitution layers.

For the difference-distribution matrices considered here, we use the activity-space framework of Beyne, Leander, and Schütt [4]. On one word, write $D^{\mathsf{S}} = P + R$. The projection P retains the zero and nonzero uniform directions, and R describes the remaining variation. In the four-word tensor product, the exact-support blocks record both the number i of residual factors and the number a of active words. The constraints $0 \le i \le a \le 4$ and $a \ge 1$ give fourteen orthogonal subspaces. Keeping all fourteen components through the recurrence allows the final conversion from the two norm to the one norm to use their separate support sizes.

The first rounds produce distributions with more structure than an arbitrary vector in one of these subspaces. For example, a single active input word produces one random word difference after the first superbox layer; the next linear layer maps it into four words that all depend on that same source. Bounds that treat these four words separately discard this relation. We retain the source variables and use additive MDS constraints, Fourier representations, and the local Gram estimates of Ren and Yu [5] to bound their joint contribution. Inputs with more active words are handled according to their number of source variables.

The last four rounds require a separate refinement. Terms with three or four residual factors share internal variables across two linear layers. We sum over these variables before taking norms, preserving the MDS constraints that connect the two layers. The other terms are bounded by products of nonnegative comparison matrices. This gives a joint four-round bound for each input and output subspace.

The bounds combine in a finite recurrence on the fourteen components. At the output, the activity component has zero total mass. Residual components on the same exact activity pattern are orthogonal, so their two norms can be combined before conversion to the one norm. We apply the recurrence to each of the four possible initial word weights, covering every nonzero input difference. Byte-table enumeration and rational evaluation, with square roots rounded upward, give the bound in (1.3).

Outline. Section 2 collects the model, the activity-space framework, and the local and MDS estimates from prior work. Section 3 states the main theorem, organizes the $4 + 2 + 4$-round composition, and proves the one-norm conversion by activity size. Section 4 sharpens the local Gram-moment and Fourier estimates. Sections 5–6 derive the refined four-round comparison bounds. Sections 7–9 analyze the actual prefix and its common-source contributions. Section 10 bounds

the last four rounds jointly, and Section 11 combines the bounds. The appendices give the finite algorithm and the rational evaluation method.

2 Preliminaries

We recall the difference-distribution framework, the AES superbox representation, and the local and MDS bounds used in the analysis.

2.1 Notation and norms

Let $V = \mathbb{F}_2^{32}$, $q = |V| = 2^{32}$, $Q = q - 1$, and $\Omega = V^4 \setminus \{0\}$, with $n = |\Omega| = q^4 - 1$. For a finite set X, write $\mathbb{R}[X]$ for the real vector space with standard basis $(\delta_x)_{x \in X}$. The space $\mathbb{C}[X]$ is defined analogously over $\mathbb{C}$. We identify these vectors with functions on X. For $A \subseteq X$, set $\delta_A = \sum_{x \in A} \delta_x$ [4, Sections 2.3 and 3.3]. The vector δ_A is unnormalized; $\delta_A/\sqrt{|A|}$ has Euclidean norm one, whereas $\delta_A/|A|$ is a probability vector. In particular, the global uniform probability vector is $\frac{\delta_\Omega}{n}$ and

$$U = \frac{\delta_\Omega \delta_\Omega^*}{n}.$$

Write $x_+ = \max\{x, 0\}$ for the positive part of a real number. All vector norms use counting measure. For $p, r \geq 1$, the induced norm is

$$\|A\|_{r,p} = \sup_{v \neq 0} \frac{\|Av\|_r}{\|v\|_p}, \qquad \|A\|_p = \|A\|_{p,p}.$$

Thus the first subscript is the output norm and the second is the input norm. We use $\|A\|_2$ for the Euclidean operator norm and $\|A\|_{\mathrm{fr}} = (\sum_{y,x} |A(y,x)|^2)^{1/2}$ for the Frobenius norm. The symbol A^* denotes the adjoint; it equals $A^\top$ for a real matrix.

For a possibly random function $F : X \to X$ on a finite binary vector space, its difference-distribution matrix is

$$D^F(b, a) = \Pr_{F,x}[F(x + a) + F(x) = b],$$

where x is uniform and independent of F. Rows index output differences and columns index input differences, including zero. For an invertible additive map A, this gives $D^A \delta_x = \delta_{Ax}$ [4, Section 2.2]. A subscript Ω denotes restriction to $\Omega \times \Omega$. Global operators are restricted to $\mathbb{R}[\Omega]$ whenever they are combined with U or the projections below; single-word operators act on $\mathbb{R}[V]$.

Fix binary coordinates on each byte and word. Write $\langle \xi, x \rangle_2$ for their binary dot product and $\chi_\xi(x) = (-1)^{\langle \xi, x \rangle_2}$. The normalized word Fourier transform is

$$\mathcal{F}(\xi, x) = q^{-1/2} \chi_\xi(x), \qquad \xi, x \in V.$$

This pairing also determines the transpose of additive word maps. The four-word transform is $\mathcal{F}^{\otimes 4}$.

We use the change of basis

$$\alpha = \sqrt{q/Q}, \qquad \beta = Q^{-1/2}, \qquad \frac{\delta_{V \setminus \{0\}}}{\sqrt{Q}} = \alpha \frac{\delta_V}{\sqrt{q}} - \beta \delta_0.$$

2.2 The independent-round-key experiment

Write $S = \mathrm{SB}$, $T = \mathrm{SR}$, and $M = \mathrm{MC}$ for the full-state maps. Let $\mathsf{s} : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ be the AES byte S-box. Sample eleven independent uniform 128-bit keys

$$K_0, K_1, \ldots, K_{10},$$

and retain this key tuple for all evaluations in the experiment. Addition is XOR. Define

$$\begin{aligned} z_0 &= x + K_0, \\ z_r &= (M \circ T \circ S)(z_{r-1}) + K_r \qquad (1 \le r < 10), \\ E_{10}(x) &= (T \circ S)(z_9) + K_{10}. \end{aligned}$$

There are ten SubBytes layers, and the last round omits MixColumns. The keys K_0 and K_{10} provide input and output whitening.

The experiment uses the AES S-box, ShiftRows, and MixColumns. It replaces the standard key schedule, which derives correlated round keys from a 128-, 192-, or 256-bit master key.

Input whitening makes the first member of a fixed plaintext pair $(x, x + a)$ uniform while preserving the difference a. At each subsequent nonlinear layer, the fresh uniform round key again makes the first member uniform, conditioned on the internal history of this fixed pair before the key addition and on the current difference. The key-averaged difference-distribution matrices therefore compose by multiplication [4, Section 2.2, Eq. (2)].

Statistical distance. For probability vectors μ, μ' on a finite set, their statistical distance is $\frac{1}{2}\sum_x |\mu(x) - \mu'(x)|$. With independent input and output whitening, the distance from pairwise independence equals half the induced one norm of the centered difference-distribution matrix [4, Definition 1 and Lemma 4]. Indeed, output whitening gives

$$\Pr[E_{10}(x) = y, E_{10}(x + a) = y + b] = N^{-1} D_{\Omega}^{E_{10}}(b, a), \qquad N = q^4.$$

Summing the absolute differences over y removes the factor N^{-1}. Maximizing over $a \neq 0$ gives the pair-channel identity used below.

2.3 Exact grouping into five superbox layers

SubBytes and ShiftRows commute. Set

$$L = T \circ M \circ T, \qquad B_k(y) = S\big((M \circ S)(y) + K_{2k-1}\big).$$

Treat the four byte columns as four words, each with alphabet size $q = 2^{32}$. The keyed recurrence has the exact form

$$\begin{aligned} Y_0 &= T(x + K_0), \\ U_k &= B_k(Y_{k-1}) \qquad (1 \le k \le 5), \\ Y_k &= LU_k + TK_{2k} \qquad (1 \le k < 5), \\ E_{10}(x) &= TU_5 + K_{10}. \end{aligned}$$

For example, the complete state after round two is $z_2 = (M \circ T)(U_1) + K_2$, so the next grouped input is $Tz_2 = LU_1 + TK_2$.

Let $S_c : V \to V$ apply s to each of the four bytes, and let $M_c : V \to V$ be one column of MixColumns. Define the random single-word superbox

$$\mathsf{S}_\kappa(y) = S_c((M_c \circ S_c)(y) + \kappa), \qquad \kappa \text{ uniform in } V.$$

We write S for this random function. Each B_k consists of four independent copies of S. Odd-indexed keys lie inside these superboxes; even-indexed keys lie between successive superbox layers. Since T preserves uniformity and independence of the keys, every word has difference-distribution matrix

$$D^{\mathsf{S}} = (D^{\mathsf{s}})^{\otimes 4}\, D^{M_c}\, (D^{\mathsf{s}})^{\otimes 4}.$$

The four-word parallel channel is $(D^{\mathsf{S}})^{\otimes 4}$. The outer linear layer has difference-distribution matrix D^L. Repeating the grouping of BLS [4, Appendix B.1] gives

$$\widetilde{D}_{\Omega}^{E_{10}} = ((D^{\mathsf{S}})^{\otimes 4} D^L)^4 (D^{\mathsf{S}})^{\otimes 4}. \tag{2.1}$$

The channel in the original coordinates satisfies

$$D_{\Omega}^{E_{10}} = D^T \widetilde{D}_{\Omega}^{E_{10}} D^T.$$

The fixed permutations on both sides preserve all norms used here, and $D^T U = U D^T = U$. Two superbox layers contain four substitutions, and five contain exactly ten.

For a finite additive alphabet V, an additive $[\ell, d]$ MDS code is an additive subgroup of V^ℓ whose restriction to any d coordinates is a bijection onto V^d. It contains q^d points. An affine MDS code is a translate of such a subgroup. These definitions require additivity, without scalar $\mathbb{F}_q$-linearity.

2.4 Word-MDS geometry of the outer layer

The outer AES layer $L = \mathrm{SR} \circ \mathrm{MC} \circ \mathrm{SR}$ satisfies

$$(Lx)_{u,c} = \sum_v (M_{\mathrm{c}})_{u,v} x_{v,c+u+v},$$

with byte indices modulo 4. Its graph $\{(x, Lx) : x \in V^4\}$ is an additive $[8, 4]$ MDS code [4, Appendix B.1]. The byte equations separate into four systems [11, Appendix C.2]. For each input word d and output word c, set

$$z_{r,d} = x_{d-r,d}, \qquad w_{r,c} = (Lx)_{r-c,c}.$$

Then

$$w_{r,c} = \sum_d (M_{\mathrm{c}})_{r-c,d-r} z_{r,d}. \tag{2.2}$$

For fixed r, this is a scalar 4×4 MDS system obtained by permuting rows and columns of M_{c}. Each word block of L is a byte permutation followed by nonzero byte scalings, which we call a byte-monomial map. Zero restrictions, information-coordinate choices, inversion, transposition, and duality preserve the separation by r. On a shortened code line, every coordinate in each system is a nonzero scalar multiple of the free byte. Thus every word is a byte permutation and nonzero byte scaling of the common word parameter. These maps and their transposes preserve byte weight.

2.5 Activity and residual subspaces

The activity and residual decomposition follows Beyne, Leander, and Schütt [4, Lemma 7 and Section 6.1]. On one word, define the orthogonal projections

$$P = \delta_0 \delta_0^* + \frac{\delta_{V\setminus\{0\}} \delta_{V\setminus\{0\}}^*}{Q}, \qquad Z = I - P.$$

The channel D^{S} isolates the zero difference and is doubly stochastic on nonzero differences, so

$$D^{\mathsf{S}} P = P D^{\mathsf{S}} = P, \qquad R := D^{\mathsf{S}} - P = Z D^{\mathsf{S}} Z.$$

For $A \subseteq \{1, 2, 3, 4\}$, let $\Omega_A = \{x \in V^4 : \mathrm{supp}(x) = A\}$. It has size $Q^{|A|}$, and its unit indicator is $\delta_{\Omega_A}/Q^{|A|/2}$. Here an active coordinate must be nonzero. We resolve this decomposition by exact support, as in [5, Section 4.2]. For residual positions $I \subseteq A$, define

$$E_{I,A} = \bigotimes_{r=1}^{4} \begin{cases} Z, & r \in I, \\ Q^{-1} \delta_{V\setminus\{0\}} \delta_{V\setminus\{0\}}^*, & r \in A \setminus I, \\ \delta_0 \delta_0^*, & r \notin A. \end{cases}$$

Distinct pairs (I, A) give orthogonal subspaces. Let $E_{i,a}$ be the sum of the projections with $|I| = i, |A| = a$. The index set is $\mathcal{S} = \{(i, a) : 0 \le i \le a \le 4,\ a \ge 1\}$. Write $\deg(i, a) = i$ and $\mathcal{A} = \{(0, a) : 1 \le a \le 4\}$. After excluding the all-zero difference, its fourteen elements are

$$(0,1),(0,2),(0,3),(0,4),\quad (1,1),\dots,(1,4),\quad (2,2),(2,3),(2,4),\quad (3,3),(3,4),\quad (4,4).$$

Write W_j for the orthogonal sum of the ranges of $E_{j,a}$ over all allowed a [4, Section 6.1]. The subspace selected by $E_{i,a}$ contains

$$n_i(a) = \binom{4}{i}\binom{4-i}{a-i}$$

fixed-position blocks. Its norm $\|E_{i,a}f\|_2$ already combines these orthogonal blocks by a sum of squares. The channel $(D^{\mathsf{S}})^{\otimes 4}$ commutes with every projection. If $\|R\|_2 \le \nu$, then

$$\|(D^{\mathsf{S}})^{\otimes 4}E_{i,a}\|_2 \le \nu^i, \qquad (D^{\mathsf{S}})^{\otimes 4}E_{0,a} = E_{0,a}.$$

If every block with fixed input and output positions has norm at most g, the aggregated block has norm at most $g\sqrt{n_i(a)n_j(b)}$. Apply the triangle inequality for each output position, Cauchy–Schwarz over input positions, and then sum the squared output norms.

2.6 Comparison bounds for the linear layer

We use BLS's MDS graph counts and shortened-code argument [4, Lemmas 9–10 and 12–13], retaining exact support sizes. Let $n = q^4 - 1$, $U = n^{-1}\delta_\Omega\delta_\Omega^*$, and $\Omega = V^4 \setminus \{0\}$. The matrix C bounds the block norms of $D^L - U$. If $i + j > 0$, the contribution of U is zero. Set

$$k = i + j, \qquad u = a + b - i - j.$$

For fixed input and output positions, there are k residual coordinates and u coordinates carrying $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$. Expand each remaining nonzero indicator using $\delta_{V\setminus\{0\}} = \delta_V - \delta_0$. If t_{free} unrestricted coordinates remain, the shortened graph code has dimension $d = \max\{k + t_{\text{free}} - 4, 0\}$. Terms with $d \le 0$ vanish by $Z\delta_0 = 0$. Terms with $d \ge k$ have a constant marginal on the residual coordinates and vanish by $Z\delta_V = 0$. Only $1 \le d \le k - 1$ remain.

For these terms, restriction to the k residual coordinates is injective and gives an $[k, d]$ MDS code. Define r_{in} and r_{out} as the base-q logarithms of the cardinalities of its input and output projection images. The MDS property gives $r_{\text{in}} = \min(j, d)$ and $r_{\text{out}} = \min(i, d)$. The nonzero row and column degrees of the incidence matrix are $q^{d-r_{\text{out}}}$ and $q^{d-r_{\text{in}}}$, respectively. The Schur test gives $q^{d-(r_{\text{in}}+r_{\text{out}})/2}$. Projection onto the residual space cannot increase the norm. Including the normalization $Q^{-u/2}$ and the inclusion–exclusion counts as in [5, Eq. (27)], then aggregating over positions, gives

$$C_{(i,a),(j,b)} := \min\left\{1, \sqrt{\frac{n_i(a)n_j(b)}{Q^u}}\sum_{d=1}^{k-1}\binom{u}{u+k-4-d}q^{d-(\min(i,d)+\min(j,d))/2}\right\}.$$

This defines the comparison coefficients involving residual factors. Out-of-range binomial coefficients are zero.

The activity-to-activity block is centered directly. Set

$$N_r^{\text{graph}} = \sum_{\ell=0}^{r}(-1)^{r-\ell}\binom{r}{\ell}q^{\max(\ell-4,0)}.$$

This counts graph-code points nonzero at exactly the specified r positions, by inclusion–exclusion and BLS's MDS intersection count [4, Lemma 10]. For $|A| = a, |B| = b$,

$$\langle\frac{\delta_{\Omega_A}}{Q^{|A|/2}}, (D^L - U)\frac{\delta_{\Omega_B}}{Q^{|B|/2}}\rangle = \frac{N_{a+b}^{\text{graph}}}{Q^{(a+b)/2}} - \frac{Q^{(a+b)/2}}{n}.$$

The count depends only on $a+b$, so the exact aggregated block norm is

$$C_{(0,a),(0,b)} = \sqrt{\binom{4}{a}\binom{4}{b}Q^{a+b}\left|\frac{N_{a+b}^{\mathrm{graph}}}{Q^{a+b}} - \frac{1}{n}\right|}.$$

Centering before taking absolute values retains the cancellation between the activity matrix and the global uniform term.

2.7 Centered operators and block comparison

If $AU = UA = U$ and $BU = UB = U$, then

$$(A-U)(B-U) = AB - U.$$

This centering identity gives the composition rule for difference-distribution matrices [4, Lemma 5 and Eq. (3)].

For a complete orthogonal decomposition $(E_s)_s$, suppose $\|E_rAE_s\|_2 \le a_{rs}$ with $a_{rs} \ge 0$. If $z_s = \|E_sx\|_2$, the triangle inequality gives

$$\|E_rAx\|_2 \le \sum_s a_{rs}z_s.$$

Summing squared lengths over the output subspaces yields $\|A\|_2 \le \|(a_{rs})\|_2$, the block-norm comparison used by Beyne, Leander, and Schütt [4, Lemma 2]. Its composition form is as follows.

Lemma 2.1 (Multiplication of block-norm bounds)**.** *Suppose two operators have nonnegative comparison matrices (a_{rt}) and (b_{ts}) for the same complete orthogonal decomposition. Their product is bounded by ordinary matrix multiplication:*

$$E_rABE_s = \sum_t (E_rAE_t)(E_tBE_s), \qquad \|E_rABE_s\|_2 \le \sum_t a_{rt}b_{ts}.$$

Proof. Insert the sum of the orthogonal projections between the operators and apply the triangle inequality and submultiplicativity to each term. □

2.8 Known local AES bounds

Let $m = 256$. The byte Fourier transform uses the same binary coordinates as the word transform. It is given by

$$(\mathcal{F}_b)_{\xi,x} = m^{-1/2}(-1)^{\langle \xi, x\rangle_2} \qquad (\xi, x \in \mathbb{F}_2^8), \qquad \mathcal{F} = \mathcal{F}_b^{\otimes 4}.$$

Thus the word transform has normalization $q^{-1/2}$. The transpose $M_{\mathrm{c}}^{\top}$ is taken with respect to the same binary pairing. Squared norms use the Euclidean inner product with counting measure.

Lemma 2.2 (Local channel bounds)**.** *For $T \in \{R, R^*, \mathcal{F}R\mathcal{F}, \mathcal{F}R^*\mathcal{F}\}$, the following bounds hold [5, Lemmas 3 and 8]:*

$$\|T\|_2 \le \nu = 11/2^{16}, \qquad \max_x \|T\delta_x\|_2^2 \le \delta = \frac{1}{5q}, \tag{2.3}$$

$$\left\|D^{\mathsf{S}}\delta_x\right\|_2^2 \le \rho = \frac{6}{5q} \quad (x \ne 0), \qquad \left\|RD^{T_0}D^{\mathsf{S}}\delta_x\right\|_2^2 \le 13/q^2. \tag{2.4}$$

In the last bound, T_0 is any invertible byte-monomial map. Keliher and Sui [2] give the two-round superbox point-probability bound

$$0 \le D^{\mathsf{S}}(y,x) \le p := \frac{53}{4q} \quad (x, y \ne 0). \tag{2.5}$$

Centering gives

$$\max_{y,x} |R(y,x)| \le r_* := \frac{49}{4q}.$$

Indeed, $p - 1/Q < 49/(4q)$ *and* $1/Q < 49/(4q)$*, and residual entries involving zero vanish.*

Byte blocks. We combine BLS's byte spectral bound [4, Section 6.4] with the row, column, and point bounds recorded by Ren and Yu [5, Section 4.1]:

$$\sigma_b = 9/128, \qquad r_b = 65/8192, \qquad p_b = 1/64, \qquad \Delta_b = r_b - 1/255.$$

For each of $D^{\mathsf{s}}, (D^{\mathsf{s}})^*, \mathcal{F}_b D^{\mathsf{s}} \mathcal{F}_b, \mathcal{F}_b (D^{\mathsf{s}})^* \mathcal{F}_b$, σ_b bounds the residual norm, r_b bounds the squared norm of every nonzero-indexed row or column, and p_b bounds entries with both indices nonzero. The entries of $\mathcal{F}_b D^{\mathsf{s}} \mathcal{F}_b$ are squares of normalized Walsh coefficients, by the differential–linear relation of Chabaud and Vaudenay [12, Lemmas 1–2]. Split each byte into its zero direction, uniform nonzero direction, and their orthogonal complement. Group the four-byte tensor factors by their number of residual factors, as in Section 2.5. For MixColumns between these five subspaces, Ren and Yu [5, Eq. (20)] give the comparison matrix

$$(c_{ij}^{\mathrm{byte}})_{i,j=0}^{4} = \begin{pmatrix} 1/10400 & 0 & 7/500 & 9/50 & 1 \\ 0 & 27/8000 & 43/500 & 1 & 1 \\ 7/500 & 43/500 & 9/10 & 1 & 1 \\ 9/50 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \qquad b_{ij}^{\mathrm{byte}} = c_{ij}^{\mathrm{byte}} \sigma_b^{i+j}. \tag{2.6}$$

Thus b_{ij}^{byte} bounds the degree-j to degree-i block of R. The degree-zero block is the centered activity compression of BLS [4, Sections 5.1–5.2], restricted to the word residual space. Its squared Frobenius norm is bounded by [5, Eq. (22) and Lemma 5]

$$h_{00} = \frac{68m^2 + 264m + 68}{(m-1)^6}, \qquad m = 256. \tag{2.7}$$

2.9 MDS code bounds

Here t is any common upper bound on the second Gram moments; its numerical value is not fixed in these lemmas.

Lemma 2.3 (MDS Hölder inequality)**.** *Let* $\mathcal{C} \subset V^\ell$ *be an additive* $[\ell, d]$ *MDS code with* $1 \le d \le \ell$*, so every projection onto* d *coordinates is bijective. For nonnegative functions* f_r*,*

$$\sum_{x \in \mathcal{C}} \prod_{r=1}^{\ell} f_r(x_r) \le \prod_{r=1}^{\ell} \left(\sum_{z \in V} f_r(z)^{\ell/d} \right)^{d/\ell}. \tag{2.8}$$

Proof. Apply Hölder over all d-element coordinate subsets as in [5, Lemma 1]. Bijectivity factors the sum for each subset, and every coordinate occurs in the same number of subsets. □

Lemma 2.4 (Gram bound for a code indicator)**.** *Let* $\mathcal{C} \subset V^\ell$ *be an additive* $[\ell, d]$ *MDS code. Let* $T_1, \dots, T_\ell$ *be matrices on* $\mathbb{C}[V]$ *satisfying* $\max_x \|T_r \delta_x\|_2^2 \le \delta$ *and* $\mathrm{Tr}((T_r^* T_r)^2) \le t$ *for every* r*. Suppose their two-sided Fourier conjugates have the same column bound. The unit code indicator* $q^{-d/2} \delta_{\mathcal{C}}$ *satisfies*

$$\left\| (T_1 \otimes \cdots \otimes T_\ell) \frac{\delta_{\mathcal{C}}}{q^{d/2}} \right\|_2 \le \Phi(\ell, d) := \sqrt{(t/q)^{\min(d, \ell-d)} \delta^{\ell - 2\min(d, \ell-d)}} \quad (0 < d < \ell). \tag{2.9}$$

Proof. Ren and Yu [5, Proposition 1] state the bound for identical factors. Their proof also gives this form with separate factors, and the affine extension below. If $d \le \ell/2$, expanding the square gives ℓ Gram entries. Apply (2.8) to $\mathcal{C}\times\mathcal{C}$ over the alphabet V^2. Each local Gram matrix satisfies $|G_{xy}| \le \delta$ and $\sum |G_{xy}|^2 \le t$. Its ℓ/d-th moment is therefore at most $\delta^{\ell/d-2}t$. Including the normalization q^{-d} gives the square of (2.9).

If $d > \ell/2$, the normalized Fourier transform maps the unit code indicator to the unit dual-code indicator, reducing to the preceding case. For an affine code, the transform introduces coordinate-wise factors of modulus one. Diagonal unitary conjugation preserves the required column norms and Gram moments, so the same bound holds.

□

Lemma 2.5 (Schur–Gram bound for an MDS coordinate map)**.** *The weighted Schur argument of Ren and Yu [5, Lemma 2] gives the following coordinate-map bound. Parametrize an additive MDS code $\mathcal{C}$ by d information coordinates $z \in V^d$, and let $c(z) \in V^\ell$ retain ℓ code coordinates. Define*

$$J_{\mathcal{C}} : \mathbb{C}^{V^d} \longrightarrow \mathbb{C}^{V^\ell}, \qquad J_{\mathcal{C}}\delta_z = \delta_{c(z)}.$$

Thus $(J_{\mathcal{C}}f)(y) = \sum_{z:c(z)=y} f(z)$, with no normalization factor. For $d \le \ell$, the map c is injective and $J_{\mathcal{C}}$ is an isometric inclusion. For $d > \ell$, every $y \in V^\ell$ has $q^{d-\ell}$ preimages. In both cases,

$$\left\| R^{\otimes\ell} J_{\mathcal{C}} \right\|_2 \le \nu^{\min(d,\ell)} \delta^{(\ell-d)_+/2} q^{(d-\ell)_+/2}. \tag{2.10}$$

The same bound holds with R, ν, δ replaced by any matrix on $\mathbb{C}[V]$ and upper bounds on its operator norm and squared column norms.

Proof. For $d \le \ell$, choose d information coordinates. Their Gram operator has norm at most ν^{2d}. Each remaining coordinate contributes a positive semidefinite Schur multiplier G' whose diagonal is at most δ. Here $\circ$ denotes entrywise multiplication, and $A \preceq B$ means that $B - A$ is positive semidefinite. Apply repeatedly the implication from $G \preceq cI$ to $G \circ G' \preceq c(I \circ G')$ [5, Lemma 2].

For $d > \ell$, the map projects onto ℓ coordinates. Every image has $q^{d-\ell}$ preimages, so summation over preimages has norm equal to the square root of this number. The residual channels contribute ν^ℓ.

□

2.10 Source distributions after four rounds

For an input difference a_0 with w active words, the first parallel superbox layer produces w independent sources. The linear layer transports their product density onto an additive $[4, w]$ MDS code. The second parallel layer acts separately on its four coordinates. In particular, for $w = 1$, let $\mu = D^{\mathsf{S}}\delta_a$, where a is the active input word, and let $T_1, \dots, T_4$ be the nonzero word blocks of L from that input position. The four-round distribution is

$$p_4 = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0} = \sum_x \mu(x) \bigotimes_{r=1}^{4} D^{\mathsf{S}} \delta_{T_r x}.$$

For residual positions $I \subseteq \{1, 2, 3, 4\}$, its corresponding component is

$$f_I = \sum_x \mu(x) \bigotimes_{r\in I} R\delta_{T_r x} \bigotimes_{r\notin I} \frac{\delta_{V\setminus\{0\}}}{Q}. \tag{2.11}$$

The same variable x occurs in every factor. Since $\mu(0) = 0$ and the T_r are invertible, the nonresidual factors are all $\frac{\delta_{V\setminus\{0\}}}{Q}$.

2.11 Known bounds for restricted sources

The following form of Ren and Yu's restriction bound keeps the source density unnormalized [5, Lemma 6]. Let X be an elementary abelian 2-group of size m_0. Feed $w \in \{1,2,3,4\}$ independent laws supported on $X \setminus \{0\}$ into an additive four-coordinate MDS map, with the other input coordinates zero. Suppose each law has squared norm at most ρ_0 and maximum point probability at most p_0, and write Y for the output. For a specified set of z zero coordinates, put $U_{w,0} = 1$ and $U_{w,z} = 0$ for $z \geq w$. The remaining event bounds are

w	$U_{w,1}$	$U_{w,2}$	$U_{w,3}$
1	0	0	0
2	ρ_0	0	0
3	ρ_0	$p_0\rho_0$	0
4	ρ_0	ρ_0^2	$p_0^2\rho_0$

. (2.12)

Let D, S be disjoint output-position sets with $|D| = z < w$ and $|S| = j \geq 1$. For an additional indicator event E on the source tuple, define $h_D(y) = \Pr[Y_S = y, Y_D = 0, E]$. If an output residual channel T has norm at most s_0 and squared column norms at most δ_0^{col}, then the proof of [5, Lemma 6(b)] gives, with $a = \min(w - z, j)$,

$$\|T^{\otimes j} h_D\|_2^2 \leq p_0^{2z} (\rho_0 s_0^2)^a (\delta_0^{\mathrm{col}})^{j-a}. \tag{2.13}$$

Indeed, any $w-z$ source coordinates parameterize the shortened code. The other probability factors cost at most p_0^z. If $j \geq w-z$, use the operator norm on information coordinates and Gram diagonal bounds on the others. If $j < w - z$, the marked marginal is bounded by a convolution of j source laws and their translates, whose squared norm is at most ρ_0^j. An indicator restriction decreases this nonnegative input density; no conditional normalization is taken. For the residual channel in Ren and Yu's statement, $\delta_0^{\mathrm{col}} = \rho_0 - 1/(m_0 - 1)$.

Common sources and convolution. For matrices $A_1, \ldots, A_j$ with $j \geq 2$, suppose $\|A_r\|_2 \leq \nu_0$, $\|A_r\|_{\mathrm{fr}}^2 \leq h_0$, and $\|A_r \delta_x\|_2^2 \leq \delta_0^{\mathrm{col}}$. Ren and Yu [5, Corollary 1] show that weights with $\|\mu\|_\infty \leq p_0$ satisfy

$$\left\| \sum_x \mu(x) \bigotimes_{r=1}^{j} A_r \delta_x \right\|_2^2 \leq \nu_0^2 p_0^2 h_0 (\delta_0^{\mathrm{col}})^{j-2}. \tag{2.14}$$

This follows by expanding the squared norm as a Schur product of Gram matrices and applying their weighted Schur bound [5, Lemma 2]. For a sum of $w \geq 2$ independently relabeled source laws on X, each of squared norm at most ρ_0, its law g satisfies

$$\|g - m_0^{-1} \delta_X\|_2^2 \leq \frac{(m_0\rho_0 - 1)^w}{m_0}. \tag{2.15}$$

This is the one-coordinate case of Dinur's Fourier marginal bound [7]. Fourier transformation makes the convolution a product; Parseval bounds the sum of squares of each source's nonconstant character expectations by $m_0\rho_0 - 1$. For $w \geq 2$, Hölder also gives

$$|g(0) - 1/m_0| \leq (m_0\rho_0 - 1)^{w/2}/m_0. \tag{2.16}$$

Byte-weight energies. Specializing (2.12) to $\rho_0 = r_b$ and $p_0 = p_b$ defines $U_{w,z}^{\mathrm{byte}}$. Let $E_{w_0,w}^{\mathrm{byte}}$ bound the squared norm of the superbox output restricted to byte weight w, for initial byte weight w_0. Keeping each output-weight term in Ren and Yu's degree bound [5, Eq. (33)] gives $E_{w_0,w}^{\mathrm{byte}} = 0$ when $z = 4 - w \geq w_0$, and otherwise

$$E_{w_0,w}^{\mathrm{byte}} = \frac{\binom{4}{w}}{255^w} (U_{w_0,z}^{\mathrm{byte}})^2 + \sum_{j=1}^{w} \frac{\binom{4}{w}\binom{w}{j}}{255^{w-j}} p_b^{2z} (r_b\sigma_b^2)^{\min(w_0-z,j)} \Delta_b^{j-\min(w_0-z,j)}. \tag{2.17}$$

The terms bound the squared norms of orthogonal exact-support components. The same formula holds in all four word-channel orientations because the byte parameters and MDS property are shared.

3 The Ten-Round Bound and Its Composition

We state the ten-round result, organize the five superbox layers, and bound the final one norm using exact activity sizes.

3.1 The one-norm bound

Theorem 3.1 (One-norm mixing of ten-round AES)**.** *Let E_{10} be the ten-round permutation with independent uniform round keys and input/output whitening defined in Section 2. Use the same sampled key tuple for every input. Its difference-distribution matrix on $\Omega = \mathbb{F}_2^{128} \setminus \{0\}$ satisfies*

$$\|D_\Omega^{E_{10}} - U\|_1 < 2^{-128}. \tag{3.1}$$

More precisely, the exact-arithmetic calculation gives $\|D_\Omega^{E_{10}} - U\|_1 \le 3017/2^{140} < 2^{-128}$.

By the whitening identity recalled in Section 2, the same bound controls output pairs. Let $\mathcal{P}_{10}$ be the averaged transition matrix of E_{10} on ordered distinct input pairs $\mathcal{X}_{\neq}^2 = \{(x, x') \in (\mathbb{F}_2^{128})^2 : x \neq x'\}$, and let every entry of $\mathcal{U}$ equal $1/[N(N-1)]$. Then

$$\|\mathcal{P}_{10} - \mathcal{U}\|_1 = \|D_\Omega^{E_{10}} - U\|_1.$$

In particular, E_{10} is ε_{10}-pairwise independent with

$$\varepsilon_{10} = \frac{1}{2}\|\mathcal{P}_{10} - \mathcal{U}\|_1 = \frac{1}{2}\|D_\Omega^{E_{10}} - U\|_1 \le \frac{3017}{2^{141}} < 2^{-129}.$$

The ten-round bound is proved in Section 11.

3.2 The ten-round composition

Let B be a nonnegative comparison matrix for $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U$ on the fourteen subspaces, and let D be diagonal with $\mathsf{D}_{(i,a),(i,a)} = \nu^i$. Then $\mathsf{D}C$ bounds $(D^{\mathsf{S}})^{\otimes 4} D^L - U$.

Since $(D^{\mathsf{S}})^{\otimes 4} U = U (D^{\mathsf{S}})^{\otimes 4} = U$ and $D^L U = U D^L = U$,

$$((D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U)(D^L - U) = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} D^L - U,$$
$$((D^{\mathsf{S}})^{\otimes 4} D^L - U)((D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U) = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U.$$

The product BC therefore bounds the four-round operator $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} D^L - U$. For an activity input $E_{0,b}$, an extra $(D^{\mathsf{S}})^{\otimes 4}$ can be removed using $(D^{\mathsf{S}})^{\otimes 4} E_{0,b} = E_{0,b}$, so $\mathsf{D}CB$ also bounds this operator. The candidate $\mathsf{D}CB$ is used only in the four activity-input columns. The other columns use BC and the joint estimates established below.

For a fixed input difference $a_0 \in \Omega$ in the grouped coordinates, the first four rounds give $p_4 = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0}$. Applying $(D^{\mathsf{S}})^{\otimes 4} D^L$ gives p_6, and applying $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} D^L$ then gives p_{10}. The recurrence thus consists of four rounds, two rounds, and four rounds. The common-source bounds in Section 8 control contributions to the same intermediate vector and may be combined with other valid bounds by taking their minimum.

The joint estimates replace specified intermediate contributions. The combined contribution from $(3,3)$ and $(3,4)$ retains three R factors and one complete P factor. The contribution from $(4,4)$ contains four R factors. Each is compared with its corresponding joint bound. The ordinary block products for states with at most two residual factors are retained. Each intermediate term containing a residual factor annihilates U. Only the activity term requires global centering.

3.3 Conversion to the one norm

Lemma 3.2 (One-norm bound from the orthogonal components)**.** *Let $q = 2^{32}$, $Q = q - 1$, and $\Omega = (\mathbb{F}_2^{32})^4 \setminus \{0\}$. Let $E_{i,a}$ be the projections with i residual coordinates and a active coordinates. If $f \in \mathbb{R}[\Omega]$ satisfies $\sum_x f(x) = 0$ and $\|E_{i,a} f\|_2 \le v_{i,a}$, then*

$$\|f\|_1 \le 2 \sum_{a=1}^{3} \sqrt{\binom{4}{a} Q^a}\, v_{0,a} + \sum_{a=1}^{4} \sqrt{\binom{4}{a} Q^a} \sqrt{\sum_{i=1}^{a} v_{i,a}^2}. \tag{3.2}$$

The activity component with four active words also satisfies

$$\|E_{0,4} f\|_2 \le \sum_{a=1}^{3} \sqrt{\binom{4}{a}}\, Q^{(a-4)/2} v_{0,a}.$$

Proof. Write $f = f_{\mathrm{act}} + f_{\mathrm{res}}$, where $f_{\mathrm{act}} = \sum_a E_{0,a} f$. Every residual factor vanishes at zero and has zero sum. Hence $\sum_x f_{\mathrm{res}}(x) = 0$, and f_{act} also has zero total mass. On an exact activity pattern A, write the activity component as c_A times the unit indicator. Its mass is $Q^{|A|/2} c_A$. There is only one pattern with four active words, so its absolute mass is at most the sum of the absolute masses of the other patterns. Therefore

$$\|f_{\mathrm{act}}\|_1 \le 2 \sum_{a=1}^{3} Q^{a/2} \sum_{|A|=a} |c_A| \le 2 \sum_{a=1}^{3} \sqrt{\binom{4}{a} Q^a}\, v_{0,a}.$$

Dividing the same mass identity by Q^2 gives the bound for the component with four active words. For each a, the vector $\sum_{i=1}^{a} E_{i,a} f$ is supported on a set of size $\binom{4}{a} Q^a$. Orthogonality bounds its squared norm by $\sum_{i=1}^{a} v_{i,a}^2$. Apply Cauchy–Schwarz on this support, sum over a, and add the activity bound to obtain (3.2). The factor $\binom{a}{i}$ is already included in the block norms and is not counted again. □

4 Refined Local AES Bounds

We refine the Gram moments and derive Fourier entry bounds and response bounds for products of residual entries.

4.1 Gram bounds from the byte decomposition

Proposition 4.1 (Gram moment bounds)**.** *The superbox residual satisfies*

$$\mathrm{Tr}(R^* R) < h = 21813/10^{10}, \qquad \mathrm{Tr}((R^* R)^2) < t = 91/10^{16}.$$

Proof. Retaining the individual byte blocks gives tighter trace bounds. Use b_{ij}^{byte} from (2.6) and h_{00} from (2.7). Within this subsection, W_j denotes the four-byte subspace with j byte residual factors. For $i + j > 0$, set

$$h_{ij} = \sigma_b^{2(i+j)} \sum_{a=i}^{4} \sum_{b=j}^{4} \frac{n_i(a) n_j(b)}{(m-1)^{a+b-i-j}} \left[\sum_{d=1}^{i+j-1} \binom{a+b-i-j}{a+b-4-d} m^{d/2} \right]^2, \quad i + j > 0. \tag{4.1}$$

For fixed support, the shortened graph contains m^d points and projects injectively onto the residual coordinates. Its incidence matrix has Frobenius norm $m^{d/2}$. Following BLS's shortened-code method [4, Lemma 13] and Ren and Yu's exact-support calculation [5, Lemma 5], apply the triangle

inequality within each block and sum squares over orthogonal support blocks. Residual projections cannot increase this norm.

For the actual byte blocks R_{ij}, we have $\|R_{ij}\|_2 \le b_{ij}^{\mathrm{byte}}$ and $\|R_{ij}\|_{\mathrm{fr}}^2 \le h_{ij}$. Set

$$H_j = \sum_i h_{ij}, \qquad B_j = \sum_i (b_{ij}^{\mathrm{byte}})^2,$$

Then $\left\|R|_{W_j}\right\|_2^2 \le B_j$. For $G = R^*R$, the blocks satisfy

$$\|G_{jk}\|_{\mathrm{fr}} \le \sum_i \min\{b_{ij}^{\mathrm{byte}}\sqrt{h_{ik}}, \sqrt{h_{ij}}b_{ik}^{\mathrm{byte}}\}, \tag{4.2}$$

$$\|G_{jk}\|_{\mathrm{fr}}^2 \le \min\{B_jH_k, B_kH_j\}. \tag{4.3}$$

The first bound uses $\|A^*B\|_{\mathrm{fr}} \le \min(\|A\|_2 \|B\|_{\mathrm{fr}}, \|A\|_{\mathrm{fr}} \|B\|_2)$. The second applies the same inequality to stacked block columns. It follows that

$$\mathrm{Tr}(R^*R) \le \sum_j H_j < h := \frac{21813}{10^{10}}, \tag{4.4}$$

$$\mathrm{Tr}((R^*R)^2) \le \sum_{j,k} \min\left\{\left[\sum_i \min\{b_{ij}^{\mathrm{byte}}\sqrt{h_{ik}}, \sqrt{h_{ij}}b_{ik}^{\mathrm{byte}}\}\right]^2, B_jH_k, B_kH_j\right\} < t := \frac{91}{10^{16}}. \tag{4.5}$$

Here t bounds the second Gram moment and $\tau = \sqrt{t}$ is its square root. Evaluating these 5×5 bounds with square roots rounded upward gives the stated inequalities. Unitary multiplication and transposition preserve both moments, so the same h, t apply in every required direction.

□

4.2 Point bounds with a Fourier transform on one side

Define the byte character sum

$$\phi(\xi, a) = \sum_b D^{\mathsf{s}}(b, a)(-1)^{\langle \xi, b\rangle_2}.$$

For nonzero $\xi, a \in \mathbb{F}_2^8$, its absolute value is at most $\eta_b = 1/8$. Use the byte parameters of Section 2.8. For $u, x \in V \setminus \{0\}$, let v be the byte weight of the mask u and w that of the input difference x. Under $\zeta \sim (D^{\mathsf{s}})^{\otimes 4}\delta_x$,

$$\sqrt{q}(\mathcal{F}D^{\mathsf{S}})(u, x) = \mathbb{E}_\zeta \prod_{\ell=1}^{4} \phi(u_\ell, (M_{\mathrm{c}}\zeta)_\ell).$$

Using the prescribed-zero bounds $U_{w,k}^{\mathrm{byte}}$ from (2.12) and expanding the product of the character bounds gives

$$c_{vw}^{\mathcal{F},\to} = \sum_{k=0}^{\min(v,w-1)} \binom{v}{k} \eta_b^{v-k}(1-\eta_b)^k U_{w,k}^{\mathrm{byte}}, \qquad |\sqrt{q}(\mathcal{F}D^{\mathsf{S}})(u, x)| \le c_{vw}^{\mathcal{F},\to}. \tag{4.6}$$

The transpose identity $(D^{\mathsf{S}})^*\mathcal{F} = \mathcal{F}\widehat{A}^* D^{M_{\mathrm{c}}^\top}\widehat{A}^*$, with $\widehat{A} = \mathcal{F}(D^{\mathsf{s}})^{\otimes 4}\mathcal{F}$, exchanges v, w. The byte channel has the same constants r_b, p_b, η_b, so we may take $c_{vw}^{\mathcal{F}} = \min(c_{vw}^{\mathcal{F},\to}, c_{wv}^{\mathcal{F},\to})$.

The case $(v, w) = (1, 1)$ is enumerated exactly. Write the counting DDT with input differences indexing rows: $\mathrm{DDT}_{a,b} = 256D^{\mathsf{s}}(b, a)$, $\mathrm{AC}_{a,\xi} = \sum_b \mathrm{DDT}_{a,b}(-1)^{\langle \xi, b\rangle_2}$. For each $c \in \{1, 2, 3\}$, enumerate $\sum_d \mathrm{DDT}_{a,d}\mathrm{AC}_{cd,\xi}$ over nonzero a and ξ. The minimum is -1344 and the maximum is 816, giving $c_{11}^{\mathcal{F}} = 21/1024$. The reverse direction $R\mathcal{F}$ is computed separately. Transposition uses $(D^{\mathsf{s}})^*$

and the coefficients $9, 11, 13, 14$ of M_c^{-1}, giving minimum -1456 and maximum 896. For cases other than $(1,1)$, the bound is below $91/4096$. Since $(\mathcal{F}P)(u,x) = -1/(\sqrt{q}Q)$ for $u, x \neq 0$,

$$|(\mathcal{F}R)(u,x)| \le m_{vw}^{\mathcal{F}} := \frac{c_{vw}^{\mathcal{F}} + 1/Q}{\sqrt{q}}, \qquad \max\{|\mathcal{F}R|, |R\mathcal{F}|\} \le m_1 := \frac{91/4096 + 1/Q}{\sqrt{q}}. \tag{4.7}$$

The actual source uses the finer forward bound $m_{vw}^{\mathcal{F}}$; the general bound for the last four rounds uses m_1, which covers both directions. These extrema are obtained by integer enumeration of the AES DDT and autocorrelation table.

4.3 Byte-weight bounds for products of three or four factors

Use the byte-weight energies $E_{w_0,w}^{\mathrm{byte}}$ in (2.17). Define $\delta_{w_0}^{\mathrm{byte}} = \sum_w E_{w_0,w}^{\mathrm{byte}} - 1/Q$ and $N_{w_0} = \binom{4}{w_0} 255^{w_0}$. The four word-level directions share the row and column squared-norm bound $\delta_{w_0}^{\mathrm{byte}}$ at byte weight w_0. Unitary Fourier multiplication preserves the relevant row or column lengths, so $\mathcal{F}R$ has the same bound. Numerically,

$$q(\delta_1^{\mathrm{byte}}, \delta_2^{\mathrm{byte}}, \delta_3^{\mathrm{byte}}, \delta_4^{\mathrm{byte}}) < (0.172257, 0.140793, 0.136766, 0.132242).$$

Set $m_{vw}^R = 1/Q$ when $v + w < 5$, and $m_{vw}^R = r_*$ otherwise. In the first case, the entry of D^{S} is zero by the MixColumns branch number, so the residual entry is $-1/Q$. For $A = \mathcal{F}R$, use $m_{vw}^A = m_{vw}^{\mathcal{F}}$. For $A = R$ or $\mathcal{F}R$ and $r \in \{3, 4\}$, consider

$$M_{r,A}(y,x) = \prod_{\ell=1}^{r} A(L_\ell y, T_\ell x),$$

where the local maps are invertible and byte-monomial, hence preserve byte weight. The Schur comparison matrix between weight blocks is

$$Z_{vw}^{(r,A)} = \sqrt{\min\{N_w (m_{vw}^A)^r, \delta_v^{\mathrm{byte}} (m_{vw}^A)^{r-2}\} \min\{N_v (m_{vw}^A)^r, \delta_w^{\mathrm{byte}} (m_{vw}^A)^{r-2}\}}. \tag{4.8}$$

Bound $r-2$ factors pointwise and apply Cauchy–Schwarz to the other two along a row or column. Alternatively, multiply the pointwise bound by the number of entries. The Schur test gives the displayed expression. For $\mu = D^{\mathsf{S}}\delta_a$ with initial byte weight w_0, set $v_w^{\mathrm{byte},(w_0)} = \sqrt{E_{w_0,w}^{\mathrm{byte}}}$. Then

$$J_{r,A} := \max_{w_0=1}^{4} \left\| Z^{(r,A)} v^{\mathrm{byte},(w_0)} \right\|_2 \quad \Longrightarrow \quad \|M_{r,A}\mu\|_2 \le J_{r,A}. \tag{4.9}$$

Exact rational evaluation gives $J_{4,R} < 2^{-106}$ and $J_{4,\mathcal{F}R} < 2^{-97}$, as well as $J_{3,R} < 2^{-78}$, $J_{3,\mathcal{F}R} < 2^{-75}$. The recurrence uses the rational bounds of (4.9).

For a matrix A on $\ell_2(V)$ and $k > 0$, define the entry power sum

$$\Sigma_k(A) = \sum_{y,x \in V} |A(y,x)|^k.$$

We next bound $\Sigma_4(\mathcal{F}R)$. The total squared row norm e_v in weight class v satisfies $0 \le e_v \le N_v \delta_v^{\mathrm{byte}}$ and $\sum_v e_v \le h$. Hence

$$\Sigma_4(\mathcal{F}R) \le \max_{\substack{0 \le e_v \le N_v \delta_v^{\mathrm{byte}} \\ \sum e_v \le h}} \sum_v e_v \max_w (m_{vw}^{\mathcal{F}})^2 < s_4 := 3 \cdot 2^{-64}. \tag{4.10}$$

This is a linear optimization in four variables. Its maximum is obtained by allocating the available total in decreasing order of the nonnegative coefficients. All coefficients and limits are rational. We also have $\Sigma_4(R) \le r_*^2 h$, $\Sigma_3(R) \le r_* h$, and $\Sigma_3(\mathcal{F}R) \le \sqrt{h s_4}$.

4.4 Point bounds with Fourier transforms on both sides

The Fourier byte channel recalled in Section 2.8 is nonnegative and isolates zero. Integer enumeration shows that all 255 nonzero rows and columns have the same second, third, and fifth power sums,

$$c_2 = \frac{65}{8192}, \quad c_3 = \frac{10583}{134217728}, \quad c_5 = \frac{48193441}{4503599627370496},$$
$$\max_{a,b\neq 0}(\mathcal{F}_b D^{\mathsf{S}} \mathcal{F}_b)(b,a) = 1/64.$$

For nonzero word inputs and outputs with total byte weight $\ell = 5, 6, 7, 8$, the MixColumns graph shortens to an $[\ell, \ell - 4]$ MDS code. MDS Hölder gives the respective point bounds for $\mathcal{F}D^{\mathsf{S}}\mathcal{F}$,

$$c_5, \quad c_3^2, \quad c_2^2 c_3, \quad c_2^4.$$

For $\ell = 7$, Hölder bounds the 7/3-power sum of any row or column of the nonzero block by $c_2^{2/3} c_3^{1/3}$. Each of the four bounds is at most c_5. Since $\mathcal{F}P\mathcal{F} = P$, centering yields

$$\max |\mathcal{F}R\mathcal{F}| \leq \max(c_5 - 1/Q, 1/Q) < m_2 := 45/q. \tag{4.11}$$

The enumeration uses only the byte Walsh table; the word-level bound follows from the MDS inequality.

The matrices $R, \mathcal{F}R, R\mathcal{F}, \mathcal{F}R\mathcal{F}$ have zero rows and columns at the zero index. Every row and column has squared norm at most δ, and the total squared norm is at most h. For example, the column bound for $\mathcal{F}R$ follows from left unitary invariance, and its row bound follows from that of $\mathcal{F}R\mathcal{F}$. The four point bounds used for the last four rounds are r_*, m_1, m_1, m_2, respectively.

Proposition 4.2 (Local data)**.** *Under Lemma 2.2, the point bounds, product-response bounds, and power-sum bounds in (4.7), (4.8)–(4.10), and (4.11) hold. The two one-sided Fourier directions use their respective byte enumerations; m_1 bounds both.*

Proof. The bounds follow from zero-event expansions, the Schur test on byte-weight blocks, and MDS Hölder. Integer enumeration gives the required byte-table extrema and moments. Rational evaluation uses the upward bounds of Appendix B. □

5 Refined Four-Round Comparison Bounds

We retain exact activity sizes when bounding the blocks of two parallel superbox layers separated by the outer linear layer.

5.1 Blocks between residual subspaces

Let B bound $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U$. An initial candidate is $B_{(i,a),(j,b)} \leq \nu^{i+j} C_{(i,a),(j,b)}$. When $i, j \geq 1$, retain the residual channels on both sides. For fixed support, let u remaining coordinates carry $\frac{\delta_V}{\sqrt{q}}$ and the others carry δ_0. The shortened graph on the residual coordinates has dimension $d = i + j + u - 4$. For $0 < d < i + j$, applying (2.10) on both sides gives

$$b_{ij}(u) = q^{[-u+(d-i)_+ + (d-j)_+]/2} \nu^{\min(i,d)+\min(j,d)} \delta^{[(i-d)_+ + (j-d)_+]/2}. \tag{5.1}$$

For the other dimensions, the zero code or the full residual marginal is annihilated. Fourier duality exchanges u with $8 - i - j - u$, so the smaller of the two bounds from (5.1) may be used. With $v = (a - i) + (b - j)$, the resulting candidate is

$$\sqrt{n_i(a) n_j(b)} \sum_{u=0}^{v} \binom{v}{u} \alpha^u \beta^{v-u} \min\{b_{ij}(u), b_{ij}(8 - i - j - u)\}. \tag{5.2}$$

Equation (5.2) gives a bound for each of the 16 pairs of positive residual counts and all activity sizes. If either endpoint has one residual factor, the bounds of Section 6 give additional candidates.

5.2 Blocks between activity and residual subspaces

Using BLS's code-indicator expansion [4, Lemmas 12–13], we refine the bound of Ren and Yu [5, Theorem 2] by applying Lemma 2.4 to each exact activity pattern. Fix $i \geq 1$ output residual factors and a active words. Set $z = 4 - a$ and $u = a - i$. For an input unit indicator of a k-dimensional coordinate subspace, expanding the nonzero restrictions on the remaining active output coordinates introduces at most u additional zero constraints. The response is bounded by

$$s(i, a; k) = Q^{-u/2} \sum_{v=0}^{u} \binom{u}{v} q^{-(z+v)/2} \Phi(i, k - z - v). \tag{5.3}$$

For these homogeneous residual-code terms, set $\Phi(i, d) = 0$ when $d \leq 0$ or $d \geq i$: the zero point and the uniform residual marginal are annihilated. This convention does not apply to an affine singleton. The factor $q^{-(z+v)/2}$ is the ratio between the original and shortened unit-indicator normalizations. Expanding every $\frac{\delta_{V \setminus \{0\}}}{\sqrt{Q}}$ in an input with exactly b active words gives

$$a(i, a; b) = \sqrt{n_i(a) \binom{4}{b}} \sum_{k=0}^{b} \binom{b}{k} \alpha^k \beta^{b-k} s(i, a; k). \tag{5.4}$$

This bounds $B_{(i,a),(0,b)}$. Taking the adjoint and using the inverse linear layer gives the reverse-direction bound. All $\Phi(1, d)$ vanish, recovering the zero blocks between activity and one residual factor proved by BLS [4, Lemma 12]. The activity-to-activity block equals the corresponding centered block of C. Together with the bounds for an endpoint with one residual factor, this defines every entry of B.

6 Affine Code Bounds for an Endpoint with One Residual Factor

We bound blocks of $(D^{\mathsf{S}})^{\otimes 4} D^{L} (D^{\mathsf{S}})^{\otimes 4}$ with one input residual factor and i output residual factors. Fix the residual positions and exact activity patterns and expand $\frac{\delta_{V \setminus \{0\}}}{\sqrt{Q}}$. The input has k factors $\frac{\delta_V}{\sqrt{q}}$; the output has t_{out} factors $\frac{\delta_V}{\sqrt{q}}$ and $z = 4 - i - t_{\text{out}}$ factors δ_0. Separate the input R, reserving its operator-norm factor ν. Once its output value x is fixed, the residual output coordinates of the $[8, 4]$ graph code satisfy an affine MDS constraint. Set

$$d = k - z, \qquad u = k + t_{\text{out}}$$

for the remaining dimension and the number of uniform contractions, respectively.

For $0 < d < i$, let $\mathcal{C}$ be the resulting affine $[i, d]$ MDS code. The Gram bound applies to its unit indicator. In low dimension, apply MDS Hölder to $\mathcal{C} \times \mathcal{C}$. In high dimension, Fourier duality adds factors of modulus one; diagonal unitary conjugation preserves column norms and second Gram moments. Set $\ell_0 = \min(d, i - d)$. For each fixed x, the output column norm is at most

$$q^{(d-u)/2} \left[(\tau^2/q)^{\ell_0} \delta^{i-2\ell_0} \right]^{1/2} = q^{(i-4)/2} \left[(\tau^2/q)^{\ell_0} \delta^{i-2\ell_0} \right]^{1/2}.$$

Summing squared column norms over at most q values of x gives the Frobenius, and hence operator-norm, bound

$$b_i(k, z) = \left[q^{i-3} (\tau^2/q)^{\ell_0} \delta^{i-2\ell_0} \right]^{1/2}.$$

The factor q^{i-3} already includes the factor q from summing over input columns.

For $d = 0$, varying x traces an $[i, 1]$ graph. The Schur–Gram argument gives the stronger bound

$$b_i(k, z) = q^{(i-4)/2}\nu\delta^{(i-1)/2}.$$

If $z > k$, only the zero point remains and the residual factors annihilate it. If $d \geq i$, the residual output marginal is uniform and is annihilated by $R\delta_V = 0$. In both cases, $b_i(k, z) = 0$.

Applying Fourier transforms to all coordinates exchanges k with $3-k$ among the three remaining input positions and exchanges z with t_{out} at the output. Both directions give valid bounds, so each term may use

$$\nu \min\{b_i(k, 4 - i - t_{\text{out}}), b_i(3 - k, t_{\text{out}})\}.$$

The leading ν belongs to the input R separated above. The additional ν in the case $d = 0$ belongs to the output-code response. Sum with weights $\binom{b-1}{k}\binom{a-i}{t_{\text{out}}}\alpha^{k+t_{\text{out}}}\beta^{a-i+b-1-k-t_{\text{out}}}$ and multiply by $\sqrt{n_i(a)n_1(b)}$ to obtain the comparison coefficient. Adjoint operators and inverse diffusion give the corresponding bound when the output has one residual factor.

Proposition 6.1 (Four-round comparison matrix)**.** *Start each entry with $\nu^{i+j}C_{(i,a),(j,b)}$. For $i, j \geq 1$, also use (5.2). If one endpoint has one residual factor, include the affine-code candidate above. If exactly one endpoint has no residual factor, include (5.4) or its adjoint version. If neither endpoint has a residual factor, use the corresponding entry of C. The matrix B obtained by taking the minimum of these candidates satisfies*

$$\|E_{i,a}((D^{\mathsf{S}})^{\otimes 4}D^L(D^{\mathsf{S}})^{\otimes 4} - U)E_{j,b}\|_2 \leq B_{(i,a),(j,b)}.$$

Proof. Take the minimum of the bounds derived in Section 5 and above. The operator $(D^{\mathsf{S}})^{\otimes 4}$ fixes the activity space pointwise, and $(D^{\mathsf{S}})^{\otimes 4}U = U(D^{\mathsf{S}})^{\otimes 4} = U$. Its activity-to-activity block therefore equals the corresponding block of $D^L - U$. □

7 The First Four Rounds by Initial Word Weight

Fix a nonzero input difference a_0 in the grouped coordinates, with $w \in \{1, 2, 3, 4\}$ active words. Set $g = D^L(D^{\mathsf{S}})^{\otimes 4}\delta_{a_0}$, $p_4 = (D^{\mathsf{S}})^{\otimes 4}g$, and $f_4 = p_4 - \frac{\delta_\Omega}{n}$. Let $g_{i,a}$ and $v^{(4)}_{i,a}$ bound the block norms of $g - \frac{\delta_\Omega}{n}$ and f_4, respectively. The bounds are uniform over initial support positions. For $w = 1$, the local computation covers all four nonzero byte weights within the initial word.

7.1 Activity coefficients and total mass

Fix an exact activity pattern A with $|A| = a$ and $z = 4 - a$. Let $\lambda_A = \sum_{x\in\Omega_A} g(x)$ and $r_a = Q^a/n$. The channel $(D^{\mathsf{S}})^{\otimes 4}$ preserves exact activity patterns, so p_4 has the same mass λ_A on Ω_A. The absolute activity coefficient is $Q^{-a/2}|\lambda_A - r_a|$. Define

$$\overline{U}_{w,z} = \begin{cases} 1, & z = 0, \\ 0, & z \geq w, \\ \rho p^{z-1}, & 1 \leq z < w. \end{cases}$$

This upper-bounds the zero-event estimates in (2.12). Hence

$$v^{(4)}_{0,a} \leq \sqrt{\binom{4}{a}Q^{-a}}\max\{r_a, |\overline{U}_{w,z} - r_a|\}. \tag{7.1}$$

For $w \geq 2$ and $a = 3$, use the additional bounds

$$e_w = 5^{-w/2}, \quad U_{w,2} = \begin{cases} 0 & w = 2, \\ p\rho & w = 3, \\ \rho^2 & w = 4, \end{cases} \quad \ell_w = (1 - e_w)/q - 3U_{w,2}, \quad u_w = (1 + e_w)/q.$$

Equation (2.16), with $q\rho - 1 = 1/5$, bounds the deviation of a specified zero-word probability from $1/q$ by e_w/q. Subtracting at most three events with two zero words gives $\lambda_A \in [\ell_w, u_w]$, with a negative lower endpoint replaced by 0. In (7.1), the maximum may therefore be replaced by $\max(|\ell_w - r_3|, |u_w - r_3|)$ when smaller. The total-mass identity further bounds $v_{0,4}^{(4)}$. Since $(D^{\mathsf{S}})^{\otimes 4}$ fixes activity vectors, $g_{0,a}$ has the same bound.

7.2 One active source word

For $w = 1$, the MDS property makes all four words nonzero after linear mixing. Residual components with $a < 4$ therefore vanish. The component with one residual factor uses (2.4). For two or more residual factors, a common-source Gram expansion gives

$$(v_{1,4}^{(4)})^2 \leq \frac{4 \cdot 13}{q^2 Q^3}, \tag{7.2}$$

$$(v_{i,4}^{(4)})^2 \leq \binom{4}{i} Q^{i-4} p^2 \min(t, \nu^2 h)(\rho - 1/Q)^{i-2}, \quad 2 \leq i \leq 4. \tag{7.3}$$

For i fixed Gram factors, apply Cauchy–Schwarz over (x, x') to two factors, giving $\mathrm{Tr}((R^\top R)^2) \leq t$. Bound the others by the squared column norm $\rho - 1/Q$. The source-probability product is at most p^2. The alternative bound $\nu^2 h$ follows from the common-source estimate (2.14). The remaining $4 - i$ projected probability factors each have squared norm $1/Q$. There are $\binom{4}{i}$ orthogonal residual-position choices, giving (7.3).

Before the final $(D^{\mathsf{S}})^{\otimes 4}$, we have

$$g_{i,4}^2 \leq \binom{4}{i} Q^{i-4} \begin{cases} \rho - 1/Q, & i = 1, \\ \rho, & i \geq 2. \end{cases} \tag{7.4}$$

For one residual factor, this is the centered source column norm. For more factors, apply (2.10) with $d = 1$ and channel $Z = I - P$, whose norm and squared column norms are at most one. The common-source embedding has norm at most one, leaving the source squared norm ρ. At this point only the projection has been applied, not the next R.

7.3 Several sources with restricted events

For $w \geq 2$, fix a and set $z = 4 - a$. Consider positive residual blocks $i \geq 1$. If $z \geq w$, these blocks vanish because the support is unreachable and $E_{i,a} \frac{\delta_\Omega}{n} = 0$. Centered activity components on unreachable supports remain: for a fixed A with $\lambda_A = 0$, the coefficient is $Q^{-a/2}(\lambda_A - r_a) = -Q^{a/2}/n$.

When $z < w$, set $\ell = \min(w - z, i)$. Apply (2.13) with the nonzero restrictions of the exact activity pattern included in the source event. This gives

$$(v_{i,a}^{(4)})^2 \leq n_i(a) Q^{i-a} p^{2z} (\rho\nu^2)^\ell \delta^{i-\ell}. \tag{7.5}$$

The factors p^{2z}, $(\rho\nu^2)^\ell$, and $\delta^{i-\ell}$ account for the zero constraints, information coordinates, and remaining residual coordinates. The factor Q^{i-a} comes from the nonzero-uniform output contractions, and $n_i(a)$ counts the orthogonal position choices.

Alternatively, expand the remaining nonzero restrictions by inclusion–exclusion. Set $b^{\mathrm{src}}_{w,i}(k) = 0$ if $k \geq w$, and otherwise use

$$b^{\mathrm{src}}_{w,i}(k) = \left(\frac{6}{5}\right)^2 \left(\frac{53}{4}\right)^k q^{-k-i/2} \nu^{\min(w-k,i)}.$$

This upper-bounds the square root of the bound (2.13), since $\rho^{\ell/2}\delta^{(i-\ell)/2} \leq q^{-i/2}(6/5)^2$ for $\ell \leq 4$. For $i = 1, k = 0$, use the stronger candidate $b^{\mathrm{src}}_{w,1}(0) = \nu\sqrt{1/(5^w q)}$: the centered source convolution has squared norm at most $1/(5^w q)$ by (2.15), followed by the residual norm factor ν. This gives

$$v^{(4)}_{i,a} \leq \sqrt{n_i(a) Q^{i-a}} \sum_{v=0}^{a-i} \binom{a-i}{v} b^{\mathrm{src}}_{w,i}(z+v). \tag{7.6}$$

Take the smaller of (7.5) and (7.6). Before the final $(D^{\mathsf{S}})^{\otimes 4}$, use

$$g^2_{i,a} \leq n_i(a) Q^{i-a} p^{2z} \rho^{\ell} (1 - 1/Q)^{i-\ell}. \tag{7.7}$$

Apply (2.13) to the same restricted density with output channel $I - P$. Its norm is one and its nonzero columns have squared norm $1 - 1/Q$. The factor ρ bounds the source density, not the channel.

Proposition 7.1 (Bounds after four rounds)**.** *Fix a nonzero difference a_0 with word weight w in the grouped coordinates. Bound the activity components by (7.1), its one-zero-event refinement, and the mass constraint of Lemma 3.2. For $w = 1$, use (7.3) and (7.4) for residual components. For $w \geq 2$, use the minimum of the valid candidates in (7.5)–(7.7). The resulting nonnegative vectors satisfy*

$$\|E_{i,a}((D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0} - \frac{\delta_\Omega}{n})\|_2 \leq v^{(4)}_{i,a}, \qquad \|E_{i,a}(D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0} - \frac{\delta_\Omega}{n})\|_2 \leq g_{i,a}.$$

Proof. The four word weights cover every nonzero input difference. The bounds treat activity components and residual components from one or several sources. Each candidate uses the same exact activity pattern and normalization. Taking coordinatewise minima and applying the mass constraint preserve the upper bounds. □

8 Bounds Retaining the Common Source

We bound the contributions produced by the actual four-round distribution. These estimates bound the actual source contribution and are sharper than an operator-norm bound on the same subspace. Write

$$q = 2^{32}, \quad Q = q - 1, \quad D^{\mathsf{S}} = P + R, \quad P = \delta_0 \delta_0^* + Q^{-1} \delta_{V\setminus\{0\}} \delta^*_{V\setminus\{0\}},$$

and let ν, δ, h, τ satisfy

$$\|R\|_2 \leq \nu, \quad \max_x \|R\delta_x\|_2^2, \ \max_y \|R^*\delta_y\|_2^2 \leq \delta, \quad \|R\|_{\mathrm{fr}}^2 \leq h, \quad \mathrm{Tr}((R^*R)^2) \leq \tau^2.$$

The same properties hold for $R, R^*, \mathcal{F}R\mathcal{F}, \mathcal{F}R^*\mathcal{F}$, including the required row and column bounds after a one-sided Fourier transform. Use $p = 53/(4q)$ and $\rho = 6/(5q)$. Each independent source $\mu = D^{\mathsf{S}}\delta_a$, for $a \neq 0$, satisfies $\|\mu\|_\infty \leq p$, $\|\mu\|_2^2 \leq \rho$, and $\mu(0) = 0$.

8.1 Local Weyl–Gram bounds

Let $p_4 = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0}$. We use the common-source representation (2.11) and the byte-monomial coordinate maps of Section 2.4.

For $\xi, \eta \in V$ and $g \in \ell_2(V)$, define translation and character multiplication by

$$(X_\eta g)(a) = g(a+\eta), \qquad (Z_\xi g)(a) = \chi_\xi(a) g(a), \qquad \chi_\xi(a) = (-1)^{\langle \xi, a \rangle_2}.$$

For $x, x' \in V$, set

$$G_{\xi,\eta}(x,x') = \sum_a R(a,x)\overline{R(a+\eta,x')}\chi_\xi(a).$$

Lemma 8.1 (Local Weyl–Gram bounds)**.** *Under the local squared-norm and moment assumptions, the entries of $G_{\xi,\eta}$ have absolute value at most δ and its Frobenius norm is at most τ. The full parameter sum and the pure translation and modulation sums satisfy (8.1) and (8.3), respectively.*

Proof. Cauchy–Schwarz on the columns gives $|G_{\xi,\eta}(x,x')| \le \delta$. Applying character orthogonality in ξ and summing over η gives

$$\sum_{\xi,\eta} |G_{\xi,\eta}(x,x')|^2 = q \|R\delta_x\|_2^2 \|R\delta_{x'}\|_2^2,$$
$$\sum_{\xi,\eta} \|G_{\xi,\eta}\|_{\mathrm{fr}}^2 \le q h^2. \tag{8.1}$$

For a unitary matrix W, set $A = RR^*$. The Schatten Cauchy–Schwarz inequality gives

$$\|R^* W R\|_{\mathrm{fr}}^2 = \mathrm{Tr}(W^* A W A) \le \mathrm{Tr}(A^2) \le \tau^2. \tag{8.2}$$

Hence $\max_{\xi,\eta} \|G_{\xi,\eta}\|_{\mathrm{fr}} \le \tau$. The pure translation and modulation sums satisfy the sharper identities

$$\sum_\eta \|R^* X_\eta R\|_{\mathrm{fr}}^2 = q \sum_y \|\mathrm{row}_y(\mathcal{F}R)\|_2^4 \le q\delta h, \qquad \sum_\xi \|R^* Z_\xi R\|_{\mathrm{fr}}^2 \le q\delta h. \tag{8.3}$$

The first identity diagonalizes translations by Fourier transformation. The second uses character orthogonality in the original coordinates. Equations (8.1)–(8.3) bound both individual Gram entries and the Frobenius sums required for several sources. □

8.2 Four residual inputs with zero and uniform contractions

Retain the first two output words and contract the third and fourth against $\frac{\delta_V}{\sqrt{q}}$ and δ_0, respectively. This defines

$$P_{V,0} : \ell_2(V^4) \to \ell_2(V^2), \qquad (P_{V,0} f)(y_1, y_2) = q^{-1/2} \sum_{z \in V} f(y_1, y_2, z, 0).$$

Let $\iota_r : V \to V^4$ insert a word at coordinate r and zero elsewhere, and let $\mathrm{pr}_r : V^4 \to V$ extract that coordinate. Set $v = L^{-1}\iota_3$ and $\kappa_4 = \mathrm{pr}_4 L$. Then $\kappa_4 v = 0$, and every word block of these two maps is invertible. Expanding the squared norm of $f \in \ell_2(V^4)$ gives

$$\begin{aligned} \|P_{V,0} D^L f\|_2^2 &= q^{-1} \sum_{\eta \in V} \sum_{a : \kappa_4 a = 0} f(a)\overline{f(a+v\eta)} \\ &= q^{-2} \sum_{\xi,\eta \in V} \sum_{a \in V^4} f(a)\overline{f(a+v\eta)}\chi_\xi(\kappa_4 a). \end{aligned} \tag{8.4}$$

The first factor q^{-1} comes from the two occurrences of the unit vector $\frac{\delta_V}{\sqrt{q}}$. The second comes from the Fourier expansion of the zero constraint.

The local translation at word r is $\eta_r = \mathrm{pr}_r(v\eta)$. More generally, for uniform contraction positions $j_1, \dots, j_k$ and translation parameters $\theta_1, \dots, \theta_k \in V$, define

$$\eta_r = \mathrm{pr}_r\, L^{-1} \sum_{s=1}^{k} \iota_{j_s} \theta_s.$$

These are the translations in the corresponding local Gram factors.

For $w = 1$, substitute the actual component $f = f_{\{1,2,3,4\}}$ from (2.11). The expansion contains four G factors with invertible coordinate changes in (ξ, η, x, x'). Bound $\mu(x)\mu(x')$ by p^2, apply four-factor Hölder, and use $\sum |G|^4 \le \delta^2 \sum |G|^2 \le \delta^2 q h^2$ to obtain

$$\|P_{V,0} D^L f_{\{1,2,3,4\}}\|_2 \le \frac{p\delta h}{\sqrt{q}}. \tag{8.5}$$

This bounds the vector before the next residual channels. Applying R on the two retained words then contributes ν^2.

For $2 \le w \le 4$, the paired sources in the squared expansion form an $[4, w]$ MDS code over V^2. For fixed (ξ, η), MDS Hölder gives a product of four $\ell_{4/w}$ norms. Each matrix has q^2 entries, and

$$\left(\sum_{x,x'} |G(x,x')|^{4/w} \right)^w \le q^{2w-4} \|G\|_{\mathrm{fr}}^4.$$

Hölder over the Weyl parameters, together with (8.1) and (8.2), gives $\sum_{\xi,\eta} \|G_{\xi,\eta}\|_{\mathrm{fr}}^4 \le q\tau^2 h^2$. Let $p_{4,W_4} = E_{4,4} p_4$ be the projection of the four-round distribution onto W_4, extended by zero at the all-zero difference when viewed in $\ell_2(V^4)$. This vector has total mass zero and need not have nonnegative entries. Retaining the normalization in (8.4) yields

$$\|P_{V,0} D^L p_{4,W_4}\|_2 \le p^w q^{w-5/2} \tau h. \tag{8.6}$$

The four words after diffusion are correlated; the bound uses only their joint MDS structure.

With one residual output word, three words are contracted. Suppose k carry $\frac{\delta_V}{\sqrt{q}}$ and $3-k$ carry δ_0, with $k = 1$ or 2. The local translation and modulation maps are surjective, and their joint map has q preimages per image. This cancels the additional normalization q^{-1}. Bounds (8.5) and (8.6) still apply, followed by one factor ν. The cases with only $\frac{\delta_V}{\sqrt{q}}$ or only δ_0 contractions are treated separately.

8.3 Uniform or zero contractions for four residual inputs

Let A be R or $\mathcal{F}R$, and choose an entry bound $m_A \ge \max_{y,x \in V} |A(y,x)|$. We use $m_R = r_*$ and $m_{\mathcal{F}R} = m_1$. The entry power sum $\Sigma_k(A)$ is defined in Section 4.

Fix the retained and contracted positions. For $d = 1, 2$, let $j_1, \dots, j_d$ be the free output positions of the kernel under consideration, and set

$$L_r^{(d)} : V^d \to V, \qquad L_r^{(d)}(y_1, \dots, y_d) = \mathrm{pr}_r\, L^{-1} \sum_{s=1}^{d} \iota_{j_s} y_s.$$

The other output coordinates are zero in this parameterization. The four maps jointly parameterize an additive $[4, d]$ MDS code, and any d coordinates give a bijection from V^d to V^d. In particular, $L_r^{(1)}$ is bijective. A full output Fourier transform exchanges zero and uniform contractions. In that representation, replace L^{-1} by $L^\top$, use the resulting free positions, and take $A = \mathcal{F}R$. The notation $L_r^{(d)}$ refers to the fixed positions and code in each case; d alone does not determine the map.

For one source and one retained output word, contracting all other words at zero gives the matrix

$$M_{4,R}(y,x) = \prod_{r=1}^{4} R(L_r^{(1)}y, T_r x).$$

Contracting them all against unit uniform vectors gives the same expression after an output Fourier transform, with R replaced by $\mathcal{F}R$ and no additional power of q. Equation (2.2) ensures byte-weight preservation. The local response bounds $\|M_{4,R}\mu\|_2$ and $\|M_{4,\mathcal{F}R}\mu\|_2$ for the actual source $\mu = D^{\mathsf{S}}\delta_a$ therefore apply.

One uniform and two zero contractions give

$$M_1(y,x) = q^{-1/2} \sum_z \prod_{r=1}^{4} R(L_r^{(2)}(y,z), T_r x),$$

where the output parameters form a $[4,2]$ MDS code. For fixed y, the variables (x,z) map bijectively to the two indices of each local factor, giving absolute row sum at most $q^{-1/2}\Sigma_4(R)$. For fixed x, MDS Hölder gives absolute column sum at most $q^{-1/2}\delta^2$. The Schur test and $\|\mu\|_2 \le \sqrt{\rho}$ yield

$$\|M_1\mu\|_2 \le q^{-1/2}\delta\sqrt{\Sigma_4(R)\rho} \le q^{-1/2}\delta r_* \sqrt{h\rho}. \tag{8.7}$$

Fourier duality gives the bound for two uniform and one zero contraction, $q^{-1/2}\delta\sqrt{\Sigma_4(\mathcal{F}R)\rho}$. Each bound is compared with (8.5), and the smaller is used.

Retaining two output words and contracting the others either all at zero or all uniformly gives a two-dimensional MDS code. With its fixed parameterization, define

$$M_{4,A}^{(2)} : \mathbb{C}[V] \to \mathbb{C}[V^2],$$
$$M_{4,A}^{(2)}((y_1,y_2),x) = \prod_{r=1}^{4} A(L_r^{(2)}(y_1,y_2), T_r x).$$

This matrix has q^2 rows and q columns, with no additional normalization. Zero contractions use $A = R$. Uniform contractions use $A = \mathcal{F}R$ and the corresponding parameterization after an output Fourier transform. For source byte weight s, suppose $|A(y,x)| \le m_s$ holds for all x in that class. The absolute row and column sums are at most $m_s^2\delta$ and δ^2. The actual source response therefore satisfies

$$\|M_{4,A}^{(2)}\mu\|_2 \le \delta^{3/2} \sum_s m_s \|\mu_s\|_2. \tag{8.8}$$

Here μ_s is the byte-weight component of the actual source, bounded by the byte-weight estimate (2.17).

For $w \ge 2$, the absolute row sum over the input $[4,w]$ code is at most $q^{w-2}\delta^2$. The absolute column sum is at most $m_A^2\delta$ for a one-dimensional output code and δ^2 for a two-dimensional output code. Multiplying by the source norm $\rho^{w/2}$ gives

$$\begin{aligned} &\text{one output word:} && m_A\delta^{3/2}(q\rho)^{w/2}/q, \\ &\text{two output words:} && \delta^2(q\rho)^{w/2}/q. \end{aligned} \tag{8.9}$$

For a mixture of zero and uniform contractions with one retained output word, omit one contraction by a unit vector and apply the two-dimensional bound. This gives another candidate to compare with (8.6) for multiple sources in $W_4 \to W_1$.

8.4 Three residual inputs and a nonzero uniform word

Fix the three residual input positions I and let ℓ be the omitted position. For one source, the remaining factor is the probability vector $\frac{\delta_{V\setminus\{0\}}}{Q}$:

$$f(a) = Q^{-1}\delta_{V\setminus\{0\}}(a_\ell)\sum_x \mu(x)\prod_{r\in I} R(a_r, T_r x).$$

In the squared expansion, this position contributes

$$G_0(\xi,\eta) = Q^{-2}\sum_{a\neq 0,\, a+\eta\neq 0}\chi_\xi(a), \qquad |G_0(\xi,\eta)| \le \begin{cases} Q^{-1}, & \xi = 0, \\ 2Q^{-2}, & \xi \neq 0. \end{cases} \tag{8.10}$$

The first case sums over between $q-2$ and Q points. The second uses the zero sum of a nontrivial character and removes at most two points.

For the mixed contraction with output in W_2, setting $\xi = 0$ leaves only translation parameters in the three residual words. Three-factor Hölder with $|G|^3 \le \delta|G|^2$ and (8.3) bounds the total third moment by $q\delta^2 h$. For $\xi \neq 0$, extending to all Weyl parameters and using (8.1) gives $q\delta h^2$. Including p^2, (8.10), and q^{-2} yields

$$\|P_{V,0}D^L f\|_2^2 \le p^2\left(\frac{\delta^2 h}{qQ} + \frac{2\delta h^2}{qQ^2}\right). \tag{8.11}$$

For W_1 with two uniform and one zero contraction, an additional translation parameter gives q preimages per image in the local moment sum. This cancels the additional normalization, leaving the same bound (8.11).

For W_1 with one uniform and two zero contractions, use

$$(P_{V,0,0}f)(y) = q^{-1/2}\sum_{z\in V} f(y,z,0,0).$$

Setting the omitted word's local modulation to zero still leaves one free modulation parameter. Each residual word's local modulation map remains surjective because the relevant 2×2 MDS minor is nonzero. The first term can therefore use the full Weyl moment:

$$\|P_{V,0,0}D^L f\|_2^2 \le p^2\left(\frac{\delta h^2}{q^2 Q} + \frac{2\delta h^2}{qQ^2}\right). \tag{8.12}$$

Bounds (8.11) and (8.12) precede the final application of R to the retained output words.

8.5 Several sources and the omitted zero word

For $w \ge 2$, separate the input blocks $(3,4)$ and $(3,3)$. In $(3,4)$, the omitted word contributes $\frac{\delta_{V\setminus\{0\}}}{Q}$ and imposes a nonzero restriction on the source. This indicator can be dropped after taking absolute values. For $w = 2,3$, the other three words form an $[3,w]$ MDS image. For $w = 4$, the projection of paired sources onto these words has q^2 preimages per image. Number the three retained words by $r = 1,2,3$. For fixed contraction parameters, let G_r denote the local Weyl–Gram matrix at word r, and let (x_r, x_r') be the corresponding word differences after the first linear layer. All cases satisfy

$$\sum_{\text{source pairs}}\prod_{r=1}^{3}|G_r(x_r,x_r')| \le q^{2w-3}\prod_{r=1}^{3}\|G_r\|_{\mathrm{fr}}. \tag{8.13}$$

For $w = 2,3$, apply MDS Hölder over an alphabet of size q^2 and compare the $\ell_{3/w}$ and ℓ_2 norms. For $w = 4$, sum independently over the three coordinates and multiply by the number of preimages. The

resulting exponent is the same. Sum $\|G_{\xi,\eta}\|_{\mathrm{fr}}^3 \le \tau\|G_{\xi,\eta}\|_{\mathrm{fr}}^2$ over the parameters of the contraction and apply (8.1)–(8.3). Before the final $(D^{\mathsf{S}})^{\otimes 4}$, the squared bounds are

$$\begin{aligned} &W_2, k=1 \text{ or } W_1, k=2: && p^{2w}q^{2w-4}\tau\left(\tfrac{\delta h}{Q}+\tfrac{2h^2}{Q^2}\right), \\ &W_1, k=1: && p^{2w}\tau h^2\left(\tfrac{q^{2w-5}}{Q}+\tfrac{2q^{2w-4}}{Q^2}\right). \end{aligned} \tag{8.14}$$

Here k counts the $\frac{\delta_V}{\sqrt{q}}$ factors among the contracted output words.

In $(3,3)$, the omitted word contributes δ_0 and the source is restricted to zero there. The source MDS code shortens to dimension $w-1$, while its density is still bounded by p^w. The omitted Gram factor is $\mathbf{1}[\eta_\ell = 0]$. For $W_2, k=1$ or $W_1, k=1$, it sets all translation parameters to zero. For $W_1, k=2$, one translation parameter remains. Together with local modulation it gives the full Weyl family; surjectivity follows from the nonzero 2×2 minor.

For $w=2$, the shortened paired-source code has dimension one, so use the entrywise third-moment bound $|G|^3 \le \delta|G|^2$. For $w \ge 3$, use (8.13) with $w-1$ and the third moment of the Frobenius norm. This gives

	$W_2, k=1$ or $W_1, k=1$	$W_1, k=2$
$w=2$	$p^4\delta^2 h/q$	$p^4\delta h^2/q^2$
$w\ge 3$	$p^{2w}q^{2w-6}\tau\delta h$	$p^{2w}q^{2w-7}\tau h^2$

. (8.15)

The zero branch is absent when $w=1$. Take square roots in (8.11)–(8.15) before multiplying by ν^i, where i counts retained output residual factors.

8.6 Contributions to the activity space

Suppose pulling a zero/uniform tensor through the linear layer gives a $[4,d]$ MDS code $\mathcal{C}$ with unit indicator $q^{-d/2}\delta_{\mathcal{C}}$. For one source and four residual factors, the responses at $d=0,4$ vanish. The others satisfy

$$A_1 = \frac{p\Sigma_4(R)}{\sqrt{q}}, \qquad A_2 = \frac{p\delta h}{q}, \qquad A_3 = \frac{p\Sigma_4(\mathcal{F}R)}{\sqrt{q}}. \tag{8.16}$$

For $d=1$, the code line gives four bijective coordinate changes. Apply Hölder over (x,y) and bound $\mu(x)$ by p. For $d=2$, fix x and apply MDS Hölder to the output code, giving a product of four column norms. Hölder over x then gives at most $\sum_x \|R\delta_x\|_2^4 \le \delta h$. For $d=3$, Fourier duality maps the unit code indicator to a unit dual-code-line indicator and changes the columns to $\mathcal{F}R\delta_x$. The source μ is unchanged, so the local matrix is not $\mathcal{F}R\mathcal{F}$.

For $j=3$ residual words, expand the remaining factor $\frac{\delta_{V\setminus\{0\}}}{Q} = (\delta_V - \delta_0)/Q$. For $j=2$, expand both remaining factors. Let k_0 count the zero coordinates chosen in this expansion, with $0 \le k_0 \le 4-j$, and put $e = d - k_0$. The terms with $e \le 0$ vanish at zero, and those with $e \ge j$ vanish by the zero column sums. Let $B_{d,j}$ bound the response for fixed residual positions. Termwise expansion gives

	$d=1$	$d=2$	$d=3$
$j=2$	$\frac{ph}{Q^2\sqrt{q}}$	$\frac{2ph}{Q^2 q}$	$\frac{ph}{Q^2 q^{3/2}}$
$j=3$	$\frac{p\Sigma_3(R)}{Q\sqrt{q}}$	$\frac{p}{Q}\left(\frac{\Sigma_3(\mathcal{F}R)}{\sqrt{q}}+\frac{\Sigma_3(R)}{q}\right)$	$\frac{p\Sigma_3(\mathcal{F}R)}{Qq}$.

(8.17)

All terms with $j=1$ vanish. Here $\Sigma_3(R) \le r_* h$, $\Sigma_3(\mathcal{F}R) \le \sqrt{h\Sigma_4(\mathcal{F}R)}$. Fourier duality introduces a factor $\sqrt{q}$ for the unnormalized sum over a two-dimensional code on three coordinates. This gives the powers of q in (8.17).

For $w \ge 2$ and four residual factors, fix the output codeword. MDS Hölder on the input $[4,w]$ code and the $\ell_{4/w}$–ℓ_2 norm comparison give q^{w-2} times four row norms. Summing over the output

code gives δh in dimension one and h^2 in dimension two. Including the density bound p^w and the unit-indicator normalization yields

$$A_1 = A_3 = p^w q^{w-5/2}\delta h, \qquad A_2 = p^w q^{w-3} h^2. \tag{8.18}$$

Dimension three uses output Fourier duality, with the same row bounds δ, h.

8.7 Exact activity patterns and position counts

The preceding contractions use $\frac{\delta_V}{\sqrt{q}}, \delta_0$. The orthogonal decomposition into 14 subspaces uses $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$, with

$$\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}} = \alpha\frac{\delta_V}{\sqrt{q}} - \beta\delta_0, \qquad \alpha = \sqrt{q/Q}, \quad \beta = Q^{-1/2}. \tag{8.19}$$

For i residual output words and exact activity size a, suppose v_k bounds the fixed-position term with k uniform contractions. The fixed-position block is bounded by

$$\sum_{k=0}^{a-i} \binom{a-i}{k} \alpha^k \beta^{a-i-k} v_k.$$

There are $n_i(a) = \binom{4}{i}\binom{4-i}{a-i}$ orthogonal output-position blocks, giving the factor $\sqrt{n_i(a)}$. The three residual input positions have four choices, combined by the triangle inequality with factor 4. Four residual inputs have only one position choice. The input and output position counts therefore enter differently.

Activity outputs can also be expanded by exact support. Let f be a fixed-position source contribution considered above. Suppose A_k bounds $|\langle q^{-k/2}\delta_{\mathcal{C}}, f\rangle|$ for every dimension-k code $\mathcal{C}$ obtained by pulling back a zero/uniform tensor through L. The aggregate activity component satisfies

$$\|E_{0,a} D^L f\|_2 \le \sqrt{\binom{4}{a}} \sum_{k=1}^{\min(a,3)} \binom{a}{k} \alpha^k \beta^{a-k} A_k. \tag{8.20}$$

The terms $k = 0, 4$ vanish by the residual zero rows, zero columns, and zero sums. Equation (8.19) holds as an identity on each exact support. For one source, multiply (8.17) by the position count $\binom{4}{j}$, sum, and add (8.16). For several sources with four residual factors, use (8.18).

Combining the bounds. For four residual inputs, combine (8.5)–(8.9), the byte-weight source-response bounds, and (8.19). For three residual inputs, combine the Weyl bounds (8.11)–(8.15) with (8.19) and the code-response bounds of Section 9. The activity contributions use (8.16)–(8.20). All powers of q, factors Q^{-1}, source densities, and position counts are fixed by these formulas before rational evaluation.

Proposition 8.2 (Middle-round bounds for the actual source)**.** *Let $p_4 = (D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4}\delta_{a_0}$, with initial word weight w. For output counts $i = 1, 2$, use (8.5)–(8.9) and the local $J_{4,A}$ candidates for four residual inputs. For three residual inputs and four active words, use (8.11), (8.12), and (8.14). For three residual inputs and three active words, use (8.15) when $w \ge 2$. These fixed-position estimates precede the final i output factors R. Multiply by ν^i, expand the activity basis using (8.19), sum over input positions, and combine orthogonal output positions by squared norms. The resulting bounds control the corresponding actual middle-round contributions. For activity outputs, aggregate (8.16)–(8.18) using (8.20).*

Proof. The squared expansions retain the source density. Four residual inputs use local Weyl–Gram bounds. Three residual inputs also retain the nonzero uniform or zero contribution of the omitted

word. Uniform or zero contractions use the product-response and Fourier moment bounds, with byte-weight preservation established in Section 2.4. Equations (8.19) and (8.20) include the position counts for residual and activity outputs, respectively. □

9 Further Bounds for Three Residual Inputs

This section gives general code-response bounds and the special case of three factors with a common source. Let $i \in \{1, 2\}$ count output residual factors. Apply R on the three residual input coordinates, with $\frac{\delta_V}{\sqrt{q}}$ or δ_0 on the remaining coordinate. The $4 - i$ other output coordinates use the same two vectors. Let k count the output factors $\frac{\delta_V}{\sqrt{q}}$. Set $c = 1$ when the remaining input factor is $\frac{\delta_V}{\sqrt{q}}$ and $c = 0$ when it is δ_0. Then $u = k + c$ is the total number of uniform factors.

9.1 A general Gram bound

Let $f \in \mathbb{R}[V^3]$ be the actual density on the three retained words before applying $R^{\otimes 3}$, with any imposed event retained at its original probability mass. Define

$$s_0 = \|f\|_2 , \qquad s_3 = \left\| R^{\otimes 3} f \right\|_2 .$$

The original graph is an $[8, 4]$ MDS code with $5 - i - u$ zero constraints. Shortening at these coordinates leaves $\ell = 3 + i$ residual coordinates forming an

$$[\ell, d] \text{ MDS code}, \qquad d = 4 - (5 - i - u) = i - 1 + u.$$

For $0 < d < \ell$, vectorize the kernel. The three input R factors become transposed factors, while the i output R factors retain their direction. Lemma 2.4 applies factorwise using the common column and Gram bounds. With $\ell_0 = \min(d, \ell - d)$, the unit code indicator has response norm at most

$$\left[(\tau^2/q)^{\ell_0} \delta^{\ell - 2\ell_0} \right]^{1/2} .$$

The u uniform contractions contribute $q^{-u/2}$, and replacing the unnormalized code indicator by a unit indicator contributes $q^{d/2}$. The Frobenius norm of the kernel is therefore at most

$$q^{(i-1)/2} \left[(\tau^2/q)^{\ell_0} \delta^{3 + i - 2\ell_0} \right]^{1/2} .$$

This also bounds the operator norm. Multiplication by the actual three-word source norm s_0 gives the response bound. For $d \le 0$ or $d \ge \ell$, use the direct operator bound below, except when the zero point or uniform marginal is explicitly annihilated.

For $w \le 3$ initially independent active words, projection onto three intermediate words is injective on the MDS code, so $s_0 \le \rho^{w/2}$. For $w = 4$, each image has at most q preimages, and Cauchy–Schwarz gives $s_0 \le q^{1/2} \rho^2$. Additional nonzero restrictions preserve these bounds. Let $D_{\mathrm{src}}(x)$ be the source density before linear mixing, $\pi_3(x)$ the projection onto the three retained words after mixing, and $b_{\mathrm{om}}(x)$ the omitted word. For a nonzero omitted word, the source density f and its unrestricted marginal g are

$$f(y) = \sum_{\pi_3(x) = y} D_{\mathrm{src}}(x) \mathbf{1}[b_{\mathrm{om}}(x) \neq 0], \qquad g(y) = \sum_{\pi_3(x) = y} D_{\mathrm{src}}(x).$$

The density f retains the event's probability mass without conditional normalization. Since $D_{\mathrm{src}} \ge 0$, we have $0 \le f \le g$ and hence $\|f\|_2 \le \|g\|_2$. For $w \le 3$, the injective bound applies. For $w = 4$, there are at most q preimages per image, giving

$$\|f\|_2^2 \le \|g\|_2^2 \le q \sum_x D_{\mathrm{src}}(x)^2 \le q\rho^4,$$

and again $s_0 \le q^{1/2}\rho^2$.

If the omitted word is zero, one source variable is determined by the other $w-1$ variables. Its point bound p gives $s_0 \le p\rho^{(w-1)/2}$.

For $i=1,k=3,c=1$, we have $d=\ell=4$. The homogeneous code fills the residual coordinates, and the zero sums of the residual factors make this term exactly zero.

A direct bound is $\nu^i s_3$, with $s_3 \le \nu^3 s_0$ in general. If $w=1$, the three residual factors share one source variable. Expanding the square and applying Hölder to the three Gram factors gives the stronger bound

$$s_3^2 \le p^2\delta\tau^2, \qquad s_3 \le \min\{p\sqrt{\delta\tau^2}, \nu^3 s_0\}.$$

The probability factor for the omitted word is expanded separately below.

9.2 Three factors with a common source

For one active source word, write the four words after mixing as $T_r X$, where each T_r is invertible and byte-monomial, and set $\mu = D^{\mathsf{S}}\delta_a$ for $a \in V \setminus \{0\}$. When $d=1$, write the residual code coordinates as $L_r y$, with $y \in V$ and additive bijections $L_r : V \to V$. The three input factors give

$$v(y) = \sum_x \mu(x) \prod_{r=1}^{3} R(L_r y, T_r x).$$

Let $\|v\|_2 \le J_{3,R}$ be the byte-weight response bound. The final i factors R act on an $[i,1]$ code embedding with norm at most $\nu\delta^{(i-1)/2}$. Retain the operator norm ν^2 of one R^*R factor; the remaining Schur factors contribute only the diagonal bound δ^{i-1} [5, Lemma 2]. The resulting bound is

$$q^{-u/2}\nu\delta^{(i-1)/2}J_{3,R}.$$

For $u=3$, Fourier duality on all coordinates gives dual dimension $\ell-d=1$. The three source factors become $\mathcal{F}R$, while the probability vector μ is unchanged. The number of uniform contractions becomes $5-i-u$, giving

$$q^{-(5-i-u)/2}\nu\delta^{(i-1)/2}J_{3,\mathcal{F}R}.$$

The final output factors may be Fourier conjugates, whose ν,δ bounds are included in the assumptions. The source matrix is $\mathcal{F}R$, rather than $\mathcal{F}R\mathcal{F}$, because the source vector is not transformed.

The byte-weight bounds apply because the maps L_r, T_r are byte-monomial, including after transposition, inversion, and dual shortening, by Section 2.4.

If the omitted word is nonzero, its probability vector is $\frac{\delta_{V\setminus\{0\}}}{Q} = (\sqrt{q}/Q)\frac{\delta_V}{\sqrt{q}} - Q^{-1}\delta_0$. Multiply the candidate with $c=1$ by $\sqrt{q}/Q$ and the candidate with $c=0$ by $1/Q$, then add by the triangle inequality. For an omitted zero word, only $c=0$ occurs, with coefficient one. Expand the output $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$ factors, multiply by $\sqrt{n_i(a)}$, and sum over the four omitted positions. The local point, squared-norm, and Gram bounds precede these position counts.

Proposition 9.1 (Candidates for three residual inputs)**.** *For a component of the actual four-round source with three residual factors, define the shortened dimension d from i,k,c as above. If $0<d<\ell$, the Gram bound applies. For a one-dimensional code or dual code, the common-source product bound also applies. When the omitted word is nonzero uniform, combine the two terms with absolute coefficients Q^{-1} and $\sqrt{q}/Q$. For an omitted zero word, retain only the zero term. The full-space case $i=1,k=3,c=1$ vanishes. In the other cases not covered by the Gram formula, use the direct operator bound. These candidates may be compared with those of Proposition 8.2 for the same contribution.*

Proof. The general bound follows by vectorizing the MDS code response. The one-dimensional cases use the three-factor source response and the output embedding bound. Byte-weight preservation makes $J_{3,R}$ and $J_{3,\mathcal{F}R}$ applicable. Expand $\frac{\delta_{V\setminus\{0\}}}{Q} = (\sqrt{q}/Q)\frac{\delta_V}{\sqrt{q}} - Q^{-1}\delta_0$ and apply the triangle inequality. □

10 Joint Bounds for the Last Four Rounds

We bound the intermediate terms with three or four residual factors across two linear layers. The nonzero unit indicator expands as

$$\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}} = \alpha\frac{\delta_V}{\sqrt{q}} - \beta\delta_0, \qquad \alpha = \sqrt{q/Q}, \quad \beta = Q^{-1/2}.$$

Assume that $R, R^*, \mathcal{F}R\mathcal{F}, \mathcal{F}R^*\mathcal{F}$ have operator norm at most ν, squared column norms at most δ, squared Frobenius norm at most h, and second Gram moment at most τ^2. The one-sided transforms $\mathcal{F}R, R\mathcal{F}$ also have squared row and column norms at most δ. For example, $\|\mathcal{F}R\delta_x\|_2 = \|R\delta_x\|_2$ and $\|R\mathcal{F}\delta_x\|_2 = \|\mathcal{F}R\mathcal{F}\delta_x\|_2$; apply the adjoint for rows. All four matrices $R, \mathcal{F}R, R\mathcal{F}, \mathcal{F}R\mathcal{F}$ have the same bounds h, τ^2.

10.1 Codes imposed by the linear layer and Fourier transforms

Use the additive MDS map from Section 2.4, $L = \mathrm{SR} \circ \mathrm{MC} \circ \mathrm{SR}$. Its graph code $\{(x, Lx) : x \in V^4\}$ is an additive $[8, 4]$ MDS code; scalar linearity of L over $\mathbb{F}_{2^{32}}$ is not assumed. The image of d free endpoint coordinates on the four intermediate coordinates is an $[4, d]$ MDS code. An additional zero constraint shortens it to dimension $d-1$. The local coordinate bijections preserve the full row, column, and moment sums.

An endpoint with i residual coordinates and k factors $\frac{\delta_V}{\sqrt{q}}$ has code dimension $i + k$. Applying Fourier transforms to its four coordinates exchanges $\frac{\delta_V}{\sqrt{q}}$ and δ_0 and preserves the residual space, since $\mathcal{F}P\mathcal{F} = P$. The alternative dimension is $4 - k$. Choosing the smaller gives

$$d = i + \min\{k, 4 - i - k\}.$$

If the dimensions agree, evaluate both Fourier choices and use the smaller bound. The remaining $d - i$ endpoint parameters are contracted against $\left(\frac{\delta_V}{\sqrt{q}}\right)^{\otimes(d-i)}$, a unit vector. Retaining them before contraction gives an operator-norm bound without an extra power of q. Omitting the endpoint residual projections can only increase this bound. The special $1 \leftrightarrow 2$ estimate below instead keeps the explicit factor $q^{-1/2}$ from this contraction.

Subscripts l and r refer to the left output endpoint and right input endpoint. For a fixed expansion term, k_l and k_r count their factors $\frac{\delta_V}{\sqrt{q}}$ before the Fourier choices, and d_l and d_r are the resulting code dimensions. Thus i residual coordinates belong to the left endpoint and j to the right.

10.2 Four intermediate residual factors

Fix the endpoint positions and Fourier choices, giving $A \in \{R, \mathcal{F}R, R\mathcal{F}, \mathcal{F}R\mathcal{F}\}$. Let $\mathcal{C}_l, \mathcal{C}_r \subseteq V^4$ be the resulting intermediate codes, with bijective parameterizations $c_l : V^{d_l} \to \mathcal{C}_l$ and $c_r : V^{d_r} \to \mathcal{C}_r$. Before the uniform endpoint contractions, the kernel is

$$\mathcal{M}_{d_l,d_r} : \mathbb{C}^{V^{d_r}} \longrightarrow \mathbb{C}^{V^{d_l}},$$

$$\mathcal{M}_{d_l,d_r}(y, x) = \prod_{r=1}^{4} A((c_l(y))_r, (c_r(x))_r).$$

This matrix has q^{d_l} output rows and q^{d_r} input columns, with no normalization factor. Take $m_A \geq \max_{x,y} |A(y,x)|$. For a fixed opposite endpoint, the absolute sum of this kernel over a $[4,d]$ MDS code is bounded by

$$S_0(A) = 0, \qquad S_1(A) = m_A^2\delta, \qquad S_d(A) = q^{d-2}\delta^2 \quad (2 \leq d \leq 4).$$

For $d = 0$, the zero row or column of A annihilates the code. For $d = 1$, bound two factors pointwise and apply Cauchy–Schwarz and the row or column squared-norm bound to the other two. For $d \geq 2$, MDS Hölder has exponent $4/d \leq 2$. Each row or column of length q satisfies $\|v\|_{4/d} \leq q^{d/4-1/2}\|v\|_2$. Multiplying the four bounds gives the claim. The Schur test yields

$$\|\mathcal{M}_{d_l,d_r}\|_2 \leq \sqrt{S_{d_l}(A)S_{d_r}(A)}.$$

The final $(D^{\mathsf{S}})^{\otimes 4}$ applies R on the i residual output coordinates, contributing ν^i; the input is an arbitrary vector in W_j.

For fixed input and output activity patterns, expand all $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$ factors and take absolute coefficients,

$$\binom{a-i}{k_l}\binom{b-j}{k_r}\alpha^{k_l+k_r}\beta^{a-i+b-j-k_l-k_r}.$$

For (i,a), there are $n_i(a) = \binom{4}{i}\binom{4-i}{a-i}$ orthogonal choices of residual positions and exact activity pattern. If every fixed-position block has norm at most g, the aggregated block has norm at most $\sqrt{n_i(a)n_j(b)}\,g$.

10.3 An improvement for two endpoints with one residual factor

Suppose each endpoint retains one residual variable and one endpoint also contracts a factor $\frac{\delta_V}{\sqrt{q}}$. For $x, y \in V$, the kernel has the form

$$M(y,x) = q^{-1/2}\sum_{z\in V}\prod_{r=1}^{4} A(L_r(y,z), T_r x).$$

Here $\{L_r(y,z)\}_r$ parametrizes an $[4,2]$ MDS code, and T_r and $z \mapsto L_r(0,z)$ are additive bijections. For fixed y, the variables (x,z) map bijectively to the two indices of each local matrix. Four-factor Hölder gives

$$\sum_x |M(y,x)| \leq q^{-1/2}\sum_{u,v}|A(u,v)|^4 \leq q^{-1/2}m_A^2 h.$$

For fixed x, MDS Hölder on the two-dimensional code gives $\sum_y |M(y,x)| \leq q^{-1/2}\delta^2$. Hence

$$\|M\|_2 \leq q^{-1/2}m_A\delta\sqrt{h}.$$

The same bound holds with the endpoints exchanged. If both endpoint codes have dimension one, use $m_A^2\delta$; if both have dimension two, use δ^2. Each Fourier direction uses its own point bound, including the distinct cases $\mathcal{F}R$ and $R\mathcal{F}$.

10.4 Three intermediate residual factors

Fix the intermediate nonresidual position. Its operator is the tensor product of three R factors and one complete P factor. This includes the sum of the components with three and four active words. Set

$$T_0(A) = 0, \quad T_1(A) = m_A\delta, \quad T_d(A) = q^{d-3/2}\delta^{3/2} \quad (2 \leq d \leq 4).$$

For a one-dimensional image on the three residual coordinates, bound one factor pointwise and use Cauchy–Schwarz on the other two. For dimensions two and three, apply three-factor MDS Hölder. In dimension four, the three coordinates are unrestricted with q preimages per image, giving the same formula.

If $d_l = 0$ or $d_r = 0$, the corresponding homogeneous endpoint code contains only the zero point, which is annihilated by the residual factors. The response is then zero. The following formulas apply to the remaining case $d_l, d_r \geq 1$.

If the endpoint Fourier choices agree, P is unchanged. Its zero term reduces both code dimensions by one, and its nonzero uniform term contributes Q^{-1}. Relaxing the nonzero restrictions and applying the Schur test gives

$$\sqrt{T_{d_l-1}T_{d_r-1}} + Q^{-1}\sqrt{T_{d_l}T_{d_r}}.$$

If exactly one endpoint is transformed, direct calculation gives

$$|(\mathcal{F}P)(u,v)| = |(P\mathcal{F})(u,v)| = \begin{cases} q^{-1/2}, & u = 0 \text{ or } v = 0, \\ q^{-1/2}Q^{-1}, & u \neq 0,\ v \neq 0. \end{cases}$$

Split the zero-row-or-column condition into two strips. Counting their intersection twice only increases the bound. Including an additional zero–zero term gives the upper bound

$$q^{-1/2}\left[\sqrt{T_{d_l-1}T_{d_r-1}} + \sqrt{T_{d_l-1}T_{d_r}} + \sqrt{T_{d_l}T_{d_r-1}} + Q^{-1}\sqrt{T_{d_l}T_{d_r}}\right].$$

Multiply by ν^i, expand $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$, and include the endpoint counts. Sum over the four positions of the intermediate P by the triangle inequality. These estimates also allow $i = 0$ or $j = 0$; a zero-dimensional endpoint response vanishes. Every intermediate term containing a residual factor annihilates the global ideal matrix. The all-activity intermediate term retains its exact centering.

Proposition 10.1 (Joint comparison for the last four rounds)**.** *Let $E_{i,a}$ be the fourteen orthogonal projections, and let B, C bound the centered block norms of $(D^{\mathsf{S}})^{\otimes 4}D^L(D^{\mathsf{S}})^{\otimes 4} - U$ and $D^L - U$, respectively. For input $s = (j,b)$ and output $r = (i,a)$, let $J^{(4)}_{rs}$ bound the term with four intermediate residual factors and $J^{(3)}_{rs}$ the terms with three residual factors and one complete P. Define them using the four-factor bounds $S_d(A)$ and three-factor bounds $T_d(A)$ above. Minimize over the independent endpoint Fourier choices, multiply by ν^i, expand $\frac{\delta_{V\setminus\{0\}}}{\sqrt{Q}}$, and include the endpoint count $\sqrt{n_i(a)n_j(b)}$. For three residual factors, sum over the four positions of P by the triangle inequality. Then*

$$\begin{aligned} &\|E_r((D^{\mathsf{S}})^{\otimes 4}D^L(D^{\mathsf{S}})^{\otimes 4}D^L - U)E_s\|_2 \\ &\quad\leq \sum_{\substack{k\in\mathcal{S} \\ \deg(k)\leq 2}} B_{rk}C_{ks} + \min\left\{\sum_{k\in\{(3,3),(3,4)\}} B_{rk}C_{ks},\ J^{(3)}_{rs}\right\} + \min\left\{B_{r,(4,4)}C_{(4,4),s},\ J^{(4)}_{rs}\right\}. \end{aligned} \tag{10.1}$$

Equation (10.1) also applies to activity endpoints $i = 0$ or $j = 0$. Its right-hand side may be compared with $(BC)_{rs}$ and any other established bound for the same block.

Proof. Insert the decomposition by intermediate residual count between the two linear layers. The component W_4 contains $R^{\otimes 4}$ and is bounded by $J^{(4)}$. The component W_3 is the sum over the four positions of $R^{\otimes 3} \otimes P$, covering both activity sizes three and four. Its complete sum is bounded by $J^{(3)}$. Each contribution also has the ordinary block-product bound, so use the smaller of the two. Retain the ordinary block products for the other components. Only the component with no residual factor needs global centering. The triangle inequality gives (10.1), with $J^{(3)}, J^{(4)}$ established by the preceding Schur and Fourier estimates. □

11 Global Composition and Proof of the Main Theorem

We combine the prefix, common-source, and joint-tail bounds in a nonnegative recurrence on the fourteen component norms.

For an input difference $a \in \Omega$ in the original coordinates, define

$$p_{a,10} = D_{\Omega}^{E_{10}} \delta_a = D^T \widetilde{D}_{\Omega}^{E_{10}} \delta_{Ta}, \qquad w(a) = \mathrm{wt}(Ta).$$

Here wt counts nonzero words. The vector $p_{a,10}$ is the ten-round output-difference distribution in the original coordinates. The recurrence uses the grouped input $a_0 = Ta$. For each initial word weight w, let

$$\Delta_w = \max_{\substack{a \in \Omega \\ w(a)=w}} \|p_{a,10} - \frac{\delta_\Omega}{n}\|_1.$$

Write $v^{(r,w)}$ for the component bounds at initial weight w, and suppress the second superscript when w is fixed.

11.1 The middle two rounds

Use $v^{(4)}$ for the first four rounds and the block-norm bounds on the centered form of g, where $g = D^L(D^{\mathsf{S}})^{\otimes 4}\delta_{a_0}$. Set $\mathsf{D}_{(i,a),(i,a)} = \nu^i$. Since $Uf_4 = 0$, the component $r = (i,a)$ of $f_6 = (D^{\mathsf{S}})^{\otimes 4} D^L f_4$ satisfies

$$E_r f_6 = \sum_s E_r((D^{\mathsf{S}})^{\otimes 4} D^L - U) E_s f_4.$$

Define the contribution from input block s as $E_r((D^{\mathsf{S}})^{\otimes 4} D^L - U) E_s f_4$. Its norm is bounded by

$$T_{rs} = \min\{(\mathsf{D}C)_{rs} v_s^{(4)},\ B_{rs} g_s\} \tag{11.1}$$

The second candidate is valid because $(D^{\mathsf{S}})^{\otimes 4}$ preserves every subspace, $E_s f_4 = (D^{\mathsf{S}})^{\otimes 4} E_s(g - \frac{\delta_\Omega}{n})$, and $U(D^{\mathsf{S}})^{\otimes 4} = U$. Thus

$$E_r((D^{\mathsf{S}})^{\otimes 4} D^L - U) E_s f_4 = E_r((D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U) E_s (g - \frac{\delta_\Omega}{n}).$$

For $i = 1, 2$, compare the contribution from $s = (4,4)$ with its common-source bound for four residual factors, and the contribution from $s = (3,4)$ with the three-factor bound having a nonzero uniform omitted word. If $w \geq 2$, also compare the contribution from $s = (3,3)$ with its zero-word bound. Sections 8 and 9 establish these candidates, including the final ν^i, the activity expansion, and all output-position counts.

Sum over s to obtain the intermediate vector. Activity rows $r = (0,a)$ admit two further candidates. The complete g has squared norm at most ρ^w. Combining its residual components by Cauchy–Schwarz gives

$$v_{0,a}^{(6)} \leq \sum_{b=1}^{4} C_{(0,a),(0,b)} v_{0,b}^{(4)} + \rho^{w/2} \left(\sum_{s:\deg(s)\geq 1} B_{(0,a),s}^2 \right)^{1/2}. \tag{11.2}$$

For $w = 1$, the common-source activity bound summed over $j = 2, 3, 4$ replaces all residual-source contributions. For $w \geq 2$, the multiple-source activity bound replaces only the contribution from $s = (4,4)$, retaining the others. Take the smallest candidate for each activity row and apply the total-mass constraint to the component with four active words.

11.2 Two bounds for the last four rounds

Let $\mathsf{cap}(v)$ replace only $v_{0,4}$ by its minimum with $\sum_{a=1}^{3}\sqrt{\binom{4}{a}Q^{a-4}}v_{0,a}$. The first computation applies the total-mass constraint after the next linear layer:

$$w^{(8-)} = \mathsf{cap}(Cv^{(6)}), \qquad v^{(10),A} = Bw^{(8-)}.$$

The superscript $8-$ denotes the point before the next superbox layer; the linear layer adds no substitution round. We write $w^{(8-,w)}$ when displaying the initial weight. For the second computation, let J be the nonnegative 14×14 comparison matrix for the complete operator $(D^{\mathsf{S}})^{\otimes4}D^{L}(D^{\mathsf{S}})^{\otimes4}D^{L}-U$. Start from BC, also allowing $\mathsf{D}CB$ in activity-input columns. The bounds $J^{(3)}$ and $J^{(4)}$ replace only the combined W_3 terms and the W_4 term, respectively. All other terms are retained. Set

$$v^{(10),B} = Jv^{(6)}, \qquad v^{(10)} = \mathsf{cap}\big(\min(v^{(10),A}, v^{(10),B})\big),$$

with coordinatewise minima. Lemma 3.2 gives the one-norm bound

$$\Delta_w \le \mathcal{L}(v^{(10)}) := 2\sum_{a=1}^{3}\sqrt{\binom{4}{a}Q^{a}}\,v_{0,a}^{(10)} + \sum_{a=1}^{4}\sqrt{\binom{4}{a}Q^{a}}\sqrt{\sum_{i=1}^{a}(v_{i,a}^{(10)})^2}. \tag{11.3}$$

The fixed output permutation D^T preserves the one norm.

11.3 Final one-norm bounds

Evaluate the block inequalities separately for $w \in \{1,2,3,4\}$ using the rational upper approximations of Appendix B. The resulting exact rational bound $\widehat{L}_w$ includes upward rounding of the square roots in (11.3), so $\Delta_w \le \mathcal{L}(v^{(10,w)}) \le \widehat{L}_w$. Set $\widehat{T}_w = \widehat{L}_w/2$. Column 2 lists a simplified ceiling for $\widehat{L}_w$; the last two columns give $\log_2 \widehat{L}_w$ and $\log_2 \widehat{T}_w$ for the exact values.

Initial weight w	Simplified ceiling for $\widehat{L}_w$	$\log_2 \widehat{L}_w$	$\log_2 \widehat{T}_w$
1	$3017/2^{140}$	-128.441513596899	-129.441513596899
2	$3187/2^{144}$	-132.362071143925	-133.362071143925
3	$2356/2^{152}$	-140.798155808897	-141.798155808897
4	$2864/2^{153}$	-141.516287972292	-142.516287972292

Exact fraction comparisons for all four initial word weights give

$$\max_{a\neq0}\|p_{a,10} - \frac{\delta_\Omega}{n}\|_1 \le \max_{1\le w\le4}\widehat{L}_w \le \frac{3017}{2^{140}} < 2^{-128}, \qquad \varepsilon_{10} = \frac{1}{2}\max_{a\neq0}\|p_{a,10} - \frac{\delta_\Omega}{n}\|_1 < 2^{-129}.$$

The statistical-distance identity follows from independent input and output whitening.

Proof of Theorem 3.1. Independent round keys allow the averaged difference channels to compose. Apply the grouping (2.1), the block multiplication of Lemma 2.1, Propositions 6.1 and 7.1, the common-source and joint four-round bounds, and the conversion of Lemma 3.2. The rational recurrence gives, for every $a \neq 0$,

$$\|(D_\Omega^{E_{10}} - U)\delta_a\|_1 \le \begin{cases} 3017/2^{140}, & w(a) = 1, \\ 3187/2^{144}, & w(a) = 2, \\ 2356/2^{152}, & w(a) = 3, \\ 2864/2^{153}, & w(a) = 4. \end{cases}$$

The four cases cover every nonzero difference, and each right-hand side is a rational upper bound. Since $\|A\|_1 = \max_a \|A\delta_a\|_1$ for any finite matrix A, this proves (3.1). Output whitening gives [4, Lemma 4]

$$\Pr[E_{10}(x) = y, E_{10}(x+a) = y+b] = N^{-1} D_\Omega^{E_{10}}(b,a).$$

The absolute deviation sum for an output-pair column is therefore $\sum_{y,b} N^{-1} |D_\Omega^{E_{10}}(b,a) - 1/n| = \|(D_\Omega^{E_{10}} - U)\delta_a\|_1$. Taking the maximum column proves equality of the two induced norms. Statistical distance is half the one-norm distance. □

12 Discussion

12.1 The one-norm bound and local probability estimates

The final quantity is the induced one norm of the centered, key-averaged difference-distribution matrix,

$$\left\| D_\Omega^{E_{10}} - U \right\|_1 = \max_{a \neq 0} \sum_{b \neq 0} \left| D_\Omega^{E_{10}}(b,a) - \frac{1}{2^{128}-1} \right|.$$

For every fixed nonzero input difference, this bound controls all output events. The sum runs over the entire output column, so it bounds the maximum differential probability and the collision probability of that column.

Euclidean norms allow the proof to use orthogonality, Gram moments, and tensor products. Conversion to the one norm must then account for support sizes. Combining all components early and multiplying by the square root of the full state-space size loses the exact activity information. Our recurrence retains all fourteen components until the endpoint, where the activity component has zero total mass and the residual components remain orthogonal.

12.2 Subspaces and common-source constraints

The orthogonal decomposition follows Beyne, Leander, and Schütt [4, Section 6.1], with the exact-support incidence bounds of Ren and Yu [5, Eq. (27)]. We keep both the number of residual factors and the number of active words throughout the composition. Words obtained from the same prefix depend on common source variables, and independently chosen worst-case vectors in their subspaces need not arise from any actual prefix.

Ren and Yu [5, Sections 3.4 and 6] use MDS constraints and Gram estimates to handle common sources. Here we retain the actual source distributions through the middle segment. MDS codes describe the linear relations between the source variables and the output words; refined Gram moments, Fourier representations, and local probability bounds control the sums along these relations.

The joint tail addresses a different loss. Inserting separate norm bounds between two consecutive diffusion layers can overestimate both parts of an expression determined by the same internal variables. We keep the joint expressions for the intermediate terms with three or four residual factors, and add the comparison bounds for the remaining terms. This retains constraints across several layers within a block.

12.3 Scope and open questions

The composition of difference-distribution matrices uses independent and uniform random round keys, so the result averages over the key tuple. The standard AES key schedule derives related round keys from one master key and does not satisfy this assumption.

Ten rounds reach the stated threshold. Whether fewer rounds suffice, and how close these estimates are to optimal, remain open. Further analysis could retain more information about the source values or operator directions, or estimate the final one norm more directly.

13 Conclusion

For ten-round AES with independent and uniform random round keys, we proved

$$\left\| D_{\Omega}^{E_{10}} - U \right\|_1 \leq \frac{3017}{2^{140}} < 2^{-128}.$$

The proof groups the ten substitution rounds exactly into five superbox layers. A fourteen-state orthogonal decomposition, local Gram moments, common-source constraints, and joint tail estimates bound the output deviation for every nonzero input difference. Exact rational calculations establish the final threshold.

Input and output whitening relate this one-norm bound to the joint output distribution of any two distinct plaintexts. The maximum statistical distance from the output pair of a uniform random permutation satisfies

$$\varepsilon_{10} = \frac{\Delta_{10}}{2} \leq \frac{3017}{2^{141}} < 2^{-129}.$$

References

[1] National Institute of Standards and Technology. *Advanced Encryption Standard (AES)*. FIPS PUB 197, 2001; updated edition, 9 May 2023. https://doi.org/10.6028/NIST.FIPS.197-upd1.

[2] Liam Keliher and Jiayuan Sui. Exact maximum expected differential and linear probability for two-round Advanced Encryption Standard. *IET Information Security*, 1(2):53–57, 2007. https://doi.org/10.1049/iet-ifs:20060161. Full version: Cryptology ePrint Archive, Paper 2005/321, 2005. https://eprint.iacr.org/2005/321.

[3] Tim Beyne. A Geometric Approach to Linear Cryptanalysis. In *Advances in Cryptology—ASIACRYPT 2021*, Part I, Lecture Notes in Computer Science, vol. 13090, pp. 36–66. Springer, 2021. https://doi.org/10.1007/978-3-030-92062-3_2. Full version: Cryptology ePrint Archive, Paper 2021/1247. https://eprint.iacr.org/2021/1247.

[4] Tim Beyne, Gregor Leander, and Immo Schütt. Pairwise Independence of AES-Like Block Ciphers. Cryptology ePrint Archive, Paper 2025/1495, revision of 27 September 2026. https://eprint.iacr.org/2025/1495.

[5] Chongxu Ren and Hongbo Yu. *Gram Moments and Pairwise Mixing of AES*. Cryptology ePrint Archive, Paper 2026/2304, 2026. https://eprint.iacr.org/2026/2304.

[6] Tianren Liu, Stefano Tessaro, and Vinod Vaikuntanathan. The t-wise Independence of Substitution-Permutation Networks. In *Advances in Cryptology—CRYPTO 2021*, Part IV, Lecture Notes in Computer Science, vol. 12828, pp. 454–483. Springer, 2021. https://doi.org/10.1007/978-3-030-84259-8_16. Full version: Cryptology ePrint Archive, Paper 2021/507. https://eprint.iacr.org/2021/507.

[7] Itai Dinur. New Techniques for Analyzing Differentials with Application to AES. In *Advances in Cryptology—EUROCRYPT 2026*, Part VI, Lecture Notes in Computer Science, vol. 16546, pp. 211–242. Springer, 2026. https://doi.org/10.1007/978-3-032-25333-0_8. Full version: Cryptology ePrint Archive, Paper 2025/1326. https://eprint.iacr.org/2025/1326.

[8] Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES—The Advanced Encryption Standard*. Information Security and Cryptography. Springer, 2002. https://doi.org/10.1007/978-3-662-04722-4.

[9] Joan Daemen and Vincent Rijmen. Understanding Two-Round Differentials in AES. In *Security and Cryptography for Networks—SCN 2006*, Lecture Notes in Computer Science, vol. 4116, pp. 78–94. Springer, 2006. https://doi.org/10.1007/11832072_6. Full report: *Two-Round AES Differentials*, Cryptology ePrint Archive, Paper 2006/039. https://eprint.iacr.org/2006/039.

[10] Tim Beyne and Vincent Rijmen. Differential Cryptanalysis in the Fixed-Key Model. In *Advances in Cryptology—CRYPTO 2022*, Part III, Lecture Notes in Computer Science, vol. 13509, pp. 687–716. Springer, 2022. https://doi.org/10.1007/978-3-031-15982-4_23. Full version: Cryptology ePrint Archive, Paper 2022/837. https://eprint.iacr.org/2022/837.

[11] Zhongfeng Niu, Kai Hu, Thomas Peyrin, Haolin Liu, and Meiqin Wang. *Pairwise Independence of Block Ciphers with Near-MDS Diffusion*. To appear.

[12] Florent Chabaud and Serge Vaudenay. Links Between Differential and Linear Cryptanalysis. In *EUROCRYPT '94*, LNCS 950, pp. 356–365. Springer, 1995. https://doi.org/10.1007/BFb0053450.

A The Fourteen-State Algorithm

This algorithm computes one-norm bounds for the four possible initial word weights from rational upper bounds on the local AES quantities. Matrix entries and vector coordinates are nonnegative, and minima are taken coordinatewise.

Inputs. Let $q = 2^{32}$ and $Q = q - 1$. Use $\mathcal{S}$, $\mathcal{A}$, and deg from Section 2.5, with the states ordered as

$$(0,1),\dots,(0,4),\ (1,1),\dots,(1,4),\ (2,2),(2,3),(2,4),\ (3,3),(3,4),(4,4).$$

The inputs are as follows.

- Section 2.8 gives the known parameters ν, δ, p, ρ. Section 4 gives the refined h, t, Fourier entry bounds, and byte-weight product responses. Here t bounds the second Gram moment, and $\tau = \sqrt{t}$.
- Use the matrix C from Section 2.6 for $D^L - U$, and the refined matrix B from Section 5 for $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} - U$. Set $\mathsf{D}_{ss} = \nu^{\deg(s)}$.
- Proposition 7.1 gives $v^{(4,w)}, g^{(w)}$, bounding the state norms of $(D^{\mathsf{S}})^{\otimes 4} D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0} - \frac{\delta_\Omega}{n}$ and $D^L (D^{\mathsf{S}})^{\otimes 4} \delta_{a_0} - \frac{\delta_\Omega}{n}$, respectively, where a_0 is a grouped input difference of word weight w.
- For $r = (i, a)$ with $i = 1, 2$, let $S_r^{(4,w)}$ be the bound for the middle-round contribution from $(4,4)$. Form it from (8.5)–(8.9) and the local bounds $J_{4,A}$, including the final factor ν^i and the activity expansion (8.19). Proposition 8.2 specifies the output-position counts.
- Let $S_r^{(3,w,4)}$ and $S_r^{(3,w,3)}$ bound the contributions from $(3,4)$ and $(3,3)$, respectively. For the former, use the applicable candidates in (8.11), (8.12), and (8.14); for the latter, use (8.15) when $w \geq 2$. Compare each with Proposition 9.1, including the final residual factors, the activity expansion (8.19), and all input- and output-position counts. The contribution from $(3,3)$ vanishes for $w = 1$.
- Let $A_a^{(1)}$ bound the activity contribution from all residual inputs for one source. Obtain it by aggregating (8.16) and (8.17) through (8.20), multiplying the contribution with j residual input words by $\binom{4}{j}$ and summing over $j = 2, 3, 4$. For $w \geq 2$, let $A_a^{(w,4)}$ be the activity contribution bound obtained from (8.18) through (8.20); it includes only input state $(4,4)$.
- Section 10 gives $J^{(3)}, J^{(4)}$, bounding the joint tail terms with three factors R and one full P, and with four factors R, respectively. When each endpoint has one residual factor, $J^{(4)}$ also takes the minimum with the special $1 \leftrightarrow 2$ bound.

The operation $\mathsf{cap}(v)$ changes only one coordinate:

$$[\mathsf{cap}(v)]_{0,4} = \min\left\{ v_{0,4}, \sum_{a=1}^{3} \sqrt{\binom{4}{a}}\, Q^{(a-4)/2} v_{0,a} \right\}.$$

All other coordinates remain unchanged. This operation applies to norm bounds for an actual centered vector, using the zero-total-mass identity in Lemma 3.2.

1. **Precompute the tail.** For $r, s \in \mathcal{S}$, define

$$G_{rs} = \sum_{\substack{k \in \mathcal{S} \\ \deg(k) \leq 2}} B_{rk} C_{ks} + \min\Big\{ B_{r,(3,3)} C_{(3,3),s} + B_{r,(3,4)} C_{(3,4),s},\, J_{rs}^{(3)} \Big\}$$
$$+ \min\{ B_{r,(4,4)} C_{(4,4),s},\, J_{rs}^{(4)} \}, \qquad J_{rs} = \min\{ (BC)_{rs}, G_{rs} \}.$$

   Only for $s \in \mathcal{A}$, further set $J_{rs} \leftarrow \min\{J_{rs}, (\mathsf{D}CB)_{rs}\}$, since $(D^{\mathsf{S}})^{\otimes 4} E_s = E_s$ holds only on the activity space. The bound $J^{(3)}$ replaces the *sum* of the $(3,3)$ and $(3,4)$ terms.

2. **Initialize the middle segment for each $w = 1, 2, 3, 4$.** Take the prefix bounds $v^{(4,w)}$ and $g^{(w)}$, and set
$$T_{rs} = \min\{(\mathsf{D}C)_{rs} v_s^{(4,w)},\ B_{rs} g_s^{(w)}\}.$$
For $r = (i, a)$ with $i = 1, 2$, replace $T_{r,(4,4)}$ by its minimum with $S_r^{(4,w)}$, and $T_{r,(3,4)}$ by its minimum with $S_r^{(3,w,4)}$. When $w \geq 2$, also replace $T_{r,(3,3)}$ by its minimum with $S_r^{(3,w,3)}$. For one source, the $(3, 3)$ component is zero.

3. **Sum the contributions and refine the activity rows.** First set $m_r = \sum_s T_{rs}$. For $r = (0, a)$, define
$$L_r = \sum_{s \in \mathcal{A}} C_{rs} v_s^{(4,w)}, \qquad Z_r = L_r + \rho^{w/2} \sqrt{\sum_{s \notin \mathcal{A}} B_{rs}^2}.$$
Set $m_r \leftarrow \min\{m_r, Z_r\}$. For $w = 1$, also take the minimum with $L_r + A_a^{(1)}$; for $w \geq 2$, take the minimum with $\sum_{s \neq (4,4)} T_{rs} + A_a^{(w,4)}$ instead. Finally, set $v^{(6,w)} = \mathsf{cap}(m)$.

4. **Bound the ten-round vector in two ways.** Set
$$w^{(8-,w)} = \mathsf{cap}(C v^{(6,w)}), \qquad v^{(10,w)} = \mathsf{cap}\big(\min\{B w^{(8-,w)}, J v^{(6,w)}\}\big).$$
The two candidates use the intermediate zero-total-mass constraint and the joint tail estimate, respectively.

5. **Compute the one-norm bound.** Evaluate $\mathcal{L}(v^{(10,w)})$ from (11.3), replacing each square root by the rational upper approximation of Appendix B. Denote the rational result by $\widehat{L}_w$, so that $\Delta_w \leq \mathcal{L}(v^{(10,w)}) \leq \widehat{L}_w$. Output $\widehat{L}_* = \max_{1 \leq w \leq 4} \widehat{L}_w$ and check by integer cross-multiplication that
$$\widehat{L}_* \leq 3017/2^{140} < 2^{-128}.$$

B Rational Evaluation

We evaluate the finite bounds with exact rational arithmetic and upper approximations to square roots.

Local estimates. Section 2.8 collects the local estimates; Section 4 refines their Gram moments. The maximum expected superbox differential probability is due to Keliher and Sui [2]. The remaining byte-table extrema and moments are obtained by integer enumeration as described in Section 4.

Square roots. For nonnegative rational x, set $\Lambda_{\mathrm{rt}} = 2^{320}$ and compute
$$z = \left\lfloor \sqrt{\lfloor x \Lambda_{\mathrm{rt}}^2 \rfloor} \right\rfloor, \qquad \mathrm{rt}(x) = \begin{cases} z/\Lambda_{\mathrm{rt}}, & (z/\Lambda_{\mathrm{rt}})^2 \geq x, \\ (z+1)/\Lambda_{\mathrm{rt}}, & \text{otherwise.} \end{cases}$$
Integer cross multiplication gives $\mathrm{rt}(x)^2 \geq x$. The local estimates use the same construction with scales 10^{24}, 2^{150}, or 10^{35}. These bounds enter the global recurrence as rational numbers.

Propagation of upper bounds. All radicands are nonnegative. Centered activity coefficients retain $Q = q - 1$ exactly, with subtraction performed before taking absolute values. Sums, products, and coordinatewise minima of nonnegative upper bounds preserve the required inequalities; each square root is replaced by its rational upper approximation. This includes the support-size factors and residual norms in (11.3). Its rational evaluation gives $\widehat{L}_w \geq \mathcal{L}(v^{(10,w)}) \geq \Delta_w$ for each initial word weight. The displayed logarithms are decimal approximations to $\log_2 \widehat{L}_w$ and $\log_2(\widehat{L}_w/2)$.