Gram Moments and Pairwise Mixing of AES

Chongxu Ren and Hongbo Yu

Department of Computer Science and Technology,
Tsinghua University, Beijing, China
rcx23@mails.tsinghua.edu.cn, yuhongbo@tsinghua.edu.cn

Abstract. Almost pairwise independence asks how closely a block cipher reproduces the joint output distribution of a random permutation on two distinct plaintexts. Existing analyses of AES with independent round keys combine bounds on the distribution after the first few rounds with contraction over later rounds. Diffusion makes local output differences dependent, and estimating local responses or successive transitions separately can lose useful information. We preserve these dependencies in two ways. First, Gram-moment estimates sum the squared overlaps between local responses whose inputs are linked by diffusion. Second, we estimate two consecutive transitions together. The compatible intermediate differences form affine fibers; summing over these fibers inside the composed operator retains the constraints shared by the two diffusion layers.

We prove $\varepsilon_{8+2k} < 2^{-92.19-41k}$ for every integer $k \geq 0$, where ε_r is the largest total variation distance, over distinct plaintext pairs, between the ciphertext-pair law and the corresponding random-permutation law. The round keys are independent and uniform, sampled once and reused across evaluations. Consequently, 10, 14, and 16 rounds suffice for statistical accuracy 2^{-128}, 2^{-192}, and 2^{-256}. The same bound controls the complete view of an algorithm making at most two adaptive classical forward queries with an independent finite private seed. The resulting family gives 41 additional bits of pairwise accuracy for every two additional rounds, starting at eight rounds.

Keywords: AES · Pairwise independence · Gram moments · MDS codes · Spectral analysis

Table of Contents

1 Introduction . . . 3
1.1 Our results . . . 5
1.2 How the proof preserves dependence . . . 6
1.3 Related work . . . 7
2 Preliminaries . . . 9
2.1 AES with independent round keys . . . 9
2.2 Difference transitions and pairwise distance . . . 10
2.3 Superboxes and MDS diffusion . . . 11
2.4 Activity, residual variation, and local norms . . . 13
3 Technical Overview . . . 14
4 The Two-Step Mixing Method . . . 18
4.1 Error profiles and total variation . . . 18
4.2 Two-step decomposition and contraction . . . 19
5 Gram Moments and MDS Fibers . . . 21
5.1 Responses to MDS codes . . . 22
5.2 Nonuniform common sources . . . 24
5.3 Encoding and restricted sources . . . 25
5.4 Two-step compositions on signed inputs . . . 27
5.5 Entry moments for thin codes . . . 31
6 Byte Estimates and Diffusion Geometry . . . 33
6.1 The byte channel and its Gram data . . . 33
6.2 AES diffusion and the activity block . . . 38
6.3 Activity frames and contracted graph responses . . . 40
6.4 Degree-two coupling . . . 43
6.5 Degree-three estimates . . . 47
6.6 Remaining block estimates . . . 53
7 Word Norms, Moments, and Sources . . . 56
7.1 Required bounds and orientations . . . 56
7.2 Columns, entries, and ordinary sources . . . 57
7.3 The word operator norm . . . 61
7.4 The word Gram moment and total energy . . . 62
7.5 One-sided Fourier moments . . . 64
8 AES Diffusion Graph Bounds . . . 67
8.1 Graph coordinates and dual representations . . . 67
8.2 One-step bounds . . . 69
8.3 Two-step bounds with fixed activity choices . . . 71
8.4 Combining the shared activity projection . . . 74
9 Recurrence Coefficients and Initialization . . . 75
9.1 From fixed positions to profile components . . . 75
9.2 Rational propagation maps . . . 76
9.3 Initial profile and two-step source estimate . . . 81

9.4 The four AES profile estimates 87
10 From the Recurrence to Pairwise Mixing 88
10.1 A support-sensitive distance bound 88
10.2 The normalized error recurrence 89
10.3 Two invariant profiles 90
10.4 The final distance bound 92
11 Conclusions 93
References 93

1 Introduction

Differential and linear cryptanalysis have shaped the design of modern block ciphers since the early 1990s [6,43]. AES, standardized in 2001 [47], was designed to provide strong protection against these methods using only 10, 12, or 14 rounds. Its wide-trail strategy combines the nonlinear byte substitution with linear diffusion so that a nonzero trail through several rounds encounters many active S-boxes [19]. This gives a mathematical basis for important design arguments. Establishing quantitative guarantees for the joint ciphertext distributions of AES, at similarly small round counts, has proved substantially harder. The widespread use of AES makes this an important problem for the theory of symmetric cryptography: we seek to understand what its concrete round function can provably guarantee.

From trails to complete differentials. The first difficulty is the passage from an individual trail to a complete differential. A trail specifies the difference after every round, whereas a differential specifies only the input and output differences. Its probability is the sum over all connecting trails. Wide-trail arguments control the individual terms, but many small terms can contribute to the same sum. Consequently, a good bound on the probability of each trail does not by itself give a comparably good bound on the probability of each differential. Nyberg and Knudsen initiated the study of rigorous differential bounds that account for this sum [49].

As in that line of work, we analyze AES with independent uniform round keys. The entire tuple is sampled once and reused for all evaluations; both plaintexts are therefore encrypted under the same keys. This idealization makes difference propagation a Markov process [39] and permits exact averaging over the keys, while retaining AES's actual substitution and diffusion. It addresses the contribution of the round function to security. Dependencies introduced by the standard key schedule, and the relation between average probabilities and those for a fixed key, require separate analysis [38,5]. Our round convention includes initial whitening and omits the final MixColumns operation; Section 2 specifies the experiment.

The years surrounding AES standardization saw a succession of results on differential probabilities and linear hulls for substitution–permutation networks (SPNs). Hong et al. and Keliher, Meijer, and Tavares developed bounds from the

substitution and diffusion layers [27,36]; Park et al. and subsequent work refined the analysis for the Rijndael round structure [50,51,34]. Keliher and Sui eventually determined the exact maximum expected differential probability $53/2^{34}$ for two AES rounds and obtained the four-round upper bound $(53/2^{34})^4$ [37]. Further work explained how the interaction between substitution and diffusion determines these probabilities [20,18]. Section 1.3 discusses this history in more detail.

These results establish guarantees about the most likely differential. They do not, by themselves, quantify how close the entire difference distribution is to uniform. For a random permutation, every nonzero output difference has the same probability at a fixed nonzero input difference. An upper bound substantially above that ideal probability can still allow many outputs to be much less likely than others. To understand convergence toward random-permutation behavior, one must control the distribution as a whole and show how this control improves when more rounds are added.

From differential bounds to pairwise independence. Almost pairwise independence formulates this distributional goal. For any two distinct fixed plaintexts, the joint ciphertext distribution over the sampled keys should be close in statistical distance to the uniform distribution on ordered pairs of distinct blocks. The distance bounds the probability difference for every event on those two outputs. It therefore gives a guarantee for the complete ciphertext pair, including all its output-difference events. This is the type of statistical control studied in decorrelation theory [54].

The broader theory of almost independent permutations and their construction by local permutations was developed by Gowers, Hoory et al., and Brodsky and Hoory [26,28,7]. For AES, the challenge is to obtain strong quantitative bounds while keeping its prescribed S-box and linear layer. Replacing these components by independent random permutations or changing the diffusion can make a proof possible, but does not establish the same guarantee for the concrete AES round function. This leads to the question studied here:

> With independent round keys, how many rounds of AES suffice to make every ciphertext-pair distribution close to that of a random permutation?

Closest quantitative bounds. Liu, Tessaro, and Vaikuntanathan (LTV) proved almost pairwise independence for SPNs with concrete substitution and diffusion layers, including AES [41]. Their Fourier-analytic extraction argument explains how substitution converts entropy in the input differences into closeness to uniformity. An initial estimate for a few rounds can then be amplified by adding rounds. For AES, their quantitative bound requires roughly nine thousand rounds to reach pairwise accuracy 2^{-128}. This established that a concrete AES round function can satisfy the distributional target, but left a large gap between provable round counts and the 10, 12, or 14 rounds of the standardized cipher.

Dinur reduced the sufficient threshold to forty rounds [21]. His analysis studies the full difference distribution through its collision probability, the sum of its squared probabilities. This quantity measures how concentrated a distribution

remains. Estimates for local substitutions and for diffusion give strong smoothing from a fixed input difference in the first few rounds. Combined with contraction in later rounds, they yield a pairwise bound. The work also obtains pointwise estimates for individual complete differentials; the pointwise and pairwise conclusions measure different aspects of convergence.

Independently and concurrently, Beyne, Leander, and Schütt (BLS) reduced the threshold to twenty rounds [4]. They bound an operator measuring how much a block of rounds can preserve a deviation from the ideal difference law. Their decomposition separates the probability of a zero/nonzero pattern, called its activity, from variation among nonzero values within that pattern. Substitution contracts the latter variation, but diffusion couples it back to activity. MDS-code expansions and local singular-value estimates control this interaction. The superbox representation groups two substitution layers into a channel on a word and permits the method to be applied to AES.

The two approaches supply complementary inputs: Dinur's estimate is particularly effective for the distribution produced by a fixed input, whereas BLS's estimate controls arbitrary signed errors during propagation. Both papers discuss this complementarity [21,4]. The D+B row in Table 1 combines Dinur's initialization with the numerical contraction refinement of the corrected BLS full version. The remaining task is to retain enough information through diffusion and repeated contraction to make these bounds effective at still smaller round counts. Our proof keeps the distinction between fixed-input initialization and signed-error propagation. It sharpens both stages by retaining overlaps between local responses and the relation between an intermediate error and the transition that produced it.

1.1 Our results

Let $E_{r,\mathbf{K}}$ be AES with r substitution layers, initial whitening, and omitted final MixColumns. The keys $\mathbf{K} = (K_0, \ldots, K_r)$ are independent uniform 128-bit strings, sampled once and retained across evaluations. Let Perm be a uniformly sampled permutation of the $N = 2^{128}$ states. We measure

$$\varepsilon_r = \max_{x \neq x'} \mathrm{TV}(\mathcal{L}(E_{r,\mathbf{K}}(x), E_{r,\mathbf{K}}(x')), \mathcal{L}(\mathsf{Perm}(x), \mathsf{Perm}(x'))) \,. \tag{1}$$

The reference law is uniform on ordered pairs of distinct states. Section 2 specifies the experiment and the AES round convention.

Theorem 1 (Mixing from eight rounds). *For every integer $k \geq 0$,*

$$\varepsilon_{8+2k} < 2^{-92.19-41k}. \tag{2}$$

Corollary 1 (The 10/14/16 round counts). *In the same model,*

$$\varepsilon_{10} < 2^{-128}, \qquad \varepsilon_{14} < 2^{-192}, \qquad \varepsilon_{16} < 2^{-256}.$$

Proof. Substituting $k = 1, 3, 4$ in Theorem 1 gives exponents 133.19, 215.19, and 256.19, respectively. □

The family begins at eight rounds because the eight- and ten-round estimates initialize the two alternating bounds used in the proof. The recurrence preserves these bounds, represented by integer vectors whose normalized outputs are below 112. The available six-round upper estimate does not satisfy the same output bound; this is a limitation of the estimate, rather than a lower bound on AES's distance. Section 10 gives the complete calculation, including the six-round comparison in Section 10.4.

The bound concerns the full ciphertext pair. It also controls the complete view of any classical algorithm making at most two adaptive forward queries, including repeats and early stopping, with a finite private seed independent of the oracle. Section 2.2 proves this consequence from whitening.

Table 1. Pairwise AES bounds with independent round keys and sufficient rounds for each accuracy. D+B combines Dinur and BLS. Our row is Theorem 1. Hurst's row is an announcement; its full proof and exact model conventions remain unchecked.

Source	Rounds	Bound	2^{-128}	2^{-192}	2^{-256}
Dinur [21]	$4+4k$	$2^{-3.82-14.65k}$	40	56	76
BLS [4]	$4+4k$	$2^{14-39.29k}$	20	28	32
D+B [21,4]	$4+4k$	$2^{-3.82-39.29k}$	20	24	32
Hurst [30]	$4+4k$	$2^{-3.82-48.03k}$	16	20	28
Ours	$8+2k$	$2^{-92.19-41k}$	10	14	16

The propagation gives 41 accuracy bits per two additional rounds. Table 1 keeps each prior bound at its own round index. The first three rows are established bounds; Hurst's estimate is an announcement whose full proof was unavailable.

Formalization. A Lean formalization of the main theorem and its supporting lemmas is available at https://github.com/chongxuren/aes-mixing-lean.

1.2 How the proof preserves dependence

Diffusion couples local responses. Group the sixteen bytes into four words, each containing four bytes. Two substitution layers with the intervening column operation form a word channel, or superbox; Section 2.3 gives the exact regrouping. Diffusion between these word channels can create dependence even when the input words are independent. If only one input word is nonzero, for example, all four output words are functions of the same word. Treating those outputs as independent would discard a constraint that the proof must respect.

Gram moments retain the overlaps. For a local channel, remove the part that records only activity. Its remaining columns are signed residual responses to the input labels. The Gram matrix records their inner products: its diagonal gives the squared size of each response, and its off-diagonal entries measure how

responses overlap. The second Gram moment sums the squares of all these inner products. It contains information that is lost when every response is replaced by its largest possible norm. Section 2.4 defines these quantities. Section 5 develops the general Gram estimates, and Sections 6–7 derive the byte and word bounds for AES. At the level of the introduction, the point is that MDS diffusion lets us average products over free coordinates while preserving the dependence between responses. The same idea is applied first to bytes and then to words.

Two transitions preserve the intermediate constraints. A one-step norm bound must allow every signed input in an intermediate space. The actual intermediate error was produced by the previous transition and satisfies additional constraints. For the larger residual degrees, we therefore estimate two transitions together. Fixing the outside labels leaves a set of compatible middle labels, an affine fiber; Section 5.4 keeps the sum over that fiber inside the operator. This preserves the shared middle labels and leads to a recurrence involving two consecutive error profiles. Proposition 1 states the general criterion, while Sections 8–9 derive its AES coefficients.

Convert to distance at the appropriate round. The error profile records both the number of residual words and the active support. Small supports have fewer possible outputs and hence a smaller conversion factor from Euclidean length to total variation. Retaining this information makes a uniform output bound possible even when some individual coordinates are large. Section 10 defines the normalized profile, checks the two starting bounds, and proves the alternating inequalities. The resulting output estimate gives Theorem 1; the 10/14/16 counts follow by substitution.

1.3 Related work

Complete differential probabilities and linear hulls. The early SPN literature developed guarantees from local S-box probabilities and the diffusion layer. Hong et al. considered optimal and near-optimal diffusion [27], and Kang et al. studied the diffusion properties used by such bounds [32]. Keliher, Meijer, and Tavares developed linear-hull bounds and applied them to Rijndael [36,35]. Park et al. analyzed the alternating byte permutation and column diffusion of Rijndael-like networks [50]; Sano et al. considered nested SPNs [53]. Chun et al. and Park et al. sharpened two-round bounds [15,51]. Cho et al. refined the differential analysis for SPNs and ARIA [13], and Keliher improved the AES estimates and showed that earlier two-round upper bounds were not tight [34].

Keliher and Sui determined the exact maximum expected differential probability $53/2^{34}$ over two AES rounds and obtained the four-round upper bound $(53/2^{34})^4$ [37]. Daemen and Rijmen explained the underlying interaction between substitution and diffusion [20]. Daemen et al. further studied expected four-round probabilities, including a simplified AES variant using the inversion S-box [18]. These results bound probabilities for specified endpoints. Our objective controls the total deviation of the whole output-pair distribution.

Affine equivalence also needs care when passing from a byte to a word. Canteaut and Roué showed that affine-equivalent S-boxes can have different two-round maximum expected differential probabilities and linear potentials when combined with fixed diffusion [9]. Their separate study of MDS-based SPNs shows that a maximum expected differential need not use the minimum possible number of active S-boxes [8]. Dunkelman et al. give a related limitation of active-S-box counting for a Feistel construction [25]. These observations support keeping the actual local maps and their diffusion constraints in the calculation.

Activity and operator analyses. Rønjom studied activity-pattern transition matrices for AES diffusion, while leaving their exact iterated agreement with the AES process unresolved [52]. We retain residual variation and its coupling to activity in the full distribution. Beyne's geometric treatment uses function spaces and operators to study cryptographic properties [3]; BLS provides the direct activity-space analysis used here [4]. Earlier operator and Fourier descriptions include the correlation matrices of Daemen, Govaerts, and Vandewalle [17] and the differential–linear relation of Chabaud and Vaudenay [11].

General indistinguishability amplification was developed by Maurer, Pietrzak, and Renner [44], with computational product theorems by Maurer and Tessaro [45]. LTV and BLS use amplification in their respective analyses. Here the final iteration is proved directly for the two-step error-profile recurrence.

Other permutation and SPN models. Kaplan, Naor, and Reingold give derandomized almost-independent permutation families without prescribing the AES round function [33]. Baignères and Vaudenay analyze AES* with independent random secret S-boxes [1]. Miles and Viola study SPN-based pseudorandom-function candidates and limited-independence guarantees [46]. Dodis et al. analyze SPNs with public random-permutation S-boxes [22]; Cogliati et al. extend this study to tweakable block ciphers [16]. Dodis, Stam, Steinberger, and Liu prove indifferentiability for confusion–diffusion networks in a public random-permutation model [23]. Those component models differ from our fixed AES S-box.

Layout graphs record the equality patterns among several evaluations of an SPN. Liu, Pelecanos, Tessaro, and Vaikuntanathan use this description to prove higher-order independence with random secret S-boxes [40]. Jain et al. study inverse walks and prove convergence to uniformity on permutations of the appropriate sign [31]. Their consequences for modified AES belong to different models from the concrete-AES comparison in Table 1. Constant-order independence of iterated local permutations does not in general imply pseudorandomness under arbitrarily many queries [24]; our theorem has the two-forward-query scope stated above. Beierle et al. give criteria excluding probability-one differentials and perfect linear approximations over two-round SPNs for every key [2]. Such fixed-key exclusions are distinct from a bound on the full output-pair law averaged over independent round keys.

Organization. Section 2 supplies the preliminaries, and Section 3 gives an informal guide to the argument. Section 4 develops the error profiles and two-step

contraction argument, and Section 5 proves the general Gram, source, and fiber estimates. Sections 6 and 7 derive the local AES bounds. Section 8 applies them to diffusion graphs; Section 9 sums supports and proves the initial estimates. Section 10 closes the recurrence and proves the mixing theorem.

2 Preliminaries

We specify the independent-key AES experiment and its difference process, then recall the superbox representation, activity decomposition, and local norms used throughout the proof.

Notation. All spaces and probability laws are finite. Laws are column vectors; e_x is a coordinate vector and $\mathbf{1}_B$ a set indicator. Norms use counting measure: $\|A\|$ is the Euclidean operator norm and $\|A\|_{\mathrm{HS}}^2 = \mathrm{Tr}(A^*A)$ is the sum of squared column norms. Here A^* is the adjoint and $A^\top$ the ordinary transpose. Write $\mathrm{Col}(\mu) = \sum_x \mu(x)^2$ and

$$\mathrm{TV}(\mu,\eta) = \tfrac{1}{2}\|\mu-\eta\|_1 = \tfrac{1}{2}\sum_x |\mu(x)-\eta(x)|.$$

2.1 AES with independent round keys

AES uses a 4×4 byte state. SubBytes (SB) applies the AES S-box, ShiftRows (SR) cyclically shifts rows, and MixColumns (MC) mixes each column [47]. Addition is bitwise XOR. Figure 1 shows the states in our experiment; composition $\circ$ is read right to left:

$$\begin{aligned} z_0 &= x + K_0, \\ z_i &= (\mathsf{MC}\circ\mathsf{SR}\circ\mathsf{SB})(z_{i-1}) + K_i \quad (1 \le i < r), \\ E_r(x) &= (\mathsf{SR}\circ\mathsf{SB})(z_{r-1}) + K_r. \end{aligned} \tag{3}$$

The independent uniform 128-bit keys are retained across queries. The last round omits MixColumns.

A word means one AES column of four bytes. We use

$$q = 2^{32}, \qquad Q = q-1, \qquad N = q^4 = 2^{128}, \qquad n = N-1.$$

The byte alphabet is $\mathbb{F}_{256} = \mathbb{F}_2[t]/(t^8+t^4+t^3+t+1)$. MixColumns applies the matrix

$$M = \begin{pmatrix} 2&3&1&1\\ 1&2&3&1\\ 1&1&2&3\\ 3&1&1&2 \end{pmatrix}, \qquad \mathsf{SR}(x)_{u,c} = x_{u,c+u},$$

where row and column indices are modulo four and the matrix entries use the standard byte encoding. The byte S-box formula is recalled in Section 6.

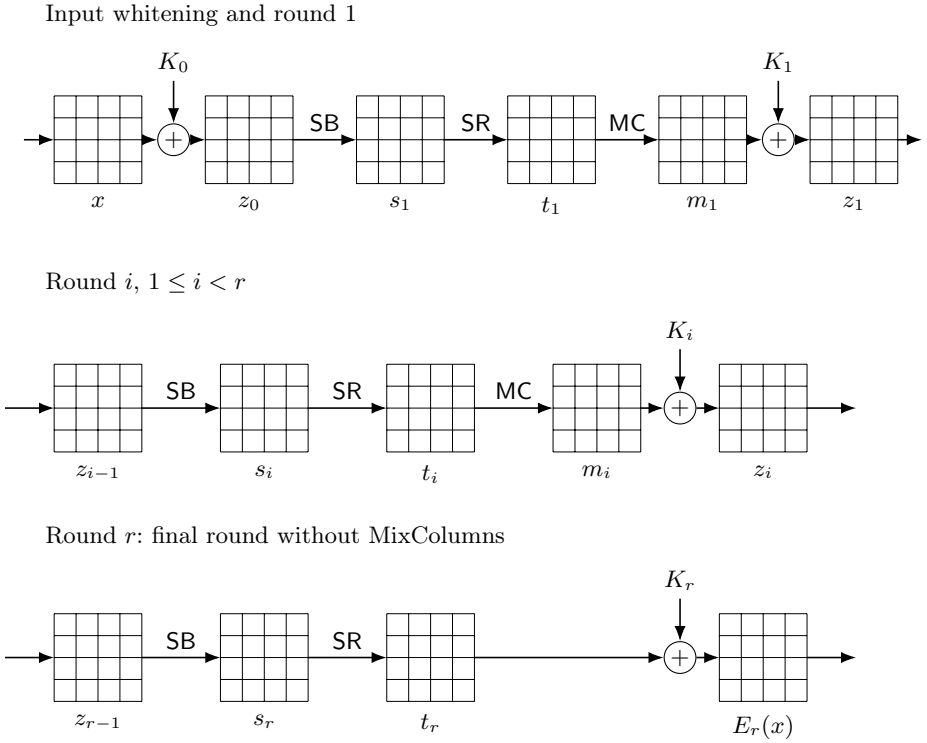

Fig. 1. AES with independent round keys. Rows show input whitening and round 1, an ordinary round i, and the final round without MixColumns. Each grid is a 4×4 byte state. Circled + denotes AddRoundKey (XOR).

2.2 Difference transitions and pairwise distance

Independent round keys make state differences a Markov chain: the next law depends only on the current difference. For a permutation s of a finite additive space V, define the normalized difference distribution table (DDT)

$$D_s(b, a) = |V|^{-1} \#\{x : s(x + a) - s(x) = b\}.$$

Its column a is the next-difference law when the first input is uniform and the difference is a. The zero difference is isolated, the nonzero block is doubly stochastic, and a change of variables gives $D_{s^{-1}} = D_s^*$.

Difference matrices need not multiply under composition without whitening. A fresh uniform key makes the first state uniform conditional on the current pair, while preserving its difference. Thus the next law depends only on that difference, and the matrices multiply [39,4]. Keys remain fixed across queries; this is an average over the retained tuple.

Distinct AES states have nonzero difference. We work on $\Omega = \mathbb{F}_{256}^{16} \setminus \{0\}$, of size $n = q^4 - 1$, with uniform law u.

Fix a round count r and write $E = E_r$. Independent input and output whitening give the exact pair identity

$$\Pr[E(x) = y, E(x+a) = y+b] = \frac{p_a(b)}{N}, \qquad \varepsilon_r = \max_{a \neq 0} \frac{1}{2} \left\| p_a - u \right\|_1, \quad (4)$$

where p_a is the output-difference law for input difference $a \in \Omega$. Input whitening removes dependence on x; output whitening makes the first ciphertext uniform independently of the difference. The ideal pair probability is $1/(N(N-1))$. Summing absolute deviations over y and $b \in \Omega$ proves the second identity [4].

Algorithms have finite private seeds, independent of the oracle, and finite outputs; auxiliary information is part of the seed.

Lemma 1 (Two adaptive queries). *The distance between the complete views of any classical algorithm making at most two forward queries to E_r or a retained uniform permutation is at most ε_r.*

Proof. Condition on the private seed. The first query, if made, is fixed, and its answer is uniform in both experiments. Conditional on that answer, stopping and the second plaintext are fixed. If the second plaintext is distinct, divide the first identity in (4) by $1/N$: the conditional second-answer distance is at most ε_r. A repeated query gives the same answer; a stopped branch adds none. For one retained uniform permutation the corresponding conditional law is uniform on the other $N-1$ outputs, by counting permutations with two specified values. Average over the common first-answer law and seed. Postprocessing the view cannot increase total variation. □

2.3 Superboxes and MDS diffusion

We use the AES *superbox* representation [20,21,4] to group two substitution layers. ShiftRows brings byte positions

$$(0, c), \quad (1, c+1), \quad (2, c+2), \quad (3, c+3)$$

into column c (indices modulo four). Figure 2 marks them for $c = 0$. Since SubBytes commutes with ShiftRows, moving ShiftRows before the first SubBytes keeps these four bytes together through SubBytes, MixColumns, key addition, and the next SubBytes. This 32-bit component is one superbox.

MDS codes and diffusion. An additive $[\ell, d]$ MDS code is a subgroup $\mathcal{C} \subseteq V^\ell$ of size q^d whose projection onto any d coordinates is bijective: any d coordinate values determine exactly one codeword. Such a set of d positions is called an *information set*.

AES diffusion follows the wide-trail strategy [19]. The graph of MixColumns is an additive $[8, 4]$ MDS code over $\mathbb{F}_{256}$: any four coordinates of (x, Mx) determine the pair. Equivalently, every square minor of M is nonzero. With w input bytes free and the rest zero, imposing $z < w$ output zeros leaves $w - z$ free values; $z \geq w$ forces zero. Imposing zeros is *shortening*; deleting coordinates is *puncturing* [42]. We use these projection properties as in Dinur [21].

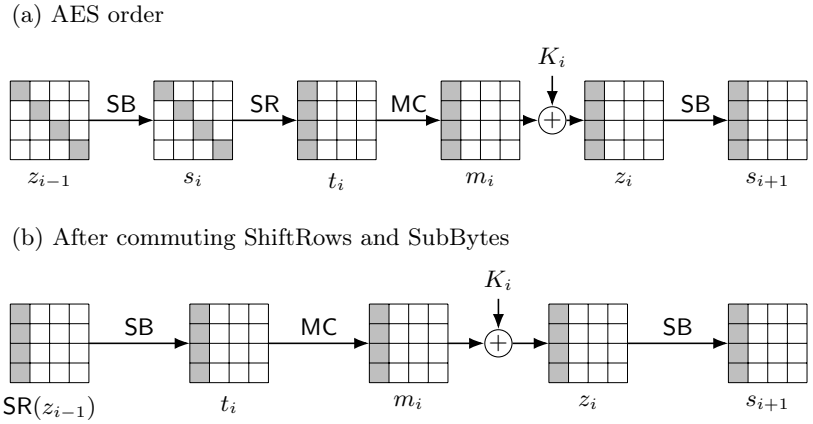

Fig. 2. One 32-bit superbox, with its four bytes shaded; unshaded bytes are arbitrary. Commuting ShiftRows and SubBytes turns (a) into the columnwise form (b). Averaging over keys gives the word difference channel K; four parallel copies form H. Here $1 \le i < r$.

Diffusion between superboxes. Between two superbox layers, the remaining AES operations mix the four words through the linear map

$$L = \mathsf{SR} \circ \mathsf{MC} \circ \mathsf{SR}, \qquad (Lx)_{u,c} = \sum_v M_{u,v} x_{v,c+u+v}.$$

Columns of x are words. Every square word-block minor of L is invertible. Thus the graph of L is an additive $[8, 4]$ MDS code over $V = \mathbb{F}_{256}^4$: any four word coordinates of (x, Lx) determine the pair. The same shortening and projection properties therefore hold on $V = \mathbb{F}_{256}^4$; scalar linearity over $\mathbb{F}_{2^{32}}$ is unnecessary.

Each individual word block of L is a byte permutation followed by nonzero byte scalings, so it preserves the number of nonzero bytes. We call such a block *byte-monomial*; Section 9.3 uses this additional property to bound the initial difference law.

Difference channels for grouped rounds. Let $K(b, a)$ be the probability that one superbox maps column difference a to b, averaged over its keys. Let H process four columns independently through K, and let Π apply L to differences: $\Pi e_a = e_{La}$. The word channel K includes all q differences. For state mixing, we restrict H and Π to the invariant nonzero space Ω.

Four rounds therefore consist of a superbox layer, linear mixing, and a second superbox layer, up to fixed input and output coordinate permutations. Writing $p_{a,4}$ for the resulting difference law from input difference a, we have

$$\mathcal{D}_4 = H\Pi H, \qquad p_{a,4} = \mathcal{D}_4 e_a. \tag{5}$$

Section 7.1 states their local estimates. For every even $r \geq 2$, the same grouping gives

$$p_{a,r} = H(\Pi H)^{r/2-1} e_a, \qquad f_{a,r} = p_{a,r} - u.$$

Thus $f_{a,r}$ is the centered error after r rounds.

Matching the AES round convention. It remains to justify the coordinate permutations used above. Put $B = \mathsf{SB} \circ \mathsf{MC} \circ \mathsf{SB}$, the four parallel superboxes with key additions temporarily suppressed. Since SR commutes with parallel substitution, for every $k \geq 1$ we have

$$(\mathsf{MC} \circ \mathsf{SR} \circ \mathsf{SB})^{\circ 2k} = \mathsf{MC} \circ \mathsf{SR} \circ B \circ (L \circ B)^{\circ(k-1)} \circ \mathsf{SR}.$$

Here $f^{\circ j}$ denotes j-fold composition, with $f^{\circ 0}$ the identity. Moving keys through fixed invertible linear maps preserves their independent uniform law. Omitting the final MixColumns leaves only an outside ShiftRows. Thus the physical input difference a is relabeled as $\mathsf{SR}(a)$, and a grouped output difference b corresponds to physical output difference $\mathsf{SR}(b)$. These fixed relabelings preserve all norms and the worst input distance [20,21,4].

2.4 Activity, residual variation, and local norms

An *activity pattern* specifies which words are nonzero; for example, $(1, 0, 1, 0)$ has two active words. Following BLS [4] and Dinur's exact-support decomposition [21], we separate pattern masses from variation among the nonzero values.

For a function f on the word space $V = \mathbb{F}_{256}^4$, define

$$(P_z f)(x) = \mathbf{1}_{x=0} f(0),$$
$$(P_c f)(x) = \mathbf{1}_{x\neq 0}\, Q^{-1} \sum_{y \neq 0} f(y),$$
$$P = P_z + P_c, \qquad P_e = I - P.$$

These are orthogonal projections. The residual $P_e f$ vanishes at zero and sums to zero. On a probability law, P preserves the mass at zero and spreads the remaining mass uniformly over nonzero words. Since the word channel fixes both activity directions, $KP = PK = P$. Its residual is therefore

$$R = K - P = KP_e, \qquad \|KP_e f\|_2 \leq \|R\|\, \|P_e f\|_2.$$

On four words, let $P_0 = P^{\otimes 4}$, restricted to Ω. It averages separately within each of the fifteen nonzero activity patterns. For instance, on $(1, 0, 1, 0)$ it averages over Q^2 states. The global uniform projection is $Uf = (\sum_x f(x))u$. A centered vector has total mass zero, so belongs to $u^\perp$. Within this space the activity and residual parts are

$$u^\perp = \mathcal{A} \oplus \mathcal{R}, \qquad P_{\mathcal{A}} = P_0 - U, \quad P_{\mathcal{R}} = I - P_0. \tag{6}$$

The first part records excess or deficit pattern mass; the second sums to zero within every pattern. The space $\mathcal{A}$ has dimension fourteen, since centering removes one global mean. The parallel channel H and its adjoint fix all activity vectors. Section 4.1 further separates residual variation by its word positions and active support.

Local norms and Gram moments. Let V be a binary vector space of size $q > 1$, with XOR as addition. For a matrix T acting on functions on V, the column Te_x is its response to input x. For AES, $T = K - P$, so this is the word-channel output minus its activity average, with entries of either sign. We use real matrices and real inner-product spaces throughout these estimates.

The *Gram matrix* of T is the matrix of inner products between these columns:

$$G = T^*T, \qquad G_{xy} = \langle Te_x, Te_y \rangle = \sum_{z \in V} T_{zx} T_{zy}.$$

Thus $G_{xx} = \|Te_x\|_2^2$ measures the squared size of one response, whereas G_{xy} measures the overlap of the responses to x and y, using unnormalized inner products. For the real residual responses in AES, a positive overlap means that their deviations reinforce in the inner product, and a negative overlap means that they oppose. The matrix is *positive semidefinite*: $f^*Gf = \|Tf\|_2^2 \geq 0$ for every f.

The *trace* $\operatorname{Tr} G = \sum_x G_{xx}$ sums the squared column sizes. We also retain the sum of the squared overlaps, called the second Gram moment. Our two local parameters are

$$\tau^2 = \operatorname{Tr}(G^2) = \sum_{x,y \in V} |G_{xy}|^2, \qquad \max_x G_{xx} \leq \delta. \tag{7}$$

The equality follows from the symmetry of G. A bound on $\operatorname{Tr}(G^2)$ may replace its exact value in every estimate below.

The operator norm $\|T\|$ is the largest amplification of Euclidean length: $\|Tf\|_2 \leq \|T\| \, \|f\|_2$ for every f. If $\|T\| \leq \nu$, then $\operatorname{Tr}(G^2) \leq \nu^2 \operatorname{Tr} G$.

Remark 1 (Relation to prior-work norms). Our residual norm $\|R\|$ is BLS's third singular value of K: the zero and uniform-nonzero directions supply its two invariant singular values. Restricting K to nonzero differences removes the zero direction; on that restriction, $\|R\|^2$ is the second Gram eigenvalue used by Dinur [4,21].

3 Technical Overview

Whitening reduces the pairwise question to the distribution of one nonzero input difference after encryption. We want to show that this distribution rapidly approaches the uniform law on nonzero blocks. Consider the four-round grouping from Section 2.3. Starting from a fixed difference, the first layer of four word channels produces a product law. Diffusion mixes the words and preserves the

Euclidean length of that law, but can make its coordinates dependent. The next local layer must therefore be analyzed without assuming a product input.

Two ideas allow us to exploit these dependencies. Within a single layer, we measure the overlaps between local responses before averaging over their common inputs. Across two successive transitions, we retain the constraints that both diffusion layers impose on the same intermediate state. Together these estimates control the signed error, the difference between the current law and the uniform law. We first explain the two estimates, then show how they give a decay factor of 2^{-41} for every two additional rounds.

Overlaps reveal the gain from averaging. Suppose two copies of the same local map receive a common uniformly chosen input label from an alphabet of size q. Write r_x for the real response vector to label x. Their joint average is $q^{-1}\sum_x r_x \otimes r_x$. The same label occurs in both factors, so averaging the two factors independently would change the expression. Define the Gram matrix by $G_{xy} = \langle r_x, r_y \rangle$. Expanding the squared length of the joint average gives

$$\left\| \frac{1}{q} \sum_x r_x \otimes r_x \right\|_2^2 = \frac{1}{q^2} \sum_{x,y} G_{xy}^2.$$

The sum of squared overlaps is the second Gram moment. It appears because both local responses depend on the same input.

This identity shows a quantitative gain in a simple case. Suppose the q response vectors are pairwise orthogonal and each has squared length δ. Only the diagonal Gram entries contribute, so the squared length of the joint average is δ^2/q. The triangle inequality applied before averaging gives only δ^2. Thus the overlaps can save a factor of q, even though all individual response lengths are the same. For AES we estimate these overlaps; we do not assume orthogonality.

An MDS code describes a more general dependence of the same kind. A uniform word of an $[\ell, d]$ MDS code is determined by any d coordinates, which are independent and uniform. If $2d \le \ell$, choose two disjoint groups of that size. In the squared-response expansion, Cauchy–Schwarz separates the products belonging to these two groups. Each group can then be summed over its free coordinates. This produces one Gram-moment factor per free coordinate; the remaining local factors use their largest squared length.

Theorem (informal). *Let the input be uniform on an additive $[\ell, d]$ MDS code over a finite vector space over $\mathbb{F}_2$ of size q, with $0 < 2d \le \ell$. Suppose each real local map has squared response lengths at most δ and second Gram moment at most τ^2. Then the squared Euclidean length of the average joint response is at most*

$$\left(\frac{\tau^2}{q^2}\right)^d \delta^{\ell-2d}.$$

Proposition 2 gives the full estimate. For initialization from fixed AES differences, the sources are nonuniform; Section 5.2 extends the common-input calculation to those laws.

Keep the parts of the error that contract differently. A local word channel treats activity and value variation differently. It preserves the point mass at zero and the uniform law on nonzero values, while contracting zero-sum variation among nonzero values. On four words, variation in one coordinate and variation in all four coordinates therefore have different local estimates. Diffusion couples these components, so we need to follow the amount of error in each of them.

We record their Euclidean lengths in an *error profile*. Each entry specifies the number of active words and the number of words carrying zero-sum variation, called the residual degree. For example, residual degree one on two active words and residual degree one on four active words have different supports, even though only one word carries variation in either case. There are fourteen such types for AES. The precise orthogonal decomposition is in Section 4.1. Retaining the support size also improves the final conversion to total variation: a vector supported on fewer states has a smaller Euclidean-to-absolute-mass factor.

Estimate two transitions before taking their norms. A component produced by one transition occupies particular directions in its output space. A separate norm bound on the next transition allows a worst-case input direction that may be different. Two coordinate projections illustrate how much information this can lose:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \|A\| = \|B\| = 1, \quad \|BA\| = 0.$$

The composition vanishes because the second map annihilates the image of the first. Multiplying their separate norms gives one. This is a reduced example of the reason to estimate a composed operator directly.

For general maps A and B, this dependence is visible in each entry of their composition:

$$(BA)_{z,x} = \sum_{y} (B)_{z,y} (A)_{y,x}.$$

Fix the input label x and output label z. Both factors must use the same intermediate label y. A large response of the first map contributes only through the response of the second map at that label. Separate norm bounds allow the two maps to attain their worst cases independently.

In the AES expansion, diffusion imposes linear relations on the byte or word labels of each transition. Fixing the outside labels leaves two sets of equations for the intermediate labels. Only assignments satisfying both sets contribute; the nonempty solution sets are affine fibers. We keep the sum of weighted local responses over each fiber inside the composed estimate. The number of compatible assignments and the overlaps between their responses can then improve the bound. Section 5.4 derives the inequalities for these signed operator sums, and Section 8 applies them to AES.

For intermediate residual degrees three and four, the concrete AES bounds benefit from retaining these common middle labels. We therefore use the composed estimates for those degrees and the one-transition estimates for degrees zero,

one, and two. To express this choice, write f_r for the error after r rounds and S for a grouped two-round transition minus the uniform projection. Let $P_{\le 2}$ and $P_{\ge 3}$ select residual degrees at most two and at least three. Inserting these complementary projections gives

$$f_{r+4} = SP_{\le 2} f_{r+2} + SP_{\ge 3} S f_r.$$

The first term uses a one-transition estimate on the current error. The second uses a two-transition estimate on the preceding error, with the high-degree middle component kept inside the composition. This identity, proved in Section 4.2 for even $r \ge 2$, explains why the recurrence uses two consecutive profiles.

From the recurrence to every later even round. We test a decay factor of 2^{-41} per two additional rounds. Let x_j be the profile after $4+2j$ rounds, with the common initial size, fixed component scales, and the candidate decay 2^{-41j} divided out. The normalization is specified in Section 10.2. The one- and two-transition estimates yield nonnegative matrices L, M such that

$$x_{j+1} \le \mathsf{L} x_j + \mathsf{M} x_{j-1} \qquad (j \ge 1).$$

Here the first matrix propagates the low-degree contribution and the second propagates the composed contribution. All inequalities between profiles are componentwise. The candidate rate is established by proving that these normalized profiles stay below fixed vectors; the alternating inequalities below provide that bound.

The large coordinates in the first few profiles alternate with the step index. We therefore use two upper profiles. Taking their coordinatewise maximum would combine large coordinates from different steps and increase the final distance estimate. The two-profile induction is as follows.

Theorem (informal). *Suppose the normalized profiles at eight and ten rounds lie below nonnegative vectors w_0 and w_1, respectively. If*

$$\mathsf{L} w_1 + \mathsf{M} w_0 \le w_0, \qquad \mathsf{L} w_0 + \mathsf{M} w_1 \le w_1,$$

then these two vectors bound every subsequent normalized profile in alternation. A common bound on their total-variation contributions gives the same geometric decay at every later even round count.

To obtain the matrices, we derive byte estimates, lift them to word channels, and combine them along the diffusion graphs (Sections 6–8). Starting the recurrence also uses the source laws generated by a fixed input difference. Those initial estimates may exploit the particular source; the operator bounds for subsequent propagation must apply to arbitrary signed errors. Section 9 assembles both kinds of estimates.

The resulting eight- and ten-round profiles satisfy the alternating bounds. The available six-round upper profile does not satisfy the same output bound, which is why the family starts at eight rounds.

Section 10 proves the two starting bounds and the alternating inequalities. Converting the upper profiles to total variation and restoring the scale gives

$$\varepsilon_{8+2k} < 2^{-92.19-41k} \qquad (k \geq 0).$$

The 10/14/16 corollary follows by substitution.

4 The Two-Step Mixing Method

We formalize the propagation argument of Section 3. An error profile records the lengths of orthogonal components, and a support-sensitive functional converts these lengths to total variation. Proposition 1 combines one-step and two-step operator estimates into a recurrence and gives conditions for its geometric decay.

4.1 Error profiles and total variation

Residual degree and active support. Let V be a finite word space of size $q > 1$ with distinguished zero, and let $b \geq 1$ be the number of words. On V, use the projections P_z, P_c, P_e of Section 2.4, with $Q = q - 1$. On $\Omega_b = V^b \setminus \{0\}$, define $P_{i,t}$, for $0 \leq i \leq t \leq b$ and $t \geq 1$, as the sum of all tensor projections having i factors P_e, $t-i$ factors P_c, and $b-t$ factors P_z. Their ranges are pairwise orthogonal and exhaust functions on Ω_b. There are $\binom{b}{t}\binom{t}{i}$ position patterns of type (i,t). Set

$$\mathbf{v}(f)_{i,t} = \|P_{i,t} f\|_2 . \tag{8}$$

The profile has $b(b+3)/2$ entries, one for each pair (i,t). Each entry is a component length; the corresponding subspace can have dimension greater than one. For example, when $b = 4$, types $(1,2)$ and $(1,4)$ both have residual degree one, but their active supports have sizes two and four.

Put $P_i = \sum_{t=\max(1,i)}^{b} P_{i,t}$. Orthogonality gives $\|P_i f\|_2^2 = \sum_t \mathbf{v}(f)_{i,t}^2$. The activity part $P_0 f$ is constant on each support pattern. If f is centered, then $P_0 f$ also has total mass zero, since every positive-degree component has a zero-sum residual factor.

For AES, $b = 4$ and $q = 2^{32}$ give fourteen types. We postpone their ordering and numerical scales to Section 9. The contraction criterion uses only the orthogonal decomposition. The distance estimate below also uses the number of states in each support.

From support profiles to total variation. Cauchy–Schwarz converts Euclidean length to absolute mass with a factor equal to the square root of the number of states in the support. For $1 \leq t \leq b$ active words, define the normalized factor

$$\gamma_t = q^{-b/2} \sqrt{\binom{b}{t}(q-1)^t}.$$

Residual components with the same support size are orthogonal, so their lengths combine in quadrature. For the activity error, zero total mass controls the all-active contribution by the contributions on smaller supports. Doubling the latter therefore accounts for the entire activity error. These observations give the monotone, positively homogeneous functional on nonnegative profile vectors

$$\Psi_b(v) = 2\sum_{t=1}^{b-1} \gamma_t v_{0,t} + \sum_{t=1}^{b} \gamma_t \sqrt{\sum_{i=1}^{t} v_{i,t}^2}. \tag{9}$$

Lemma 2 (Support-sensitive distance). *For every centered function f on Ω_b,*

$$\tfrac{1}{2}\|f\|_1 \le \frac{q^{b/2}}{2}\Psi_b(\mathbf{v}(f)).$$

Proof. Put $v = \mathbf{v}(f)$. For each t, the positive-degree components are orthogonal and are supported on the same union of $\binom{b}{t}(q-1)^t$ states. Their sum therefore satisfies

$$\left\|\sum_{i=1}^{t} P_{i,t} f\right\|_1 \le q^{b/2}\gamma_t \sqrt{\sum_{i=1}^{t} v_{i,t}^2}$$

by Cauchy–Schwarz. Summing over the support sizes gives the residual term.

The activity part is constant on each pattern and has total mass zero. On the all-active pattern it has one constant value, so

$$\begin{aligned}\|P_{0,b} f\|_1 &= \left|\sum_{x\in\Omega_b} (P_{0,b} f)(x)\right| \\ &\le \sum_{t=1}^{b-1} \|P_{0,t} f\|_1 .\end{aligned}$$

Cauchy–Schwarz bounds the contribution at each smaller support size by $q^{b/2}\gamma_t v_{0,t}$. Twice their sum bounds the full activity ℓ_1 norm. Adding the activity and residual estimates and dividing by two proves the claim. □

For AES, the weights for one, two, or three active words are much smaller than the all-active weight. Keeping these weights allows larger profile coordinates on small supports while still obtaining a small distance. Section 10 specializes this functional to the AES profiles.

4.2 Two-step decomposition and contraction

We estimate selected intermediate components through two consecutive transitions, keeping their middle projection inside the composed operator. The remaining components use one-transition estimates. The following criterion applies to any linear operator with an orthogonal decomposition; it requires no probabilistic assumption on the intermediate vectors. All vector inequalities below are componentwise.

Proposition 1 (Two-step contraction). (i) Operator bounds. *Let S be a linear operator on a finite-dimensional Hilbert space, and let $P_1, \dots, P_m$ be orthogonal projections onto pairwise orthogonal subspaces with sum I. Choose $\mathcal{H} \subseteq \{1, \dots, m\}$ and put $P_{\mathcal{H}} = \sum_{\beta \in \mathcal{H}} P_\beta$. Suppose A, B are nonnegative $m \times m$ matrices satisfying*

$$\|P_\alpha S P_\beta\| \le A_{\alpha\beta} \quad (\beta \notin \mathcal{H}), \qquad \|P_\alpha S P_{\mathcal{H}} S P_\beta\| \le B_{\alpha\beta},$$

with $A_{\alpha\beta} = 0$ for $\beta \in \mathcal{H}$. For $g_n = S^n g_0$, the profiles $x_{n,\alpha} = \|P_\alpha g_n\|_2$ obey

$$x_{n+2} \le A x_{n+1} + B x_n \qquad (n \ge 0). \tag{10}$$

(ii) Alternating bounds. *Let a nonnegative sequence obey* (10). *Suppose $w_0, w_1 \in \mathbb{R}^m_{>0}$, $\rho > 0$, $c \ge 0$, and*

$$\begin{aligned} \rho A w_1 + B w_0 &\le \rho^2 w_0, \quad x_0 \le c w_0, \\ \rho A w_0 + B w_1 &\le \rho^2 w_1, \quad x_1 \le c\rho w_1. \end{aligned} \tag{11}$$

Then, for every integer $n \ge 0$,

$$x_{2n} \le c\rho^{2n} w_0, \qquad x_{2n+1} \le c\rho^{2n+1} w_1.$$

Proof. For part (i), insert $I - P_{\mathcal{H}} + P_{\mathcal{H}} = I$ to obtain

$$g_{n+2} = S(I - P_{\mathcal{H}}) g_{n+1} + S P_{\mathcal{H}} S g_n.$$

Project onto P_α and expand both input vectors with $\sum_\beta P_\beta = I$. The first term uses only $\beta \notin \mathcal{H}$; in the second, keep the middle projection inside the composed operator. The triangle inequality and the stated block bounds give (10).

For part (ii), the two initial bounds start an induction. From the bounds at indices $2n, 2n+1$, obtain

$$\begin{aligned} x_{2n+2} &\le c\rho^{2n}(\rho A w_1 + B w_0) \le c\rho^{2n+2} w_0, \\ x_{2n+3} &\le c\rho^{2n+1}(\rho A w_0 + B w_1) \le c\rho^{2n+3} w_1. \end{aligned}$$

The second line uses the newly obtained even-index bound. Nonnegativity of A, B justifies every substitution. □

Using the criterion. For a proposed ρ, the two invariant conditions in (11) are linear inequalities in w_0, w_1. Once these weights are chosen, both initial conditions hold with

$$c = \max_{1 \le \alpha \le m} \left\{ \frac{x_{0,\alpha}}{(w_0)_\alpha}, \frac{x_{1,\alpha}}{\rho (w_1)_\alpha} \right\}.$$

Setting $w_0 = w_1 = w$ recovers the single-vector test $(\rho A + B) w \le \rho^2 w$.

The profile bound also controls every componentwise monotone, positively homogeneous map $\Theta : \mathbb{R}^m_{\ge 0} \to \mathbb{R}_{\ge 0}$:

$$\Theta(x_n) \le c\rho^n \max\{\Theta(w_0), \Theta(w_1)\}.$$

Thus $\rho < 1$ gives geometric decay of the output. For centered errors with the product-space profiles of Section 4.1, we may take $\Theta = \Psi_b$ and apply Lemma 2 to bound total variation.

Specialization to AES. Write $T = H\Pi$ and $S = T - U$, with U the uniform projection of Section 2.4. Both H and Π preserve u. The grouped law in Section 2.3 gives

$$f_{a,r+2} = Tf_{a,r} = Sf_{a,r} \qquad (r \geq 2 \text{ even}), \tag{12}$$

since each full error has total mass zero. We suppress a when a bound holds uniformly over nonzero differences.

An individual activity component $P_{0,t}f$ can have nonzero mass even when f is centered. Using S subtracts the uniform part of its image. These subtractions cancel when the components are recombined, since $Uf = 0$. We therefore use S for the component bounds in Proposition 1.

For AES, choose the intermediate components of residual degrees three and four. In the notation of Proposition 1, this means $P_{\mathcal{H}} = P_{\geq 3}$, where

$$P_{\leq 2} = \sum_{i \leq 2, t} P_{i,t}, \qquad P_{\geq 3} = \sum_{i \geq 3, t} P_{i,t}.$$

These are complementary orthogonal projections on functions on Ω. A residual coordinate has zero total sum, so $UP_{\geq 3} = P_{\geq 3}U = 0$.

Lemma 3 (Memory identity). *For every nonzero input difference and even $r \geq 2$,*

$$f_{r+4} = SP_{\leq 2}f_{r+2} + SP_{\geq 3}Sf_r, \qquad SP_{\geq 3}S = TP_{\geq 3}T. \tag{13}$$

Proof. The operator decomposition in the proof of Proposition 1, with $P_{\mathcal{H}} = P_{\geq 3}$, gives the first identity. For the second, expand $S = T - U$. Both products of $P_{\geq 3}$ with U vanish, leaving $TP_{\geq 3}T$. □

The high-degree restriction is on the intermediate component; the input and output components may have any residual degree. The composed block can therefore be estimated using the two diffusion layers together, as in Section 5.4.

Section 10.2 removes the desired factor 2^{-41} per grouped transition by normalizing the profiles. The AES induction then uses Proposition 1 with $\rho = 1$, so its two inequalities describe an invariant region. The starting profiles at eight and ten rounds and the two alternating weight vectors are established in Section 10.

5 Gram Moments and MDS Fibers

The recurrence of Proposition 1 requires estimates for both its initial profiles and its propagation operators. For additive MDS diffusion, the first task leads to responses to uniform codewords and nonuniform common sources. The second requires bounds that hold for every signed error vector. We obtain these by retaining information coordinates and summing over the affine fibers shared by two transitions.

The code expansions and Frobenius estimates build on BLS [4]; the use of Gram traces and MDS marginals also appears in Dinur's analysis [21]. The estimates below retain the second Gram moment of Section 2.4 within the code average.

5.1 Responses to MDS codes

For a uniform codeword, its information coordinates are independent and uniform, but the remaining coordinates are determined by them. Thus a product of local responses cannot in general be averaged by multiplying ℓ separate averages. We expand the squared response into Gram entries and average over information sets. The second Gram moment then measures overlaps between distinct labels without replacing each overlap by its largest possible value.

Diffusing a coordinate subspace with d free words produces such a code (Section 2.3). We first use its unit indicator; the probability-law normalization follows immediately after Proposition 2.

The unit indicator $c_{\mathcal{C}} = q^{-d/2}\mathbf{1}_{\mathcal{C}}$ has Euclidean norm one. Write $T^{\otimes\ell}$ for the separate action on the coordinates, so $(T^{\otimes\ell})_{y,x} = \prod_i T_{y_i,x_i}$ for $x, y \in V^\ell$. We bound its squared response $\left\|T^{\otimes\ell}c_{\mathcal{C}}\right\|^2$.

Two dependent coordinates. For $\ell = 2, d = 1$, the code has the form (L_1x, L_2x) for two bijections L_1, L_2. Its response is the sum of the product vectors $Te_{L_1x} \otimes Te_{L_2x}$, each with coefficient $q^{-1/2}$. Both factors use the same label x. Squaring pairs the term indexed by x with the term indexed by y; the inner product of two product vectors is the product of their local inner products. Taking absolute values and applying Cauchy–Schwarz gives

$$\begin{aligned}\|(T \otimes T)c_{\mathcal{C}}\|_2^2 &\le q^{-1} \sum_{x,y} |G_{L_1x,L_1y}|\,|G_{L_2x,L_2y}| \\ &\le q^{-1} \sum_{u,v} |G_{uv}|^2 = \tau^2/q.\end{aligned}$$

The factor q^{-1} comes from squaring the unit-indicator normalization. Each bijection only relabels the sum over pairs of source labels, so both Cauchy–Schwarz factors are τ. For $2d \le \ell$, the same argument uses two disjoint information sets of size d: each replaces one of the two coordinates in this example. The remaining $\ell - 2d$ factors are bounded by δ. This yields one factor τ^2/q per free direction.

Product averages over homomorphisms of finite Abelian groups also arise in the discrete Hölder–Brascamp–Lieb inequalities studied by Christ [14]. Here two disjoint information sets suffice for the Cauchy–Schwarz argument. We give that specialization explicitly in Proposition 2, including the Fourier-dual case.

Fourier duality for larger dimensions. For $d > \ell/2$, the code has fewer constraints than free coordinates. A Fourier change of basis lets us average over those constraints instead, by exchanging the code with its dual [42]. Fix a nondegenerate symmetric binary bilinear pairing $[\omega, x]$ on V, such as the dot product, and set $\mathcal{F}_{\omega,x} = q^{-1/2}(-1)^{[\omega,x]}$. These normalized parity functions are orthonormal, so $\mathcal{F} = \mathcal{F}^* = \mathcal{F}^{-1}$ preserves lengths. The matrix $\mathcal{F}T\mathcal{F}$ represents T in this basis; its columns need the same bound only when $d > \ell/2$. The dual has $q^{\ell-d}$ codewords, so this change of basis lets the proof use the smaller of the number of free coordinates and the number of constraints.

Proposition 2 (Gram-moment bound for an MDS-code input). *Let V be a finite vector space over $\mathbb{F}_2$ with $q = |V| > 1$. Let T be a real matrix indexed by V, and set $\tau = \sqrt{\mathrm{Tr}((T^*T)^2)}$, the square root of its second Gram moment. Let $\mathcal{C} \subseteq V^\ell$ be an additive $[\ell, d]$ MDS code with $0 < d < \ell$, and put $a = \min(d, \ell - d)$. Suppose $\|Te_x\|_2^2 \le \delta$ for every $x \in V$. If $d > \ell/2$, also suppose $\|\mathcal{F}T\mathcal{F}e_x\|_2^2 \le \delta$ for every x. Then*

$$\left\|T^{\otimes \ell} c_{\mathcal{C}}\right\|^2 \le \left(\frac{\tau^2}{q}\right)^a \delta^{\ell - 2a}. \tag{14}$$

The same bound holds for $T_1 \otimes \cdots \otimes T_\ell$ if every T_i has second Gram moment at most τ^2 and obeys the same column bounds, including its Fourier orientation when needed.

The uniform probability law on the code is $q^{-d/2} c_{\mathcal{C}}$. Its squared response is therefore q^{-d} times the bound in (14). For $2d \le \ell$, this replaces each factor τ^2/q by τ^2/q^2, giving the probability-law estimate in Section 3.

Proof (of Proposition 2). We prove the mixed-map statement and put $G_i = T_i^* T_i$. Cauchy–Schwarz on two columns gives $|(G_i)_{uv}| \le \sqrt{(G_i)_{uu}(G_i)_{vv}} \le \delta$.

The case $d \le \ell/2$. Choose disjoint information sets I, J, each of size d. Expand the squared norm and bound the $\ell - 2d$ Gram entries outside $I \cup J$ by δ. Cauchy–Schwarz in the pair (x, y) gives

$$\begin{aligned}\left\|\left(\bigotimes_i T_i\right) c_{\mathcal{C}}\right\|^2 &\le q^{-d}\delta^{\ell-2d} \sum_{x,y\in\mathcal{C}} \prod_{i\in I} |(G_i)_{x_i,y_i}| \prod_{i\in J} |(G_i)_{x_i,y_i}| \\ &\le q^{-d}\delta^{\ell-2d} \prod_{B\in\{I,J\}} \left(\sum_{x,y\in\mathcal{C}} \prod_{i\in B} |(G_i)_{x_i,y_i}|^2\right)^{1/2}.\end{aligned}$$

For either information set B, the coordinate pairs (x_i, y_i) range freely and independently over $(V \times V)^d$. Hence

$$\sum_{x,y\in\mathcal{C}} \prod_{i\in B} |(G_i)_{x_i,y_i}|^2 = \prod_{i\in B} \sum_{u,v\in V} |(G_i)_{uv}|^2 \le \tau^{2d}.$$

The product of the two square roots is at most τ^{2d}. This proves (14) with $a = d$.

The case $d > \ell/2$. Define the dual code

$$\mathcal{C}^\perp = \{\omega \in V^\ell : \textstyle\sum_i [\omega_i, x_i] = 0 \text{ for every } x \in \mathcal{C}\}.$$

Summing a parity function over the subgroup $\mathcal{C}$ gives zero unless it is identically one there. Consequently $\mathcal{F}^{\otimes\ell} c_{\mathcal{C}} = c_{\mathcal{C}^\perp}$, where $|\mathcal{C}^\perp| = q^{\ell-d}$. The dual is an additive $[\ell, \ell - d]$ MDS code: a dual codeword vanishing on $\ell - d$ coordinates is supported on d coordinates, and the bijective projection of $\mathcal{C}$ there lets us vary those d entries independently. Nondegeneracy of the pairing then forces every entry of

the dual codeword to be zero. Each projection of the dual onto $\ell - d$ coordinates is therefore injective, and its size makes it bijective. Since $\mathcal{F}$ preserves lengths,

$$\left\| \left(\bigotimes_i T_i \right) c_{\mathcal{C}} \right\| = \left\| \left(\bigotimes_i (\mathcal{F} T_i \mathcal{F}) \right) c_{\mathcal{C}^\perp} \right\|.$$

The transformed map has Gram matrix $\mathcal{F} G_i \mathcal{F}$. Since $\mathcal{F}^2 = I$, its second moment is $\mathrm{Tr}(G_i^2) \le \tau^2$. The assumed Fourier column bounds supply δ. Apply the first case to the dual code, now with $a = \ell - d$. □

5.2 Nonuniform common sources

A nonuniform common source assigns a weight $\mu(x)$ to each local product response. Expanding the squared norm pairs source values x, y and multiplies their Gram entries, now with weights $\mu(x)\mu(y)$. Such sources occur in the initialization of Section 9.3.

The *Schur product* is entrywise multiplication, $(A \odot B)_{xy} = A_{xy} B_{xy}$. Write $A \succeq 0$ for positive semidefiniteness and $A \preceq B$ for $B - A \succeq 0$. The Schur product theorem [29] preserves positive semidefiniteness under this product, yielding the following bound.

Lemma 4 (Weighted Schur bound). *Let $j \ge 2$. If $G_1, \dots, G_j \succeq 0$ and $G_1 \preceq \nu^2 I$, then for every vector μ indexed by their columns,*

$$\mu^*(G_1 \odot \cdots \odot G_j)\mu \le \nu^2 \sum_x |\mu(x)|^2 \prod_{i=2}^{j} (G_i)_{xx}. \tag{15}$$

Proof. Write $G_1 = A^* A$, with $\|A\| \le \nu$, and choose Gram vectors $(G_i)_{xy} = \langle b_{i,x}, b_{i,y} \rangle$ for $i \ge 2$. Put $w_x = \bigotimes_{i=2}^{j} b_{i,x}$. Keep the source label until the last map is applied:

$$\begin{aligned} \mu^*(G_1 \odot \cdots \odot G_j)\mu &= \left\| (A \otimes I) \sum_x \mu(x) e_x \otimes w_x \right\|^2 \\ &\le \nu^2 \sum_x |\mu(x)|^2 \, \|w_x\|^2 . \end{aligned}$$

The last sum is exact before applying A, because distinct source labels are orthogonal. Since $\|w_x\|^2 = \prod_{i=2}^{j} (G_i)_{xx}$, this is (15). □

The Schur bound uses the operator norm of one map and the column energies of the others. A point-probability bound gives a second estimate: retain two Gram factors in the sum and use their second moments. This preserves the gain from averaging their overlaps even for a nonuniform source.

Corollary 2 (Shared source with a point-probability bound). *Let $A_i : \mathbb{R}^V \to \mathcal{H}_i$ be matrices mapping into finite-dimensional inner-product spaces, with Gram matrices $G_i = A_i^* A_i$. Suppose $\operatorname{Tr}(G_i^2) \le \tau^2$ and $(G_i)_{xx} \le \delta$ for every i, x. If $j \ge 2$ and $\|\mu\|_\infty \le p$, then*

$$\left\| \sum_x \mu(x) \bigotimes_{i=1}^{j} A_i e_x \right\|^2 \le p^2 \tau^2 \delta^{j-2}. \tag{16}$$

Here $\|\mu\|_\infty = \max_x |\mu(x)|$. Each product uses the same label x in every coordinate, so the sum represents a common source. For a probability law, p bounds each point probability; signed real weights are also allowed.

Proof. Expand the squared norm and take absolute values. Bound each source weight by p and each of the last $j-2$ Gram entries by δ. Cauchy–Schwarz for the two remaining factors gives

$$\left\| \sum_x \mu(x) \bigotimes_{i} A_i e_x \right\|^2 \le p^2 \delta^{j-2} \sum_{x,y} |(G_1)_{xy} (G_2)_{xy}| \le p^2 \delta^{j-2} \sqrt{\operatorname{Tr}(G_1^2) \operatorname{Tr}(G_2^2)}.$$

This proves the claim. □

5.3 Encoding and restricted sources

Encoding arbitrary coefficients. An information set lets us retain independent labels while applying the local maps to dependent coordinates. This gives an operator bound for any source coefficients, including the restricted laws considered next.

Lemma 5 (Encoding through an information set). *Suppose projection of $C \subseteq V^\ell$ onto d coordinates is bijective. If $\|A_i\| \le \nu$ and $\|A_i e_x\|_2^2 \le \delta$ for all i, x, the map*

$$E_C f = \sum_{x \in C} f(x) \bigotimes_{i=1}^{\ell} A_i e_{x_i}$$

satisfies $\|E_C\| \le \nu^d \delta^{(\ell-d)/2}$.

Proof. Index the codewords by their information coordinates. The tensor product of the maps on these coordinates has norm at most ν^d. Each remaining coordinate contributes a Gram factor whose diagonal is at most δ. Apply Lemma 4 with the first Gram matrix bounded by $\nu^{2d} I$: the squared response is at most $\nu^{2d} \delta^{\ell-d} \sum_{x \in C} |f(x)|^2$. If there are no remaining coordinates, the tensor-product norm gives this bound directly. □

Restricting a probability law. The common-source bounds also apply when diffusion forces some outputs to zero. A marginal keeps selected coordinates and sums over the others. We retain the probability of each zero event in its marginal, so the restricted law need not have mass one. This follows the MDS marginal method of Dinur [21]; the form below is used for both word-column estimates in Section 7.2 and four-round initialization in Section 9.3.

Lemma 6 (MDS zero events and restricted responses). *Let V be an elementary abelian 2-group of size m. Feed $w \in \{1,2,3,4\}$ independent laws supported on $V \setminus \{0\}$ into w positions of an additive four-coordinate MDS map, with the other inputs zero. Each input law has collision at most ρ and largest point probability at most p. Write Y for the output tuple.*

(a) Specified zero events. *Any specified set of $z \geq w$ output zeros is impossible. For a specified set $\mathcal{Z}$ of $z = 1,2,3$ positions, $\Pr[Y_{\mathcal{Z}} = 0] \leq U_{w,z}$, where*

$$U_{w,z} = \begin{cases} \rho^{\min(z,w-z)} p^{\max(0,2z-w)}, & 1 \leq z < w, \\ 0, & z \geq w. \end{cases} \tag{17}$$

The other output coordinates are unrestricted. If $\rho \leq p$, the simpler bound $U_{w,z} \leq \rho p^{z-1}$ holds whenever $1 \leq z < w$.

(b) Residual response. *Let A isolate zero and be doubly stochastic on nonzero symbols. Let P fix the mass at zero and average the other entries over $V \setminus \{0\}$, as in Section 2.4. Suppose $R = A - P$ has norm at most s_0 and each nonzero column of A has collision at most ρ. Put $\Delta_\rho = \rho - 1/(m-1)$. For disjoint output sets $\mathcal{Z}, S$ with $|\mathcal{Z}| = z < w$ and $|S| = j \geq 1$, define*

$$h_{\mathcal{Z}}(y) = \Pr[Y_S = y,\ Y_{\mathcal{Z}} = 0].$$

With $d = w - z$ and $a = \min(d, j)$,

$$\left\| R^{\otimes j} h_{\mathcal{Z}} \right\|_2^2 \leq p^{2z} (\rho s_0^2)^a \Delta_\rho^{j-a}. \tag{18}$$

The same bound holds with $h_{\mathcal{Z}}(y) = \Pr[Y_S = y,\ Y_{\mathcal{Z}} = 0,\ \mathcal{E}]$ for any further indicator event $\mathcal{E}$ on the input tuple. These marginals are not renormalized following restriction.

Proof. Part (a): probabilities of zero events. If $z \geq w$, the zero equations force all active inputs to zero by MDS, contradicting their support. Hence suppose $z < w$ and put $d = w - z$. Write μ_I for the product law on an input set I. First suppose $z \leq d$. Choose disjoint input sets I, J of size z and fix all remaining inputs. The zero equations determine the J coordinates as an affine bijection L of the I coordinates: any z input columns of the selected MDS equations are invertible. The conditional probability of the zero event is therefore at most

$$\sum_{x \in V^z} \mu_I(x) \mu_J(L(x)) \leq \|\mu_I\|_2 \|\mu_J\|_2 \leq \rho^z.$$

Averaging over the fixed inputs preserves this bound.

If $d \le z$, choose disjoint input sets I, J of size d. Each parameterizes the zero-solution space bijectively. Bound the other $w-2d = z-d$ probability factors by p, then apply Cauchy–Schwarz on that space to the two product laws. This gives $p^{z-d}\rho^d$. Together the two cases prove (17). The simpler bound replaces all but one collision factor by p.

Part (b): the residual response. After z outputs are fixed to zero, any $d = w-z$ active input coordinates parameterize the solution space. The other z input probability factors are at most p. Denote the restricted density on the free coordinates by $\eta_{\mathcal{Z}}$. It is pointwise at most p^z times the product of their input laws, so

$$\|\eta_{\mathcal{Z}}\|_2^2 \le p^{2z}\rho^d.$$

The same domination holds with the additional indicator $\mathcal{E}$. The marginal $h_{\mathcal{Z}}$ sums this density over all source points giving each marked output tuple. All comparisons with the product law are made at this stage, before the signed residual maps act.

If $j \ge d$, the marked outputs lie in an additive $[j, d]$ MDS code: any d of them identify the source point. The density need not be uniform on this code. Transport $\eta_{\mathcal{Z}}$ to it; the bijection preserves its squared norm. Every residual column has squared norm at most Δ_ρ, since the nonzero columns lose the uniform energy $1/(m-1)$ and the zero column vanishes. Lemma 5 therefore gives

$$\left\|R^{\otimes j} h_{\mathcal{Z}}\right\|_2^2 \le s_0^{2d}\Delta_\rho^{j-d}\,\|\eta_{\mathcal{Z}}\|_2^2,$$

which is (18) with $a = d$. Each output beyond the information set contributes a column-energy factor Δ_ρ.

If $j < d$, fix any $d-j$ free inputs. The map from the remaining j inputs to the j marked outputs is bijective: together these prescriptions specify d coordinates of the shortened MDS code. Under the unrestricted free product law, the marked marginal is therefore the convolution of a bijective image of j independent inputs with the contribution of the remaining inputs. Its collision is at most ρ^j, since averaging translates cannot increase the Euclidean norm. Pointwise domination gives $\|h_{\mathcal{Z}}\|^2 \le p^{2z}\rho^j$; applying $R^{\otimes j}$ proves the same formula with $a = j$. □

5.4 Two-step compositions on signed inputs

To bound a two-step operator, fix the input and output labels outside the composition and sum over the compatible middle labels. Each transition restricts these labels to an affine fiber. The normalized fiber maps below express the composition as an operator on arbitrary coefficients, as required by Proposition 1.

The operator determined by the fibers. Let $C \subseteq V^\ell$ be an additive MDS code. Suppose it is the middle projection of a larger shortened MDS graph, and assume this projection is injective. Thus each middle label determines a unique outside label. In the AES application, the graph dimension is at most the number of middle coordinates, so the MDS property supplies this injectivity (Section 8.1).

If the outside projection has rank I and C has dimension s, a fixed admissible outside label x leaves a fiber C_x of size q^a, $a = s - I$. Shortening the graph at these outside coordinates makes every nonempty fiber an affine MDS code on the middle coordinates. In particular, any a of its middle coordinates parameterize the fiber. We use this property for both codes and all fiber partitions below. Let X_C be the set of admissible outside labels and let $x(u)$ be the outside label associated with $u \in C$. Write U_C for the coordinate inclusion from $\ell_2(C)$ to $\ell_2(V^\ell)$. Define $J_C : \ell_2(X_C) \to \ell_2(C)$ by

$$(J_C f)(u) = q^{-a/2} f(x(u)), \qquad u \in C.$$

The normalization is determined by the fiber size:

$$\|J_C f\|_2^2 = \sum_{x \in X_C} \sum_{u \in C_x} q^{-a} |f(x)|^2 = \sum_{x \in X_C} |f(x)|^2,$$
$$(J_C^* g)(x) = q^{-a/2} \sum_{u \in C_x} g(u).$$

Thus J_C is an isometry and its adjoint takes a normalized sum over each fiber. These are identities for arbitrary signed or complex coefficients; no uniformity assumption on the source is needed.

Assume all local maps A_j have norm at most ν, row and column squared norms at most δ, and total squared norm at most h. For two codes $C, D \subseteq V^\ell$ with dimensions s, t and outside-fiber dimensions a, b, define their normalized overlap by

$$\mathcal{O}_{C,D} = J_C^* U_C^* (A_1 \otimes \cdots \otimes A_\ell) U_D J_D. \tag{19}$$

Its entry at outside labels x, y is

$$\mathcal{O}_{C,D}(x, y) = q^{-(a+b)/2} \sum_{\substack{u \in C_x \\ v \in D_y}} \prod_{j=1}^{\ell} A_j(u_j, v_j).$$

Thus both sets of middle constraints remain inside the sum. The factor $q^{-(a+b)/2}$ comes from the two unit fiber indicators.

Information-set estimates. We first estimate this overlap by keeping information coordinates as orthogonal labels. The resulting encoder norms involve products of local row or column energies summed over a fiber. For nonnegative integers n, a, define

$$F(n, a) = \begin{cases} \delta^{n-a} h^a, & a \le n, \\ q^{a-n} h^n, & a > n. \end{cases} \tag{20}$$

Lemma 7 (Collision sum on an affine MDS fiber). *Let $g_1, \ldots, g_n$ be nonnegative functions on V satisfying $g_j \le \delta$ and $\sum_x g_j(x) \le h$. On an affine MDS fiber of dimension a, observed in n distinct coordinates, the sum of $\prod_{j=1}^n g_j(u_j)$ is at most $F(n, a)$.*

Proof. If $a \le n$, choose a of the observed coordinates as an information set. Bound the other factors by δ. The sum over the chosen coordinates then factors into a sums, each at most h. If $a > n$, every observed tuple has q^{a-n} preimages in the fiber; summing its n factors gives the other branch. □

For example, a one-dimensional fiber observed in two coordinates costs $F(2,1) = \delta h$: one factor is summed and the other is bounded pointwise. The squared normalization in J_C divides this by q. This is the origin of the factor $\delta h/q$ below.

If $s + t \ge \ell$, we can choose an information set for each code so that their union covers all middle coordinates. Their intersection has $s + t - \ell$ coordinates; only these require local operator norms. On the remaining coordinates, Lemma 7 supplies the energy estimates.

Lemma 8 (Information-set overlap). *If $s + t \ge \ell$, then*

$$\|\mathcal{O}_{C,D}\| \le \nu^{s+t-\ell} \sqrt{\frac{F(\ell - s, a) F(\ell - t, b)}{q^{a+b}}}. \tag{21}$$

Proof. Partition the positions into A, B, E of sizes $\ell - t$, $s + t - \ell$, and $\ell - s$. Then $A \cup B$ is an information set for C, and $B \cup E$ is one for D. Write $e_{v_B} = \bigotimes_{j \in B} e_{v_j}$, and similarly for other coordinate sets; an empty tensor factor is 1.

We factor the operator into two encoders and the local maps on B. For outside labels $y \in X_D$ and $x \in X_C$, define

$$V_D e_y = q^{-b/2} \sum_{v \in D_y} \left(\bigotimes_{j \in A} A_j e_{v_j} \right) \otimes e_{v_B} \otimes e_{v_E},$$

$$V_C e_x = q^{-a/2} \sum_{u \in C_x} e_{u_A} \otimes e_{u_B} \otimes \left(\bigotimes_{j \in E} A_j^* e_{u_j} \right).$$

Retaining the information labels makes the energy calculation exact. Different $v \in D$ have different (v_B, v_E), so their terms are orthogonal; the columns for different y also have disjoint supports. Thus Lemma 7 gives

$$\|V_D\|^2 = \max_y q^{-b} \sum_{v \in D_y} \prod_{j \in A} \|A_j e_{v_j}\|_2^2 \le q^{-b} F(\ell - t, b).$$

Similarly, the retained (u_A, u_B) determine $u \in C$. Using row energies, which are the column energies of A_j^*, gives $\|V_C\|^2 \le q^{-a} F(\ell - s, a)$.

For each pair of outside labels, expansion of the two fiber sums gives

$$\mathcal{O}_{C,D} = V_C^* \left(I_A \otimes \bigotimes_{j \in B} A_j \otimes I_E \right) V_D.$$

The adjoint in the left encoder supplies the local entries on E; the right encoder supplies those on A. The intervening maps supply the entries on B, giving the displayed entry of (19). The middle map has norm at most $\nu^{|B|} = \nu^{s+t-\ell}$. Multiplying the three norm bounds proves the claim. □

For two four-coordinate planes, $s = t = 2$. If both outside projections leave a one-dimensional fiber, the bound is $\delta h/q$. Without these averages the same factorization gives δ^2. The ratio $h/(q\delta)$ measures the gain from retaining the free middle value. This gain is available because the incidence maps impose the average.

From fiber sums to operator norms. Another estimate uses the absolute entries of the raw kernel. For each left fiber, sum over that fiber and the whole right code; do the analogous sum for each right fiber. After normalization, these are row and column sums of a nonnegative majorant of the overlap. The usual row–column norm bound therefore applies.

Lemma 9 (A norm bound from fiber sums). *Let $W \geq 0$ be indexed by $C \times D$. Suppose these sets are partitioned into fibers C_x, D_y of sizes q^a, q^b, with normalized indicator maps J_C, J_D as above. If*

$$\sum_{u \in C_x,\, v \in D} W(u,v) \leq \rho_a, \qquad \sum_{u \in C,\, v \in D_y} W(u,v) \leq \kappa_b$$

for every outside label, then every matrix B with $|B(u,v)| \leq W(u,v)$ satisfies

$$\|J_C^* B J_D\| \leq q^{-(a+b)/2} \sqrt{\rho_a \kappa_b}. \tag{22}$$

Proof. The nonnegative matrix $\widetilde{W} = J_C^* W J_D$ has entries

$$\widetilde{W}(x,y) = q^{-(a+b)/2} \sum_{u \in C_x,\, v \in D_y} W(u,v).$$

Since the fibers partition the two sets, its row and column sums are at most $q^{-(a+b)/2}\rho_a$ and $q^{-(a+b)/2}\kappa_b$. Weighted Cauchy–Schwarz gives, for every vector f,

$$\begin{aligned} \|\widetilde{W} f\|_2^2 &\leq \sum_x \left(\sum_y \widetilde{W}_{xy} \right) \left(\sum_y \widetilde{W}_{xy} |f_y|^2 \right) \\ &\leq q^{-a-b} \rho_a \kappa_b \|f\|_2^2. \end{aligned}$$

Finally, $|J_C^* B J_D f| \leq \widetilde{W}|f|$ proves the same norm bound for signed or complex kernels. □

Singleton fibers give the usual row–column estimate. When a fiber is the whole code, its partial sum becomes the total kernel sum. The factor $q^{-(a+b)/2}$ always comes from the normalized indicator maps, with no assumption on the input law. Section 5.5 computes these partial sums for lines and planes.

5.5 Entry moments for thin codes

When the two code dimensions sum to less than the middle length, their information sets cannot cover all middle coordinates. For the line and plane cases below, fourth moments of local entries give the partial sums needed by Lemma 9. We keep the codes and normalized fiber maps of Section 5.4, with $q = |V|$.

A line has MDS dimension one and a plane dimension two. Every nonempty outside fiber is assumed to be an affine MDS code of its stated dimension, as in Section 5.4. In particular, each coordinate of a one-dimensional fiber is a bijection onto V; the line–plane estimate uses this property. The local maps have row and column squared norms at most δ, and total squared norm at most h. For the thin-code estimates below, assume the additional fourth-moment bounds, with nonnegative parameters,

$$\begin{aligned} \max_x \sum_y |A_j(x,y)|^4 \le r_4^2\delta, \qquad \max_y \sum_x |A_j(x,y)|^4 \le r_4^2\delta, \\ \sum_{x,y} |A_j(x,y)|^4 \le s_4^2 h. \end{aligned} \tag{23}$$

Different maps may be used at different coordinates, provided they obey the same inequalities. Ordinary entry bounds $|A_j(x,y)| \le p$ imply (23) with $r_4 = s_4 = p$. One-sided Fourier maps can instead be handled directly by their fourth moments.

Both estimates use the following finite-sum Hölder inequality. For nonnegative functions $f_1, \dots, f_\ell$ on a common finite set and $\eta, \tau \ge 0$, suppose

$$\sum f_j^2 \le \eta, \qquad \sum f_j^4 \le \tau^2\eta.$$

Then

$$\sum \prod_{j=1}^{\ell} f_j \le \eta\tau^{\ell-2} \qquad (\ell = 3, 4). \tag{24}$$

For three factors, use Hölder exponents $2, 4, 4$; for four, use $4, 4, 4, 4$. The resulting products of moment bounds are $\eta^{1/2}(\tau^2\eta)^{1/2}$ and $\tau^2\eta$, respectively.

Two lines.

Lemma 10 (Two lines). *Let C, D be lines in V^ℓ, $\ell \in \{3, 4\}$, with fiber dimensions $a, b \in \{0, 1\}$, and put $e = \ell - 2$. Set*

$$\rho_0 = \delta r_4^e, \qquad \rho_1 = h s_4^e.$$

Then

$$\|\mathcal{O}_{C,D}\| \le \sqrt{\frac{\rho_a \rho_b}{q^{a+b}}}. \tag{25}$$

Proof. Use the raw absolute kernel $K(u, v) = \prod_{j=1}^{\ell} |A_j(u_j, v_j)|$. Every coordinate of either line is a bijection of its parameter. We apply (24) to rows, columns and

the total sum. For a row or column, apply this with $(\eta, \tau) = (\delta, r_4)$. For the total sum, both line parameters vary and each coordinate pair ranges bijectively over V^2; apply it with $(\eta, \tau) = (h, s_4)$.

On either side, a fiber is a singleton or the entire line. Its partial sum is therefore bounded by ρ_0 or ρ_1, respectively. Lemma 9 gives (25). □

A line and a plane.

Lemma 11 (A line and a plane). *Let C be a line and D a plane in V^4. For outside-fiber dimensions $a \in \{0, 1\}$ and $b \in \{0, 1, 2\}$, set*

$$\rho_a = \delta^{2-a} h^a, \qquad (\kappa_0, \kappa_1, \kappa_2) = (\delta r_4^2, h s_4^2, \delta h).$$

Then

$$\|\mathcal{O}_{C,D}\| \le \sqrt{\frac{\rho_a \kappa_b}{q^{a+b}}}. \tag{26}$$

The adjoint version exchanges the two codes.

Proof. Put $K(u, v) = \prod_{j=1}^{4} |A_j(u_j, v_j)|$ on $C \times D$, and let $g_j(x) = \sum_y |A_j(x, y)|^2$ be the row energy of the jth local map. Thus $g_j \le \delta$ and $\sum_x g_j(x) \le h$.

Left fibers: a point or the whole line. For a fixed $u \in C$, Cauchy–Schwarz separates the two information pairs $\{1, 2\}$ and $\{3, 4\}$ of the plane:

$$\begin{aligned}\sum_{v \in D} K(u, v) &\le \prod_{B \in \{\{1,2\},\{3,4\}\}} \left(\sum_{v \in D} \prod_{j \in B} |A_j(u_j, v_j)|^2 \right)^{1/2} \\ &= \prod_{j=1}^{4} g_j(u_j)^{1/2} \le \delta^2.\end{aligned}$$

The equality uses bijectivity of each information-pair projection. To sum this bound over u, cap the last two factors by δ and use Cauchy–Schwarz on the first two. Every coordinate of the line ranges bijectively over V, so

$$\sum_{u \in C, v \in D} K(u, v) \le \delta \sum_{u \in C} \sqrt{g_1(u_1) g_2(u_2)} \le \delta \sqrt{\sum_x g_1(x) \sum_x g_2(x)} \le \delta h.$$

The row and total estimates are precisely $\rho_0 = \delta^2$ and $\rho_1 = \delta h$.

Right fibers: a point, a line or the whole plane. For a singleton right fiber, each coordinate of the left line ranges bijectively over V. Apply (24) with four factors and $(\eta, \tau) = (\delta, r_4)$ to obtain $\kappa_0 = \delta r_4^2$. For a one-dimensional right fiber D_y, each coordinate pair ranges bijectively over V^2 on $C \times D_y$. The same inequality with $(\eta, \tau) = (h, s_4)$ gives $\kappa_1 = h s_4^2$. For the full right plane, the same partial sum is the total sum, already bounded by $\kappa_2 = \delta h$. Lemma 9 now gives (26) for every fiber choice. Taking the adjoint exchanges the two codes and their fiber dimensions. □

Section 8 applies these estimates to the AES graphs, including the Fourier orientations and activity-frame coefficients. The required local moments are established first for bytes in Section 6 and then for words in Section 7.

6 Byte Estimates and Diffusion Geometry

The word channel has two byte layers separated by MixColumns. This section derives two kinds of estimates for its byte-degree blocks: operator norms, which control the response to an arbitrary input vector, and block energies, the squared Hilbert–Schmidt norms that sum the squared responses over an orthonormal input basis. Section 7.3 assembles the operator bounds into a bound for the whole word channel's residual part. Section 7.4 combines them with the block energies to bound its Gram moment and total energy. The byte contraction and Gram data, together with the codes imposed by diffusion, supply the inputs to these block calculations.

Substitution already gives strong contraction at high residual degrees. The blocks coupling activity to degrees two and three need more care: their responses can overlap, and the zero/uniform activity frame is not orthogonal. We express these responses as shortened MDS codes and compare their Gram matrix with the activity-frame metric. Coincident lines control degree two; signed line–plane overlaps give the sharper degree-three bounds. The remaining blocks use row counts and the general code-energy estimate in Section 6.6.

6.1 The byte channel and its Gram data

The byte estimates used throughout this section are

$$s = 9/128, \qquad r_b = 65/8192, \qquad p_b = 1/64. \tag{27}$$

Here s bounds the norm after removing the zero and uniform-nonzero directions, r_b bounds the sum of squared probabilities in a nonzero column, and p_b bounds its largest entry. Thus s controls arbitrary signed inputs, whereas r_b and p_b describe a fixed input difference. We need these same parameters for the channel, its transpose, and their additive Fourier transforms; each orientation is justified below.

All three come from the inversion underlying the AES S-box. In the byte encoding of Section 2, AES uses

$$s_{\mathrm{AES}}(x) = v \oplus \mathrm{rot}_1(v) \oplus \cdots \oplus \mathrm{rot}_4(v) \oplus \mathtt{63}_{16}, \qquad v = x^{-1}, \quad 0^{-1} = 0,$$

where rot_j is cyclic left rotation by j bits [47]. The lemma below first counts the possible output differences of inversion. The resulting formula depends only on the product of the input and output differences. This makes a multiplicative Fourier basis suitable for bounding the norm. The invertible affine output map then transfers the bounds to AES. Nyberg [48] established the differential uniformity of inversion; Daemen and Rijmen [20] give its difference counts in the AES analysis. Propagation characteristics of inversion and their relation to Kloosterman sums were studied by Charpin, Helleseth, and Zinoviev [12]. The character-sum argument below follows BLS [4]; their stated lemma assumes odd

field degree κ, whereas AES uses $\kappa = 8$. We give the calculation for both parities below.

Lemma 12 (Residual contraction of inversion). *Let $m = 2^\kappa$ with $\kappa \geq 1$, and let D_{inv} be the difference channel of inversion on $\mathbb{F}_m$, with $0^{-1} = 0$. Let P_b project onto the zero and uniform-nonzero directions. Then*

$$\|D_{\mathrm{inv}} - P_b\| \leq \frac{\sqrt{m}+2}{m}.$$

For even κ, each row and column of the nonzero block has one entry $4/m$, $m/2-2$ entries $2/m$, and $m/2$ zeros. Both conclusions persist after any invertible affine transformation of the output.

Proof. Counting differences. Let $\mathrm{tr}(x) = x + x^2 + \cdots + x^{2^{\kappa-1}}$ be the absolute field trace from $\mathbb{F}_m$ to $\mathbb{F}_2$. This field trace is distinct from the matrix trace Tr used for Gram moments. For nonzero differences a, b, put $c = ab$ and write the first input as at. Outside $t \in \{0, 1\}$, the difference equation is $t^2 + t = c^{-1}$. The map $t \mapsto t^2 + t$ has kernel $\{0, 1\}$ and image equal to the trace-zero subspace. Indeed, trace vanishes on its image, and both spaces have size $m/2$: the trace polynomial is nonzero of degree $m/2 < m$. The two excluded inputs both give $c = 1$. Hence

$$D_{\mathrm{inv}}(b, a) = d(ab), \qquad d(c) = \frac{2}{m}\left(\mathbf{1}_{\mathrm{tr}(c^{-1})=0} + \mathbf{1}_{c=1}\right).$$

For even κ, $\mathrm{tr}(1) = 0$; counting the nonzero trace-zero elements gives the stated histogram.

Bounding residual variation. Write $G = \mathbb{F}_m^*$ for the nonzero field elements under multiplication. A multiplicative character is a function $\chi : G \to \mathbb{C}$ with $|\chi(x)| = 1$ and $\chi(xy) = \chi(x)\chi(y)$. These are the Fourier modes of the multiplicative group; dividing their vectors by $\sqrt{m-1}$ gives an orthonormal basis. The constant character describes the uniform direction, and all other characters have sum zero. A change of variable $c = ab$ shows that D_{inv} sends the χ vector to λ_χ times the $\bar{\chi}$ vector, where $\lambda_\chi = \sum_c d(c)\chi(c)$. Thus it suffices to bound these scalar multipliers. The constant character has $\lambda_1 = 1$. For nonconstant χ, put $\psi(t) = (-1)^{\mathrm{tr}(t)}$ and $g(\chi) = \sum_{t \in G} \chi(t)\psi(t)$. Then

$$\lambda_\chi = \frac{g(\bar{\chi}) + 2}{m}, \qquad |g(\chi)|^2 = \sum_{t \in G} \chi(t) \sum_{v \in G} \psi((t-1)v) = m.$$

For the last equality, the inner sum is $m - 1$ at $t = 1$ and -1 otherwise, while $\sum_t \chi(t) = 0$. Thus $|\lambda_\chi| \leq (\sqrt{m}+2)/m$. Since character conjugation permutes an orthonormal basis, this bounds the residual operator norm. At $m = 2$ the residual space is zero.

Transferring the bound to an affine output map. If the output is transformed by $x \mapsto Ax + c$ with A invertible and $\mathbb{F}_2$-linear, the difference channel becomes $\Pi_A D_{\mathrm{inv}}$. The permutation Π_A relabels output differences, preserving lengths and the two directions selected by P_b. This proves both invariances. □

The AES linear part is $I + R + R^2 + R^3 + R^4$, where R rotates a byte by one bit. It is invertible: its polynomial has value one at $X = 1$, so it is coprime to $X^8 - 1 = (X + 1)^8$, and $R^8 = I$. Lemma 12 at $m = 256$ gives $s = 9/128$. The histogram gives $r_b = (4^2 + 126 \cdot 2^2)/256^2$ for every nonzero row or column, and $p_b = 4/256$. This proves (27), the standard AES byte parameters reported by Dinur [21].

Write D for the AES byte difference channel and $R_b = D - P_b$ for its residual part, where P_b projects onto the zero and uniform-nonzero directions as in Lemma 12.

The same bounds in the Fourier basis. The Fourier orientation is needed for the dual-code argument in Proposition 2. This is the additive Fourier transform on byte differences, rather than the multiplicative basis used inside the inversion proof. Write $\mathcal{F}_b$ for its matrix, with entry $256^{-1/2}(-1)^{\mathrm{tr}(uv)}$ at (u, v). Define the ordinary Walsh table

$$W_{uv} = \sum_{x \in \mathbb{F}_{256}} (-1)^{\mathrm{tr}(ux + v s_{\mathrm{AES}}(x))}.$$

The differential–linear Fourier relation of Chabaud and Vaudenay [11] expresses the transformed DDT through squared Walsh coefficients. Expanding the DDT over pairs $(x, x + a)$ gives, with the corresponding index convention,

$$(\mathcal{F}_b D \mathcal{F}_b)_{v,u} = W_{uv}^2/256^2 =: \widehat{D}_{v,u}.$$

Indeed, the sum over the input difference changes variables to two independent byte inputs, and the two character sums multiply. Thus $\widehat{D}$ is nonnegative and fixes the same zero and uniform-nonzero directions as D. The required row and column data follow from symmetry and orthogonal invariance, without summing the Walsh table. For inversion, if $u, v \neq 0$, the substitution $x = vz$ shows

$$W_{uv}^{\mathrm{inv}} = \sum_z (-1)^{\mathrm{tr}(uvz + z^{-1})}.$$

Thus every nonzero row and every nonzero column has the same multiset of squared values, since multiplication by a nonzero field element is a bijection. An affine output map only relabels the output mask and changes the signs of Walsh entries; it preserves these squared multisets.

Let $\widehat{D} = \mathcal{F}_b D \mathcal{F}_b$. Fourier conjugation preserves the squared Hilbert–Schmidt norm. The isolated zero entry contributes one in each basis, and the 255 nonzero columns of D each have collision r_b. If $\widehat{r}$ denotes the common nonzero row or column collision of $\widehat{D}$, then

$$1 + 255\widehat{r} = \left\| \widehat{D} \right\|_{\mathrm{HS}}^2 = \|D\|_{\mathrm{HS}}^2 = 1 + 255 r_b, \qquad \widehat{r} = r_b.$$

The row and column sums equal one: $\widehat{D}$ is nonnegative and $\mathcal{F}_b$ exchanges the zero direction with the full constant direction, both fixed by D. This also proves that zero remains isolated.

For the largest entry we use the classical nonlinearity bound for inversion recalled by Nyberg [48], who attributes it to Carlitz and Uchiyama [10]: every nonzero component of $x \mapsto x^{-1}$ on $\mathbb{F}_{2^\kappa}$ has distance at least $2^{\kappa-1} - 2^{\kappa/2}$ from every affine Boolean function. At $\kappa = 8$ this is 112, equivalent to $|W_{uv}| \le 256 - 2 \cdot 112 = 32$ for nonzero masks. Affine output changes preserve this bound. We use this nonlinearity result only for the largest-entry estimate; the collision equality above follows independently from norm invariance. Consequently, for any fixed nonzero row or column mask,

$$|W_{uv}| \le 32, \qquad \sum W_{uv}^2 = 256^2, \qquad \sum W_{uv}^4 = 256^4 r_b = 65 \cdot 2^{19}. \tag{28}$$

The probability bound for $\widehat{D}$ is therefore $32^2/256^2 = 1/64$. Together with the norm invariance, this proves the common parameters s, r_b, p_b for $D, D^*, \widehat{D}, \widehat{D}^*$. We can therefore use the same three byte parameters in all four word-channel orientations of Lemma 21. Lemma 13 supplies the additional Gram-overlap estimates.

Byte Gram data from character sums. The inversion estimates above [48] control individual byte response lengths. The code estimates also depend on overlaps between responses. For inversion, each overlap is determined by two field traces. The resulting Gram kernel is unchanged by Fourier conjugation, which gives common overlap and moment bounds in all four orientations.

Put $m = 256$, $\psi(x) = (-1)^{\mathrm{tr}(x)}$, and

$$\mathcal{K}(z) = \sum_{x \neq 0} \psi(x + z/x) \qquad (z \neq 0).$$

Define the common denominator and Gram bounds by

$$L_b = 4177920, \qquad \Delta_b = \frac{65}{8192} - \frac{1}{255} = \frac{16766}{L_b},$$
$$\gamma_{b,-} = \frac{319}{L_b}, \qquad \gamma_{b,+} = \frac{191}{L_b}, \qquad T_b = \frac{571447}{2^{27}}.$$

The diagonal value Δ_b is a residual column's squared length. Off-diagonal overlaps lie between $-\gamma_{b,-}$ and $\gamma_{b,+}$. The second Gram moment is bounded by T_b. These are byte quantities; the word second moment τ^2 is derived from them in Section 7.4. The exact kernel and row counts also supply the cubic and mixed Fourier estimates of Lemma 18.

Lemma 13 (Byte Gram data). *Let the family of four residual orientations be*

$$\mathcal{T}_b = \{R_b, R_b^*, \mathcal{F}_b R_b \mathcal{F}_b, \mathcal{F}_b R_b^* \mathcal{F}_b\}.$$

For $R \in \mathcal{T}_b$, put $G = R^* R$. Then

$$G_{0x} = G_{x0} = 0, \quad G_{xx} = \Delta_b \ (x \neq 0), \quad -\gamma_{b,-} \le G_{xy} \le \gamma_{b,+} \ (x \neq y),$$
$$\mathrm{Tr}(G^2) \le T_b.$$

The four corresponding byte channels have collision $65/8192$ *at every nonzero column and point probabilities at most* $1/64$.

Let D_{inv} *be the inversion difference channel and* $G_* = (D_{\mathrm{inv}} - P_b)^2$. *Every such* G *is a conjugate of* G_* *by an invertible binary-linear relabeling. Moreover,* $\mathcal{F}_b G_* \mathcal{F}_b = G_*$, *and on nonzero labels*

$$\begin{aligned} (G_*)_{xy} &= a_b \mathbf{1}_{x=y} + b_b\{\psi(x/y) + \psi(y/x)\} + c_b, \\ a_b &= \frac{m+4}{m^2}, \quad b_b = \frac{2}{m^2}, \quad c_b = -\frac{1}{m(m-1)}. \end{aligned} \tag{29}$$

Every nonzero row of G_* *has off-diagonal numerators* $191, -64, -319$ *over denominator* L_b, *with multiplicities* $h, 252-2h, h+2$, *respectively, for a common integer* $0 \le h \le 70$. *Finally,* $-33 \le \mathcal{K}(z) \le 31$ *for* $z \neq 0$.

Proof. The inversion channel is symmetric. We compute its Gram kernel and row multiplicities first, then use coordinate relabelings to obtain the four AES orientations.

The overlap kernel. For distinct nonzero x, y, the trace conditions in the product of its columns become two distinct nonzero binary linear forms in the reciprocal output label. They are independent over $\mathbb{F}_2$, so their common kernel has $m/4$ elements, of which $m/4-1$ are nonzero. The two exceptional inversion differences contribute the remaining terms. Expanding the formula in Lemma 12 gives

$$(G_*)_{xy} = \frac{4}{m^2}\left(\mathbf{1}_{\mathrm{tr}(x/y)=0} + \mathbf{1}_{\mathrm{tr}(y/x)=0} - 1\right) - \frac{1}{m(m-1)}.$$

Its diagonal is Δ_b. Replacing each trace-zero indicator by $(1+\psi)/2$ therefore gives the kernel (29). The three off-diagonal values are $191, -64, -319$, divided by 4177920, proving their stated bounds.

The Walsh bound (28), obtained from the classical nonlinearity estimate for inversion [48], also controls the character sum $\mathcal{K}$ defined above. The inversion Walsh sum with masks $(1, z)$ is $1 + \mathcal{K}(z)$: the term at zero is one, and the other terms are the displayed sum. Thus $|1 + \mathcal{K}(z)| \le 32$, and hence

$$-33 \le \mathcal{K}(z) \le 31.$$

No sharper lower endpoint is needed below.

Row multiplicities and the second moment. For $z = 1$, let h count the labels $r \neq 0, 1$ with $\mathrm{tr}(r) = \mathrm{tr}(1/r) = 0$. Among these labels the two single-zero cases each have size $126 - h$, and the no-zero case has size $h+2$. For each fixed nonzero row label, the other nonzero column labels give each ratio in $\mathbb{F}_m \setminus \{0, 1\}$ exactly once. The three Gram numerators therefore have multiplicities $h, 252-2h, h+2$ in every nonzero row. Thus $\mathcal{K}(1) = 4h - 249$ and $h \le 70$. The row-square sum of G_* is increasing in h, because $191^2 - 2 \cdot 64^2 + 319^2 > 0$. It follows that

$$\mathrm{Tr}(G_*^2) \le \frac{255}{4177920^2}\{16766^2 + 70 \cdot 191^2 + 112 \cdot 64^2 + 72 \cdot 319^2\} = T_b. \tag{30}$$

The four orientations. If $\Pi_r e_x = e_{rx}$, with $\Pi_r e_0 = e_0$, then the trace-pairing Fourier matrix satisfies $\mathcal{F}_b \Pi_r \mathcal{F}_b = \Pi_{r^{-1}}$. The Gram kernel is a symmetric multiplicative convolution and has zero row sums. Thus its expansion as a linear combination of the Π_r has coefficients whose sum is zero, and coefficients at r, r^{-1} agree. Consequently

$$\mathcal{F}_b G_* \mathcal{F}_b = G_*. \tag{31}$$

Let A be the invertible binary linear part of the AES output affine map. Then $R_b = \Pi_A(D_{\mathrm{inv}} - P_b)$, so the two ordinary Grams are G_* and $\Pi_A G_* \Pi_A^*$. Fourier conjugation sends Π_A to $\Pi_{(A^\dagger)^{-1}}$, where $A^\dagger$ is the trace-pairing adjoint, defined by $\mathrm{tr}(uAx) = \mathrm{tr}((A^\dagger u)x)$. The other two Grams are therefore also permutation conjugates of G_*. This proves all diagonal, off-diagonal, moment and relabeling assertions in all four orientations.

The collision and point-probability claims for the four byte channels were established in (27) and (28) above. Equivalently, the collision is $\Delta_b + 1/255 = 65/8192$ from the diagonal Gram calculation. This completes the stated estimates in every orientation. □

Byte autocorrelation. Charpin, Helleseth, and Zinoviev [12] prove that every nonzero component of inversion in even dimension $n \geq 3$ has maximum absolute autocorrelation $2^{(n+2)/2}$. At $n = 8$ this is 32. Invertible affine changes preserve that maximum, so the statement applies to the AES S-box and its inverse. In matrix notation, with $N_b = 256D$ and $W_b = 16\mathcal{F}_b$, it gives

$$|(N_b W_b)_{xy}|,\ |(N_b^* W_b)_{xy}| \leq 32 \qquad (x, y \neq 0). \tag{32}$$

6.2 AES diffusion and the activity block

We first separate the word channel by the number of residual byte coordinates. The block with no residual coordinates depends only on support counts in the MDS graph. Its calculation also applies to the word-level diffusion used later in the recurrence.

Decomposing the word channel. Let F_i project onto byte residual degree i: exactly i coordinates use $I - P_b$ and the other coordinates use P_b. These are byte degrees within one word; the projections P_j in Section 4.1 instead count residual words in the full state. The range of F_0 records all sixteen byte activity patterns. It contains the two word invariants selected by P, so its centered part $F_0 - P$ has dimension fourteen. Put

$$A = D^{\otimes 4}, \qquad K = A\Pi_M A, \qquad R_{ij} = F_i R F_j \quad (0 \leq i, j \leq 4).$$

Substitution preserves these degrees, fixes activity, and contracts degree i by at most s^i. Thus a raw diffusion block from degree j to degree i receives a factor s^{i+j}. We will retain the actual byte Grams to sharpen the bounds on blocks coupling activity and residual degrees and on the second moment. This organization follows the residual-degree method of BLS [4]. The byte parameters $s, \Delta_b, \gamma_{b,-}, T_b$ are defined in Section 6.1.

One activity calculation at both scales. The MDS support counts depend only on the alphabet size, so one calculation gives the activity norm and energy at both byte and word scale. Grouping support patterns by size reduces the operator to four dimensions. Let L_m be an additive four-coordinate MDS map on an alphabet of size m, and put $Q_m = m - 1$. Any four coordinates of $(L_m x, x)$ determine x. Inclusion–exclusion therefore gives the number of graph points on each specified exact support of size $0 \le r \le 8$:

$$N_r(m) = \sum_{t=0}^{r} (-1)^{r-t} \binom{r}{t} m^{\max(t-4,0)}. \tag{33}$$

Indeed, a prescribed set of t possibly nonzero positions leaves $\max(t-4, 0)$ free symbols; inclusion–exclusion removes the points with an additional zero. Thus $N_0 = 1$ and

$$\begin{aligned} N_1 = N_2 = N_3 = N_4 = 0, \qquad & N_5 = Q_m, \\ & N_6 = Q_m(m-5), \\ & N_7 = Q_m(m^2 - 6m + 15), \\ & N_8 = Q_m(m^3 - 7m^2 + 21m - 35). \end{aligned} \tag{34}$$

Let $P_{\mathrm{act},m}$ average within exact support patterns and let $P_{\mathrm{inv},m}$ project onto the zero state and the uniform nonzero vector. Put $T_m = P_{\mathrm{act},m} \Pi_{L_m} P_{\mathrm{act},m} - P_{\mathrm{inv},m}$. As in the support-count method of BLS [4], the compressed kernel depends only on the two support sizes. Differences between patterns of the same size are killed.

On the four nonzero support weights, use coordinates recording total mass. The matrix C_m of $P_{\mathrm{act},m} \Pi_{L_m} P_{\mathrm{act},m}$ has entries

$$(C_m)_{ab} = \binom{4}{a} \frac{N_{a+b}(m)}{Q_m^b} \qquad (1 \le a, b \le 4).$$

With $(\mathsf{W}_m)_{aa} = \sqrt{\binom{4}{a} Q_m^a}$, the matrix $\mathsf{W}_m^{-1} C_m \mathsf{W}_m$ is symmetric and acts on orthonormal weight indicators. Its centered eigenvalues therefore determine the Euclidean norm of T_m.

The uniform nonzero distribution gives eigenvalue 1. A second eigenvector follows by substituting (34):

$$C_m \begin{pmatrix} 1 \\ -3 \\ 3 - m \\ Q_m \end{pmatrix} = \frac{4}{Q_m^2} \begin{pmatrix} 1 \\ -3 \\ 3 - m \\ Q_m \end{pmatrix}.$$

Its coordinates sum to zero, so it is already centered. To find the two remaining eigenvalues, note that entries with $a + b \le 4$ vanish. Reversing the four columns makes the matrix triangular, with diagonal $4/Q_m^3, 6/Q_m^2, 4/Q_m, 1$; the reversal has positive sign. Reading the original diagonal entries at support weights three and four also gives

$$\det C_m = \frac{96}{Q_m^6}, \qquad \operatorname{Tr} C_m = 1 - \frac{6}{Q_m^2} - \frac{20}{Q_m^3}.$$

After removing eigenvalues 1 and $4/Q_m^2$, the remaining two have sum $-10(m+1)/Q_m^3$ and product $24/Q_m^4$. They are therefore the roots of

$$\lambda^2 + \frac{10(m+1)}{Q_m^3}\lambda + \frac{24}{Q_m^4} = 0.$$

The projection $P_{\mathrm{inv},m}$ removes the uniform nonzero direction. The quadratic roots are negative; the larger absolute root dominates $4/Q_m^2$. Hence

$$\|T_m\| = \frac{5(m+1) + \sqrt{m^2+98m+1}}{Q_m^3} < \frac{6m+54}{Q_m^3}, \tag{35}$$

where the last inequality uses $\sqrt{m^2+98m+1} < m+49$. Summing the squares of the three centered eigenvalues also gives

$$\|T_m\|_{\mathrm{HS}}^2 = \frac{16}{Q_m^4} + \frac{100(m+1)^2}{Q_m^6} - \frac{48}{Q_m^4} = \frac{68m^2+264m+68}{Q_m^6}. \tag{36}$$

The substitutions and their adjoints fix activity, so $R_{00} = T_{256}$, and $P_A\mathcal{D}_4P_A = T_q$ on centered activity vectors. Substitution at $m = 256$ and $m = q = 2^{32}$ yields

$$\|R_{00}\| < 6.3/256^2, \qquad q\|R_{00}\|_{\mathrm{HS}}^2 < 71, \tag{37}$$

$$\|P_A\mathcal{D}_4P_A\| \le 6.01/q^2. \tag{38}$$

For the two norm comparisons it suffices to check $10\cdot 256^2(6\cdot 256+54) < 63\cdot 255^3$ and $100q^2(6q+54) < 601(q-1)^3$; the trace comparison is $q(68\cdot 256^2 + 264\cdot 256 + 68) < 71\cdot 255^6$. All use integers. Forward and inverse diffusion have the same support counts, so the bounds hold in both directions and for arbitrary inputs. □

6.3 Activity frames and contracted graph responses

For the blocks with residual coordinates, exact support counts alone do not describe the responses. Zero and uniform activity vectors are more useful: they fix a graph coordinate or sum over it, producing the shortened codes to which Section 5 applies. Their nonorthogonality must be included when estimating norms.

Use the two unit vectors e_0 and $u_0 = m^{-1/2}\mathbf{1}$ for activity. They are not orthogonal: their inner product is $\xi = m^{-1/2}$. For r activity coordinates, $\mathcal{V}_r$ maps a coefficient vector indexed by subsets $A \subseteq [r]$ to a linear combination of product vectors: its column at A uses u_0 in positions A and e_0 elsewhere. For instance, with two activity coordinates the four columns are $e_0\otimes e_0$, $u_0\otimes e_0$, $e_0\otimes u_0$, and $u_0\otimes u_0$. Here "uniform" includes zero: these are not exact nonzero-pattern indicators. The frame Gram matrix $\mathsf{G}_r = \mathcal{V}_r^*\mathcal{V}_r$ and inverse have the explicit entries

$$(\mathsf{G}_r)_{AB} = \xi^{|A\triangle B|}, \qquad (\mathsf{G}_r^{-1})_{AB} = c_0^r(-\xi)^{|A\triangle B|}, \qquad c_0 = \frac{m}{m-1}. \tag{39}$$

They are tensor powers of the corresponding two-by-two matrices. For coefficients a, the synthesized input has squared norm $a^*\mathsf{G}_4 a$; its response under a block B has squared norm $a^*\mathcal{V}_4^* B^* B \mathcal{V}_4 a$. We must compare these two quadratic forms, since coefficient length alone is not input length. Multiplying the compressed Gram by G_4^{-1} makes that comparison. The result is similar to the restriction of B^*B to activity, so its largest eigenvalue is the desired squared norm. An absolute row-sum bound, including one after positive diagonal scaling, bounds this eigenvalue.

We use the block-norm comparison and the activity-frame singular-value bound of BLS [4] as established ingredients. Their frame consists of exactly the zero and normalized uniform vectors defined above; the remaining factors form orthonormal residual bases. The next lemma states the specialization needed for our scalar arrays. Its proof only combines those ingredients and counts the orthogonal choices of residual positions.

Lemma 14 (Activity arrays). *Consider an operator block B with a input and b output activity coordinates, supported on those activity subspaces: $B = P_{\mathrm{act,out}} B P_{\mathrm{act,in}}$. Here each P_{act} is the activity projection on the specified coordinates and the identity on the residual coordinates. Leave all residual coordinates unchanged and synthesize the activity coordinates with $\mathcal{V}_a, \mathcal{V}_b$. Suppose a nonnegative scalar matrix F bounds, entry by entry, the operator norms of the blocks of $\widetilde{B} = \mathcal{V}_b^* B \mathcal{V}_a$. Then*

$$\|B\| \le (1 - m^{-1/2})^{-(a+b)/2} \, \|F\| \, . \tag{40}$$

If the same bound holds for each choice of i input and j output residual positions among four coordinates, the complete degree block is bounded by an additional factor $\sqrt{\binom{4}{i}\binom{4}{j}}$.

Proof. The block-norm comparison of BLS [4] gives $\left\|\widetilde{B}\right\| \le \|F\|$: the triangle inequality bounds each output length by F applied to the input lengths. Their activity-frame bound gives $\|\mathcal{V}_a^{-1}\| = (1 - m^{-1/2})^{-a/2}$ on its image, and similarly for $\mathcal{V}_b$. Applying these two bounds to $\widetilde{B} = \mathcal{V}_b^* B \mathcal{V}_a$ proves (40). Distinct residual-position spaces are orthogonal. Their scalar comparison matrix has every entry equal to the common bound, and its norm is that bound times $\sqrt{\binom{4}{i}\binom{4}{j}}$. □

In this MDS-code method of BLS [4], many activity choices have the same response pattern. The scalar array retains that symmetry; Section 6.6 uses it to compute the array norms by grouping subsets of equal size.

Eliminating the empty and full frame coordinates. The degree-two and degree-three blocks annihilate two frame vectors. Eliminating their coefficients sharpens the metric comparison: those coefficients affect the input length even though they contribute no response. Use $m = 256$ and the byte frame above. Write $P_{e,b} = I - P_b$ for the byte residual projection. For $A \subseteq [4]$, let b_A use the uniform frame vector exactly in positions A, and write $\mathcal{V} = \mathcal{V}_4$ for the synthesis map with

metric G_4 from (39). For $B_j = F_j \Pi_M F_0$ of positive residual degree, the columns at $A = \varnothing, [4]$ vanish. We minimize the input norm over these two coefficients while keeping the other fourteen fixed. The resulting effective metric has, for $1 \le |A|, |B| \le 3$, with $w = |A|$, $v = |B|$, and $h = |A \cap B|$,

$$(\mathsf{G}_{\text{eff}})_{AB} = \xi^{w+v-2h} - \frac{\xi^{w+v} + \xi^{8-w-v} - \xi^{8+w-v} - \xi^{8-w+v}}{1-\xi^8}. \tag{41}$$

This is the Schur complement of the endpoint metric $\left(\begin{smallmatrix} 1 & \xi^4 \\ \xi^4 & 1 \end{smallmatrix}\right)$. Consequently, $\mathcal{V}^* B_j^* B_j \mathcal{V} \prec t^2 \mathsf{G}_{\text{eff}}$ on the fourteen retained columns proves $\|B_j\| < t$. The minimization ensures the bound holds for every choice of the omitted coefficients.

Contracting a graph against frame vectors. We use the graph-shortening and residual-projection construction of BLS [4], with the normalization made explicit below. The graph consists of the four input coordinates and their four outputs under MixColumns. Contracting an activity position means fixing it to zero or summing it against the normalized uniform vector. Mark j output positions as residual. Among the other $8-j$ graph positions, let $\mathcal{U}$ be the positions contracted against the uniform frame vector; all remaining positions are fixed to zero. The MDS vanishing lemma of BLS [4] applies with four input coordinates, j residual positions and $8 - j - |\mathcal{U}|$ zero positions. It leaves only $0 < d = j + |\mathcal{U}| - 4 < j$. In this surviving case, projection onto the marked positions is injective and gives an MDS code $\mathcal{C}_{\mathcal{U}}$. Its contracted vector is

$$v_{\mathcal{U}} = m^{(j-4)/2} P_{e,b}^{\otimes j} c_{\mathcal{C}_{\mathcal{U}}}, \qquad c_{\mathcal{C}_{\mathcal{U}}} = m^{-d/2} \mathbf{1}_{\mathcal{C}_{\mathcal{U}}}. \tag{42}$$

The normalization follows directly: the $|\mathcal{U}|$ uniform factors contribute $m^{-|\mathcal{U}|/2}$, while converting the code sum into a unit indicator contributes $m^{d/2}$. To recover the squared norm in the unmarked output activity positions, contract their frame indices against the inverse metric. For h such positions this inverse has entries $c_0^h(-\xi)^{|S \triangle T|}$. These inverse-metric coefficients enter both the degree-two and degree-three estimates.

Damped responses on the same frame. The same contraction applies when residual coordinates occur on either side of the graph and carry byte residual maps. Mark k coordinates and let t activity choices be uniform. Fixing the other $8 - k - t$ coordinates to zero, or *shortening* the code, leaves dimension $d = k - 4 + t$. As in (42), only $0 < d < k$ contributes. Replacing the residual projections by byte residual maps gives

$$v_\alpha = m^{(k-4)/2} (R_1 \otimes \cdots \otimes R_k) c_{\mathcal{C}_\alpha}. \tag{43}$$

Here $c_{\mathcal{C}}$ is a unit code indicator; input residual legs use R_b^* and output residual legs use R_b. The factors are fixed within a marked position pattern. This is just the vectorization of that matrix block: a matrix's Hilbert–Schmidt norm is the Euclidean norm of its entries.

6.4 Degree-two coupling

Our goal here is

$$\|F_2 \Pi_M F_0\| < 71/5000.$$

The byte channels will then multiply this bound by s^2. The difficulty is that different activity choices can produce the same line on the two residual coordinates, so their responses cannot be treated as orthogonal. Here $m = 256$, $Q_b = m - 1 = 255$, $\xi = m^{-1/2} = 1/16$, and $c_0 = m/(m-1) = 256/255$.

Degree two: coincident lines. Only $|\mathcal{U}| = 3$ survives in (42); its code is a line in the two marked positions. Write v_ℓ for the response of a line with nonzero slope ℓ. The residual projection kernel used by BLS [4] is, on nonzero labels, $P_{e,b}(x, y) = \mathbf{1}_{x=y} - 1/Q_b$; zero labels contribute nothing. Multiplying these kernels gives

$$m^4 \langle v_\ell, v_{\ell'} \rangle = m Q_b \mathbf{1}_{\ell=\ell'} - m. \tag{44}$$

Thus the overlap depends only on whether the slopes coincide. Number the four inputs $0, 1, 2, 3$ and the outputs $4, 5, 6, 7$. For example, mark outputs $4, 5$. The three choices

$$\mathcal{U} = \{0, 1, 6\}, \{1, 2, 7\}, \{3, 6, 7\},$$

with all other unmarked graph coordinates fixed to zero give the same line $\mathcal{C}_\mathcal{U} = \{(z, z) : z \in \mathbb{F}_{256}\}$. They therefore produce the same residual response. The symbolic calculation below finds all such coincidences; the norm bound then uses only their equality classes.

A symbolic classification of the slopes. Replace the AES byte 2 by an indeterminate t, so that byte 3 becomes $t + 1$. Over $\mathbb{F}_2(t)$ put

$$M(t) = \begin{pmatrix} t & t+1 & 1 & 1 \\ 1 & t & t+1 & 1 \\ 1 & 1 & t & t+1 \\ t+1 & 1 & 1 & t \end{pmatrix}.$$

Cyclic symmetry leaves the marked output pairs 45 and 46. For pair 45, the unmarked positions are $\{0, 1, 2, 3, 6, 7\}$; for pair 46, they are $\{0, 1, 2, 3, 5, 7\}$. The table (45) uses these graph labels directly: for example, 016 means that input positions $0, 1$ and output position 6 use the uniform vector. The other unmarked positions are fixed to zero.

Lemma 15 (Symbolic slope classes). *Let α lie in a field of characteristic two and have algebraic degree greater than six over $\mathbb{F}_2$. The graph of $M(\alpha)$ is MDS. Its slope classes for each marked pair are exactly the classes in* (45)*, evaluated at $t = \alpha$. In particular, this classification applies to AES.*

Proof. Use the five polynomial factors

$$a = t, \quad b = t + 1, \quad c = t^2 + t + 1, \quad d = t^3 + t + 1, \quad e = t^3 + t^2 + 1.$$

They are distinct monic irreducibles: c, d, e have no root in $\mathbb{F}_2$, which suffices in degrees two and three. Direct multiplication gives

$$\widetilde{M} = M(t)^{-1} = \mathrm{circ}(ac, d, e, bc),$$

where each row is the preceding row shifted one place to the right. For example, the first diagonal product is $a^2c + b^2c + d + e = c + (t^2 + t) = 1$; the other three entries in that product row vanish. For rows u, v and columns i, j, write $\Delta_{uv}(ij) = M_{ui}M_{vj} + M_{uj}M_{vi}$. The two representative rows of minors are

ij	01	02	03	12	13	23
$\Delta_{01}(ij)$	c	c	b	c	1	a
$\Delta_{02}(ij)$	1	b^2	c	c	a^2	1

.

Rotation supplies all other two-by-two minors. The inverse identity implies $\det M(t) = 1$: its determinant has a polynomial inverse and therefore is a nonzero constant. The three-by-three minors are the entries of $\widetilde{M}$. Thus every square minor is a nonzero polynomial of degree at most three. They stay nonzero at α, proving MDS.

Every slope now follows from one of three formulas. Let u, v be the marked matrix rows and J the uniform input positions. If $J = \{i\}$, both unmarked outputs are uniform and the slope is M_{vi}/M_{ui}. If $J = \{i, j\}$, exactly one unmarked output is zero; call its row w. Its equation has kernel vector (M_{wj}, M_{wi}), so the slope is $\Delta_{vw}(ij)/\Delta_{uw}(ij)$. Finally, if $J = \{0, 1, 2, 3\} \setminus \{i\}$, both unmarked outputs are zero. Writing the input as $\widetilde{M}y$, its ith coordinate gives slope $\widetilde{M}_{iu}/\widetilde{M}_{iv}$. These cases cover $4 + 12 + 4 = 20$ choices per marked pair.

Substitution gives the complete table below. Several triples in one entry form one slope class; singleton entries have no coincidences.

pair 45				pair 46				
triples	slope	triples	slope	triples	slope	triples	slope	
012	d/e	037	a/c	012	$d/(bc)$	057	a^{-1}	
013	$e/(bc)$	067, 126	a^{-1}	013	$e/(ac)$	123	ac/e	
016, 127, 367	1	123	ac/d	015	a/c	125	b/a	
017, 267	b	136	b^2/c	017	b/c	157	b^{-1}	(45)
023	b/a	137	c/a^2	023	bc/d	235	c/b	
026	a^2	167	a/b	025, 137	c	237	c/a	
027	b^{-2}	236	c/b	027, 135	c^{-1}	257	a	
036	c^{-1}	237	c	035, 127	1	357	b	
				037	a/b			

The entries in (45) are reduced quotients of the five irreducibles, so unique factorization makes different displayed quotients unequal within each marked pair.

Before cancellation, every numerator and denominator has degree at most three. Two unequal slopes therefore have a nonzero cross-product difference of degree at most six. It cannot vanish at α. Thus evaluation preserves all the classes and creates no new ones. The AES generator has degree eight, from $\mathbb{F}_{256} = \mathbb{F}_2[t]/(t^8 + t^4 + t^3 + t + 1)$.

For pair 45, the triples $016, 127, 367$ give respectively $\Delta_{13}(01)/\Delta_{03}(01) = c/c$, $\Delta_{12}(12)/\Delta_{02}(12) = c/c$, and $M_{13}/M_{03} = 1$. Rotating inputs and outputs gives all six marked pairs; reversing a pair inverts its slopes and preserves their equality classes. □

From coincidences to two activity groups. The equality classes of Lemma 15 define a nonnegative overlap matrix after a change of signs. Grouping its rows into odd subsets and pairs gives a two-dimensional upper bound. The weights 3 and 8 will make its spectral-radius estimate a pair of rational inequalities.

Lemma 16 (A slope criterion for degree two). *Consider a circulant four-coordinate MDS map on $\mathbb{F}_{256}$ with the same slope classes as in* (45)*. Then*

$$\|F_2 \Pi_M F_0\|^2 < \frac{66}{5m^2}, \qquad \|F_2 \Pi_M F_0\| < \frac{71}{5000}. \tag{46}$$

The reverse block $F_0 \Pi_M F_2$ obeys the same bounds.

Proof. Retain only the equality term in (44). The discarded constant kernel is the Gram matrix of the same scalar response for every line. Contracting it against the positive inverse output metric preserves positivity; its minus sign permits its removal in an upper bound.

Change the sign of input coefficient A by $(-1)^{|A|}$, using a diagonal sign matrix Σ. For a marked output pair R, let S, T be subsets of its other two outputs. Their sizes must be $3 - |A|, 3 - |B|$, respectively. Therefore $|S \triangle T| \equiv |A| + |B| \pmod 2$: the sign change removes every negative inverse-frame coefficient. Define the nonnegative matrix

$$\mathcal{J}_{AB} = \sum_{R} \sum_{\substack{|S|=3-|A| \\ |T|=3-|B|}} \xi^{|S \triangle T|} \mathbf{1}_{\mathrm{slope}_R(A \cup S) = \mathrm{slope}_R(B \cup T)}.$$

The sums range over all six marked output pairs; invalid subset sizes contribute zero. The Gram upper bound is

$$m^2 \Sigma \mathcal{V}^* B_2^* B_2 \mathcal{V} \Sigma \preceq c_0 \mathcal{J}. \tag{47}$$

Its factor is $c_0^2 Q_b/m = c_0$: the inverse metric on the two unmarked outputs contributes c_0^2, and the scaled equality term contributes Q_b/m. Also

$$\Sigma (\mathsf{G}_{\mathrm{eff}})^{-1} \Sigma = c_0^4 K_{\mathrm{dist}}, \qquad (K_{\mathrm{dist}})_{AB} = \xi^{|A \triangle B|}.$$

The inverse effective metric is the retained principal submatrix of G_4^{-1}. Hence $m^2 \|B_2\|^2$ is bounded by the spectral radius of $c_0^5 K_{\mathrm{dist}} \mathcal{J}$. This product is similar

to the positive semidefinite Gram comparison obtained from (47) by the inverse square root of the sign-changed metric.

Let $\mathcal{O}$ be the eight input subsets of size one or three and $\mathcal{M}$ the six pairs. Count row sums into these two groups. The diagonal of $\mathcal{J}$ is $6\binom{2}{3-|A|}$, giving $6, 12, 6$ for sizes $1, 2, 3$. The nontrivial classes supply the remaining entries:

input	$\mathcal{O}$	$\mathcal{M}$
singleton	6	4ξ
adjacent pair	4ξ	$12 + 3\xi^2$
opposite pair	0	$12 + 4\xi^2$
triple	6	0

Pair types refer to the four-cycle of input positions. For a direct check, the 45 class $\{016, 127, 367\}$ contains inputs $\{0, 1\}, \{1, 2\}, \{3\}$ with uniform output sets $\{6\}, \{7\}, \{6, 7\}$. Its pair–pair entry is ξ^2 and its two singleton–pair entries are ξ. All three 45 classes together give four singleton–adjacent-pair coincidences and one coincidence between adjacent pairs. Their four rotations give four cross incidences at each singleton and adjacent pair, and two neighbors at each adjacent pair. The three 46 classes give one adjacent-pair coincidence and two opposite-pair coincidences. Their two rotations add one neighbor at each adjacent pair and multiplicity four between the opposite pairs. No class contains an input triple. This proves the table.

The corresponding row sums of K_{dist} follow just from subset distances:

$$\widehat{K}_{\mathrm{dist}} = \begin{pmatrix} 1 + 6\xi^2 + \xi^4 & 3\xi + 3\xi^3 \\ 4\xi + 4\xi^3 & 1 + 4\xi^2 + \xi^4 \end{pmatrix}, \qquad \widehat{J} = \begin{pmatrix} 6 & 4\xi \\ 4\xi & 12 + 4\xi^2 \end{pmatrix}.$$

For example, an odd subset has six other odd subsets at distance two and one at distance four; its pair neighbors have distances one and three, three of each. A pair has four odd neighbors at each of those distances, and four other pairs at distance two and one at distance four. Thus $\widehat{K}_{\mathrm{dist}}$ gives the exact group row sums of K_{dist}, while $\widehat{J}$ bounds those of $\mathcal{J}$.

For nonnegative matrices, group row bounds compose: apply the second matrix to a vector constant on each group, then apply the first to its group bounds. Give odd inputs weight 3 and pairs weight 8. Since $\widehat{J}(3, 8)^\top = (20, 775/8)^\top$ and $c_0^5 < 51/50$, the entire estimate reduces to two rational inequalities at $\xi = 1/16$:

$$c_0^5 \widehat{K}_{\mathrm{dist}} \widehat{J} \begin{pmatrix} 3 \\ 8 \end{pmatrix} < \begin{pmatrix} 79/2 \\ 211/2 \end{pmatrix} < \frac{66}{5} \begin{pmatrix} 3 \\ 8 \end{pmatrix}.$$

For hand checking, the two entries of $\widehat{K}_{\mathrm{dist}}(20, 775/8)^\top$ are $1268255/32768$ and $54216455/524288$. Let $P = c_0^5 K_{\mathrm{dist}} \mathcal{J}$ and $W = \mathrm{diag}(w)$, where $w_A = 3$ on $\mathcal{O}$ and $w_A = 8$ on $\mathcal{M}$. The preceding inequalities give

$$\sum_B (W^{-1} P W)_{AB} = \frac{(Pw)_A}{w_A} < \frac{66}{5}.$$

Since W is positive diagonal, this similar matrix is nonnegative and has the same spectral radius as P. Its maximum row sum bounds that common spectral radius

strictly below $66/5$, proving the squared bound. Finally $66/(5\cdot 256^2) < (71/5000)^2$ gives the norm bound.

For the reverse block, additive Fourier conjugation preserves each degree projection and sends Π_M to $\Pi_{M^{-\top}}$. Reversing the four coordinates conjugates a circulant matrix to its transpose, hence conjugates M^{-1} to $M^{-\top}$. Together with adjoints, these isometries identify the forward and reverse block norms. □

The resulting damped blocks satisfy

$$\sqrt{q}\,\|R_{20}\|,\ \sqrt{q}\,\|R_{02}\| < 256^2(9/128)^2(71/5000) = 4.6008 < 4.601. \tag{48}$$

6.5 Degree-three estimates

Three residual positions leave five activity positions in the MixColumns graph. The surviving subsets have size two or three, giving ten lines and ten planes on the residual coordinates. Their pairwise overlaps serve two different estimates. Comparing the response Gram with the input-frame metric bounds the operator norm; contracting it with the inverse frame metric gives the total squared Hilbert–Schmidt norm. Both calculations need the sign of a contained line–plane overlap.

Put $m = 256$, $q = m^4$, $\xi = 1/16$, and $c_0 = 256/255$. At each marked position fix one permitted byte residual map R_i; different positions may use different orientations. For an activity choice α, use the contracted vector from (43). In the three-coordinate case, writing $d = |\alpha| - 1$ makes its normalization explicit:

$$v_\alpha = m^{-(d+1)/2} \sum_{x\in\mathcal{C}_\alpha} \bigotimes_{i=1}^{3} R_i e_{x_i}, \qquad \Gamma_{\alpha\beta} = q\,\langle v_\alpha, v_\beta\rangle\,.$$

The byte maps remain fixed when comparing activity choices. Some positive overlaps will enter with negative inverse-frame coefficients. Keeping a lower bound on these overlaps preserves the resulting cancellation; replacing every overlap by its absolute upper bound would lose it.

Lemma 17 (Three-coordinate overlaps). *For any three marked graph positions, surviving activity choices α, β (each of size two or three), and any fixed choices $R_i \in \mathcal{T}_b$, the following bounds hold, with $d_{\mathrm{H}} = |\alpha \triangle \beta|$:*

d_{H}	*Case*	*Overlap bound*
1	*contained*	$c_- < \Gamma_{\alpha\beta} < c_+$
2	*equal dimension*	$\lvert\Gamma_{\alpha\beta}\rvert < e_1$
3	*transverse*	$\lvert\Gamma_{\alpha\beta}\rvert < f$
4	*equal dimension*	$\lvert\Gamma_{\alpha\beta}\rvert < e_2$

Here the constants are exact rationals:

$$e_1 = .00305, \quad e_2 = .023, \quad f = .053, \quad c_- = .0185, \quad c_+ = .1165.$$

At even distances both responses come from lines or both from planes. At distance one the line is contained in the plane; its positive lower overlap will be retained in

both the operator and Gram-moment estimates. At distance three they intersect only at zero. Lemma 18 supplies the diagonal and complementary cases by cubic moments.

Proof. Write $G_i = R_i^* R_i$ for the byte Grams. For a line the coefficient of its unnormalized indicator in (43) is m^{-1}; for a plane it is $m^{-3/2}$. The four geometric cases below are the inputs needed by Lemma 19 and Lemma 20.

Two lines or two planes ($d_{\mathrm{H}} = 2, 4$). For $d_{\mathrm{H}} = 2$, the sum of the two lines is a two-dimensional MDS code. Their three coordinate-equality sets are disjoint on nonzero parameter pairs, each of size $m-1$. For $d_{\mathrm{H}} = 4$ the lines are distinct, but two equality sets may coincide. The byte diagonal and off-diagonal bounds give

$$m^2(m-1)\{3\Delta_b\gamma_{b,-}^2 + (m-4)\gamma_{b,-}^3\} < e_1,$$
$$m^2(m-1)\{\Delta_b^2\gamma_{b,-} + \Delta_b\gamma_{b,-}^2 + (m-3)\gamma_{b,-}^3\} < e_2.$$

The second expression also covers three distinct equality sets, since the difference is $m^2(m-1)\gamma_{b,-}(\Delta_b - \gamma_{b,-})^2 \geq 0$. Shortening the original MDS graph proves the distance-two assertion; Fourier duality gives the same bounds for two planes.

A contained line and plane ($d_{\mathrm{H}} = 1$). The line lies in the plane. Their equal nonzero pairs contribute $z_* = m^{3/2}(m-1)\Delta_b^3$. The plane has $3(m-1)$ words of weight two and $(m-1)(m-2)$ of weight three. Each difference has m preimages in the line–plane pair space, so the remaining contribution has absolute value at most

$$h_* = m^{5/2}(m-1)\{3\Delta_b\gamma_{b,-}^2 + (m-2)\gamma_{b,-}^3\}.$$

Substitution gives $z_* - h_* > .0185$ and $z_* + h_* < .1165$, the stated interval.

A transverse line and plane ($d_{\mathrm{H}} = 3$). The line and plane intersect only at zero and sum to the full three-coordinate space. Its weight enumerator gives

$$m^{3/2}\{(\Delta_b + (m-1)\gamma_{b,-})^3 - \Delta_b^3\} < f.$$

For both cases the original shortenings share a zero graph coordinate; their sum has dimension at most three, so projection onto the marked triple is injective and preserves these intersections. All scalar comparisons use the rational byte data in Lemma 13. Its uniform bounds apply separately to each marked position, including adjoints on input positions. □

Cubic and complementary overlaps. The geometric bounds above leave two cases: a response paired with itself, and a line paired with the complementary plane. Hölder's inequality reduces the first to an absolute cubic Gram moment and the second to a cubic moment of the mixed Gram–Fourier kernel. Put

$$\begin{aligned} S_3 &= \frac{m-1}{L_b^3}\big(16766^3 + 70\cdot 191^3 + 112\cdot 64^3 + 72\cdot 319^3\big), \\ a_0 &= \frac{1081}{1000}, \qquad g_F = \frac{13276}{4L_b}, \qquad b_0 = \frac{111}{500}. \end{aligned} \tag{49}$$

Lemma 18 (Cubic and complementary overlaps). *For every byte Gram $G = R^*R$, $R \in \mathcal{T}_b$,*

$$\sum_{x,y} |G_{xy}|^3 \le S_3, \qquad \max_{x,y} |(G\mathcal{F}_b)_{xy}| \le g_F, \qquad m^2 S_3 < a_0, \qquad m^2 g_F T_b < b_0.$$

For the damped vectors of Lemma 17, $q\,\|v_\alpha\|_2^2 < a_0$. If the line and plane activity choices are complementary among the five activity positions, then $|q\,\langle v_\alpha, v_\beta\rangle\,| < b_0$.

Proof. Lemma 13 gives the three off-diagonal multiplicities $h, 252-2h, h+2$, with $h \le 70$. The coefficient of h in their absolute cubic sum is $191^3 - 2\cdot 64^3 + 319^3 > 0$. Summing over the $m-1$ nonzero rows therefore gives S_3. Relabeling preserves this sum, as does passing to another allowed Gram. For an explicit rational comparison, use

$$\Delta_b^3 < 6.4627 \cdot 10^{-8}, \quad \gamma_{b,+}^3 < 9.56 \cdot 10^{-14},$$
$$(64/L_b)^3 < 3.6 \cdot 10^{-15}, \quad \gamma_{b,-}^3 < 4.46 \cdot 10^{-13}.$$

The weighted sum with coefficients $1, 70, 112, 72$ is below $6.467 \cdot 10^{-8}$. Multiplication by $m^2(m-1) = 16711680$ gives $1.0807443456 < a_0$. All decimals in this calculation are exact terminating rationals.

For the mixed kernel, take $x, y \ne 0$ and put $z = xy$. Summing the inversion Gram formula of Lemma 13 against an additive character gives

$$(G_*\mathcal{F}_b)_{xy} = \frac{1}{16}\left[a_b\psi(z) + b_b\{\mathcal{K}(z) - 1 + m\mathbf{1}_{z=1}\} - c_b\right]. \tag{50}$$

For $z = 1$ the bracket is positive even at $\mathcal{K}(1) = -33$, and its upper bound at 31 is $16g_F$. For $z \ne 1$, its absolute value is at most $a_b + 34b_b - c_b < 16g_F$. The zero row and column vanish. If $G = \Pi_L G_* \Pi_L^*$ for an invertible binary-linear map L, Fourier duality gives

$$G\mathcal{F}_b = \Pi_L G_* \mathcal{F}_b \Pi_{L^\dagger},$$

where $L^\dagger$ is the trace-pairing adjoint. Both outside maps permute labels, so the same entry bound holds for every allowed G. Moreover,

$$m^2 g_F T_b = \frac{7586530372}{34225520640} < \frac{111}{500};$$

the cross-products are $3793265186000 < 3799032791040$.

A damped line can be written as

$$v = m^{-1} \sum_x R_1 e_{\lambda_1 x} \otimes R_2 e_{\lambda_2 x} \otimes R_3 e_{\lambda_3 x}, \qquad \lambda_1, \lambda_2, \lambda_3 \ne 0.$$

Its squared norm is a sum of products of three relabeled Gram kernels. Hölder's inequality gives $q\,\|v\|_2^2 \le m^2 S_3 < a_0$. A damped plane is the Fourier transform of a damped dual line before the three byte maps are applied. Its kernels are $\mathcal{F}_b G_i \mathcal{F}_b$, which are again allowed Grams, so the same argument applies. For a

complementary line–plane pair, the corresponding kernels are $G_i\mathcal{F}_b$, up to scalar relabelings. Since

$$\sum_{x,y} |(G_i\mathcal{F}_b)_{xy}|^3 \le g_F \, \|G_i\mathcal{F}_b\|_{\mathrm{HS}}^2 = g_F \, \mathrm{Tr}(G_i^2) \le g_F T_b,$$

Hölder gives the overlap bound $m^2 g_F T_b < b_0$. The three byte maps may be different; no Fourier commutation is used. □

The degree-three operator bound. We now compare the response Gram with the effective input metric from Section 6.3. Splitting input subsets by parity separates two tasks: absolute row sums control the responses within each class, while the signed overlap bounds control the coupling between classes. Throughout this calculation, $\xi = 1/16$ and $c_0 = 256/255$.

Lemma 19 (Degree-three coupling). *The word blocks satisfy* $\sqrt{q}\,\|R_{30}\| < 33/10$ *and* $\sqrt{q}\,\|R_{03}\| < 33/10$.

Proof. Let V_4 synthesize the tensor frame $[e_0, m^{-1/2}\mathbf{1}]^{\otimes 4}$ on the input activity space. Its Gram matrix, indexed by subsets of $[4]$, is

$$(\Gamma_4)_{AB} = \xi^{|A \triangle B|}.$$

Set $G_{\mathrm{in}} = qV_4^* R_{30}^* R_{30} V_4$. Its empty and full columns vanish because R kills the zero and constant word vectors. Let $\mathcal{S} = \{A : 1 \le |A| \le 3\}$, and split it into $\mathcal{O} = \{A : |A| \in \{1,3\}\}$ and $\mathcal{E} = \{A : |A| = 2\}$. Eliminating the empty and full coordinates of Γ_4 gives its Schur complement Γ_{eff}. Put

$$\lambda = (33/10)^2 = 1089/100, \qquad \kappa = 4c_0 = 1024/255, \qquad L = \lambda\Gamma_{\mathrm{eff}} - (G_{\mathrm{in}})_{\mathcal{S}\mathcal{S}}.$$

Positivity of L gives the required norm bound. We will obtain diagonal lower bounds $6I$ and $3I/2$, and a squared norm below $25/3$ for the block between them. A final Schur complement then leaves the positive margin $I/9$.

Fix one of the four marked output triples. One output activity bit ϵ remains, and its response at input subset A can be nonzero only when $2 \le |A| + \epsilon \le 3$. Write $\gamma_{A\epsilon,B\eta} = q\,\langle v_{A\epsilon}, v_{B\eta}\rangle$ for the overlap of the damped vectors in (43), with a vanishing response contributing zero. The inverse metric on the remaining output coordinate gives

$$c_0 \sum_{\epsilon,\eta\in\{0,1\}} (-\xi)^{\epsilon\oplus\eta} \gamma_{A\epsilon,B\eta} \tag{51}$$

as its contribution to $(G_{\mathrm{in}})_{AB}$. Summing the four triples replaces c_0 by κ in a uniform bound.

For an odd subset, only one bit survives. It has three other subsets of its own size, three noncomplementary subsets of the opposite odd size, and one complement. Lemmas 17 and 18 bound their overlaps by e_1, ξf, and ξb_0, respectively, and its diagonal by a_0. A pair has four pair neighbors at distance two and one at

distance four. Equal output bits in (51) contribute twice e_1 or e_2; unequal bits contribute twice ξf or ξb_0. Its diagonal is at most $2a_0$: the two cross terms have a negative coefficient and a positive contained line–plane overlap, so they may be removed in an upper bound. Consequently,

$$\sum_{B\in\mathcal{O}} |(G_{\mathrm{in}})_{AB}| < \kappa[a_0 + 3e_1 + \xi(3f + b_0)] < \kappa\frac{9}{8} \qquad (A \in \mathcal{O}),$$
$$\sum_{B\in\mathcal{E}} |(G_{\mathrm{in}})_{AB}| < \kappa[2a_0 + 8e_1 + 8\xi f + 2e_2 + 2\xi b_0] < \kappa\frac{229}{100} \qquad (A \in \mathcal{E}). \tag{52}$$

The brackets are exactly 1.1139625 and 2.28665.

To compare these response bounds with input lengths, we first bound the full metric on each parity class, then subtract the correction from eliminating the empty and full subsets. Let Σ multiply the coordinate at subset A by $(-1)^{|A|}$. The average

$$\frac{1}{2}(\Gamma_4 + \Sigma\Gamma_4\Sigma) = \frac{1}{2}\left[\begin{pmatrix}1 & \xi\\ \xi & 1\end{pmatrix}^{\otimes 4} + \begin{pmatrix}1 & -\xi\\ -\xi & 1\end{pmatrix}^{\otimes 4}\right]$$

removes the mixed-parity entries. Tensor products of $(1, 1)$ and $(1, -1)$ diagonalize both summands. The eigenvalues of their average are

$$1 + 6\xi^2 + \xi^4, \qquad 1 - \xi^4, \qquad (1 - \xi^2)^2.$$

The last is the smallest. Hence every principal block on a single parity class is at least $(1 - \xi^2)^2 I$.

Eliminating the empty and full subsets subtracts $\mathsf{C}E^{-1}\mathsf{C}^*$, where E is their 2×2 metric and C couples the retained subsets to these endpoints. We have

$$E = \begin{pmatrix}1 & \xi^4\\ \xi^4 & 1\end{pmatrix}, \qquad \left\|E^{-1}\right\| = (1 - \xi^4)^{-1}.$$

On $\mathcal{O}$, the coupling has four rows (ξ, ξ^3) and four rows (ξ^3, ξ); on $\mathcal{E}$, it has six identical rows (ξ^2, ξ^2). Thus

$$\|\mathsf{C}_{\mathcal{O}}\|^2 = 4\xi^2(1 + \xi^2)^2, \qquad \|\mathsf{C}_{\mathcal{E}}\|^2 = 12\xi^4.$$

At $\xi = 1/16$, the tensor lower bound exceeds $127/128$, while the two correction norms are at most

$$\frac{4\xi^2(1 + \xi^2)^2}{1 - \xi^4} = \frac{1}{64}\frac{257}{255} < \frac{3}{128}, \qquad \frac{12\xi^4}{1 - \xi^4} = \frac{12}{65535} < \frac{1}{128}.$$

Consequently,

$$(\Gamma_{\mathrm{eff}})_{\mathcal{O}\mathcal{O}} \succ \frac{31}{32}I, \qquad (\Gamma_{\mathrm{eff}})_{\mathcal{E}\mathcal{E}} \succ \frac{63}{64}I.$$

The absolute row bounds in (52) bound the norms of the corresponding response blocks. Since $\kappa < 201/50$, we obtain

$$\begin{aligned} L_{\mathcal{O}\mathcal{O}} &\succ \left(\lambda\frac{31}{32} - \frac{201}{50}\frac{9}{8}\right) I = \frac{19287}{3200} I \succ 6I, \\ L_{\mathcal{E}\mathcal{E}} &\succ \left(\lambda\frac{63}{64} - \frac{201}{50}\frac{229}{100}\right) I = \frac{242247}{160000} I \succ \frac{3}{2} I. \end{aligned} \tag{53}$$

The diagonal margins in (53) leave room for the coupling between parity classes. Here it matters that L is a difference of two Grams: a positive response overlap reduces a positive metric entry. The mixed-parity correction from eliminating the empty and full subsets is

$$p_\times = \frac{\xi^3(1+\xi^2)}{1+\xi^4}, \qquad \xi^3 \le p_\times \le \xi^3 + \xi^5,$$

so $(\Gamma_{\text{eff}})_{AB} = \xi^{|A \triangle B|} - p_\times$ between the classes. Call the two subsets comparable when one contains the other. In (51), comparable subsets give a contained line–plane overlap and an equal-dimension overlap at distance two; noncomparable subsets give a transverse overlap and an equal-dimension overlap at distance four. Therefore

$$\begin{aligned} \kappa(c_- - \xi e_1) < (G_{\text{in}})_{AB} < \kappa(c_+ + \xi e_1) \quad & \text{for comparable subsets,} \\ |(G_{\text{in}})_{AB}| < \kappa(f + \xi e_2) \quad & \text{otherwise.} \end{aligned}$$

In the comparable case the entry of L is positive, since

$$\lambda(\xi - \xi^3 - \xi^5) = \frac{71088831}{104857600} > \frac{2}{3} > \frac{1}{2} > \kappa(c_+ + \xi e_1).$$

Its absolute value is thus at most $a = \lambda(\xi - \xi^3) - \kappa(c_- - \xi e_1)$. For a noncomparable pair it is at most $b = \lambda\xi^5 + \kappa(f + \xi e_2)$. One common comparison gives

$$\begin{aligned} a + b &= \lambda(\xi - \xi^3 + \xi^5) + \kappa[f - c_- + \xi(e_1 + e_2)] \\ &= \frac{71091009}{104857600} + \frac{23122}{159375} < \frac{17}{25} + \frac{3}{20} = \frac{83}{100} < \frac{5}{6}. \end{aligned} \tag{54}$$

An odd subset has three comparable and three noncomparable pairs; a pair has four comparable and four noncomparable odd subsets. Hence the cross-parity absolute row sums are below $5/2$ and $10/3$, respectively.

The row–column norm bound therefore gives

$$\|L_{\mathcal{O}\mathcal{E}}\|^2 < \frac{5}{2}\frac{10}{3} = \frac{25}{3}.$$

Together with (53), this controls the final Schur complement:

$$L_{\mathcal{E}\mathcal{E}} - L_{\mathcal{E}\mathcal{O}} L_{\mathcal{O}\mathcal{O}}^{-1} L_{\mathcal{O}\mathcal{E}} \succ \left(\frac{3}{2} - \frac{25/3}{6}\right) I = \frac{1}{9} I.$$

Thus $L \succ 0$. The Schur-complement criterion now gives $G_{\text{in}} \prec \lambda\Gamma_4$. Since V_4 spans the input activity space, $\sqrt{q}\,\|R_{30}\| < 33/10$. For R_{03}, vectorize the marked input legs instead. Their maps are adjoint byte residuals, which belong to $\mathcal{T}_b$, so the same overlap bounds and metric comparison apply. □

The degree-three energy bound. For a fixed set of three residual positions, we need the total squared length of the block's entries. Vectorization treats input and output positions together, so the Hilbert–Schmidt estimate uses all five activity coordinates at once. The same signed overlaps remain useful because the inverse metric has negative coefficients at odd distances.

Lemma 20 (Degree-three energy). *Every fixed marked-position block of total degree three has squared Hilbert–Schmidt norm below* $22/q$.

Proof. Use the five-coordinate activity frame. There are ten line and ten plane choices. Its inverse Gram has entries $c_0^5(-\xi)^{|\alpha \triangle \beta|}$. For each line or plane, the equal-size neighbors have counts six and three at distances two and four. The opposite-size neighbors have counts three, six, and one at distances one, three, and five. Contracting the response Gram with this inverse metric therefore gives the upper bound

$$E_3 \le 20c_0^5 \left[a_0 + 6\xi^2 e_1 + 3\xi^4 e_2 - 3\xi c_- + 6\xi^3 f + \xi^5 b_0\right]. \tag{55}$$

At distance one the inverse-metric coefficient is negative, so we use the positive lower overlap c_- from Lemma 17. At distances three and five we use the negative endpoints of the symmetric overlap intervals. The bracket equals $1.077681635665893554687 5 < 1.077682$. The binomial expansion of $(1 + 1/255)^5$ gives $c_0^5 < 1.01977$. Thus

$$E_3 < 20(1.01977)(1.077682) = 21.9797554628 < 22.$$

These counts and bounds apply to any fixed marked positions and their allowed byte maps. □

6.6 Remaining block estimates

For transfers between positive residual degrees, the number of graph extensions bounds each activity-frame entry. These counts depend only on the number of uniform positions, so the scalar comparison arrays compress by subset size. Lemma 14 converts their norms into diffusion bounds; multiplying by the byte contractions then gives the remaining entries for the word operator comparison.

Use $m = 256$. If a block has u activity coordinates in total, its two inverse synthesis maps contribute the factor $(1 - 1/16)^{-u/2}$ in Lemma 14.

For degrees (i, j), put $k = i + j$. A frame entry with t uniform activity choices shortens the MDS graph to dimension $d = k - 4 + t$. When $0 < d < k$, its matrix on the marked output and input coordinates has at most $m^{d-\min(i,d)}$ ones per row and $m^{d-\min(j,d)}$ ones per column. There is at most one graph extension per entry, since the marked coordinates determine the shortened codeword.

For a fixed row, the compatible graph points form its *fiber*: the fixed marked outputs impose $\min(i, d)$ independent conditions, leaving $d-\min(i, d)$ free symbols. Fixing a column gives the analogous count. For example, $d = 2$, $i = 1$, $j = 2$ leaves at most m ones in a row and one in a column. The matrix norm is at

most the square root of those two counts. Including the activity normalization therefore gives the scalar entry bound

$$m^{-t/2+d-(\min(i,d)+\min(j,d))/2}.$$

All other dimensions vanish after the residual projections. Lemma 14 now combines these scalar entries, with factor $\sqrt{\binom{4}{i}\binom{4}{j}}$ for the residual-position choices. For $(i,j)=(1,1)$ only $t=3$ survives. For $(1,2)$ the values are $t=2,3$, each with bound m^{-1}. For $(2,2)$ the values $t=1,2,3$ have bounds $m^{-1/2}, m^{-1}, m^{-1/2}$, respectively. Taking the norm of these three activity-choice arrays gives

$$\begin{aligned}
&\|F_1 \Pi_M F_1\| \le 12(16/15)^3 m^{-3/2},\\
&\|F_1 \Pi_M F_2\| \le \sqrt{24(8+2\sqrt{13})}(16/15)^{5/2}/m,\\
&\|F_2 \Pi_M F_2\| \le 3(16/15)^2\{3/m+\sqrt{m^{-2}+16/m}\} < .9.
\end{aligned}$$

Compressing activity arrays by subset size. Each scalar array depends only on the sizes of its indexing subsets. This lets us compute its norm on a smaller space without any loss. For nonnegative integers a, b, write $K_{S,T} = f(|S|+|T|)$, where $S \subseteq [a]$, $T \subseteq [b]$, and f is real-valued. Let column i of U_a be the indicator of the size-i subsets divided by $\sqrt{\binom{a}{i}}$. These columns are orthonormal. Thus

$$K = U_a \widehat{K} U_b^*, \qquad \widehat{K}_{ij} = \sqrt{\binom{a}{i}\binom{b}{j}}\, f(i+j), \qquad \|K\| = \left\|\widehat{K}\right\|.$$

Indeed, the array kills vectors whose coefficients sum to zero within each size class; its action on the orthogonal complement is exactly $\widehat{K}$. We apply this identity to the three cases above.

For degrees $(1,1)$, take $a=b=3$ and $f(t)=\mathbf{1}_{t=3}$, leaving the factor $m^{-3/2}$ outside. The compressed matrix is anti-diagonal with entries $1,3,3,1$, so its norm is 3. The four residual-position choices and the frame factor $(16/15)^3$ give the first displayed bound.

For degrees $(1,2)$, take $a=3$, $b=2$ and $f(t)=\mathbf{1}_{t\in\{2,3\}}$, leaving m^{-1} outside. The Gram matrix $\widehat{K}^*\widehat{K}$ preserves the symmetric and antisymmetric subspaces under complementation of the two input positions. In size coordinates, use the orthonormal basis $(e_0+e_2)/\sqrt{2}, e_1$ for the symmetric subspace. The restriction is

$$\begin{pmatrix} 4 & 6 \\ 6 & 12 \end{pmatrix}, \qquad \text{with eigenvalues } 8 \pm 2\sqrt{13}.$$

The antisymmetric vector $(e_0-e_2)/\sqrt{2}$ has eigenvalue 4. Hence the array norm is $\sqrt{8+2\sqrt{13}}$. Multiplication by $\sqrt{\binom{4}{1}\binom{4}{2}} = \sqrt{24}$ and $(16/15)^{5/2}/m$ gives the second bound.

For degrees $(2,2)$, take $a=b=2$ and set $f(1)=f(3)=A=m^{-1/2}$, $f(2)=B=m^{-1}$, and $f(0)=f(4)=0$. On the same symmetric basis, the

compressed array itself is

$$\begin{pmatrix} B & 2A \\ 2A & 2B \end{pmatrix}.$$

Its antisymmetric eigenvalue is $-B$. For nonnegative A, B, the larger eigenvalue of the displayed matrix dominates the absolute values of all other eigenvalues, giving

$$\|K\| = \frac{3B + \sqrt{B^2 + 16A^2}}{2}.$$

The six residual-position choices and the frame factor $(16/15)^2$ give the third bound. These calculations depend only on subset sizes and the alphabet size, so they apply to every marked-position pattern and every MDS orientation, and can be reused at word scale.

Multiplying by $m^2 s^{i+j}$ bounds these three scaled entries by $1.16, 2, 1.45$, respectively. Every other positive-degree transfer uses the elementary diffusion norm one. Activity/degree-one transfer is zero because the shortened dimension cannot lie strictly between zero and one. Together with the degree-three activity block, these estimates give the comparison matrix of Lemma 23 in Section 7.3. □

Fixed-position energies outside degree three. For each $k \in \{2, 4, 5, 6, 7, 8\}$, let E_k bound q times the squared Hilbert–Schmidt norm of a block with k fixed residual positions across the input and output. We derive these bounds from code energies and the activity frame. Degree three is treated directly in Lemma 20; the final block-energy matrix and word Gram moment are proved in Lemma 25.

For $0 < d < k$, put $a = \min(d, k - d)$ and define

$$\zeta_{k,d} = m^k (T_b/m)^a \Delta_b^{k-2a}. \tag{56}$$

The mixed-map version of Proposition 2 and (43) give $q\|v_\alpha\|^2 \le \zeta_{k,d}$ for an activity choice whose shortening has dimension d. In particular $\zeta_{k,d} = \zeta_{k,k-d}$.

The squared Hilbert–Schmidt norm is a contraction against the inverse frame metric, as in the degree-three energy proof. Here the individual code energies suffice: Cauchy–Schwarz bounds their overlaps, and taking absolute values gives a row-sum estimate. Fix the k marked residual positions, put $\ell_{\mathrm{act}} = 8 - k$, and let $\mathcal{S}$ be the activity subsets $\alpha \subseteq [\ell_{\mathrm{act}}]$ for which $0 < k - 4 + |\alpha| < k$. The other contracted vectors vanish. Here the $\ell_{\mathrm{act}} = 8 - k$ activity positions include both input and output coordinates of the diffusion graph. Their product frame therefore uses $\mathsf{G}_{\ell_{\mathrm{act}}}^{-1}$. The squared Hilbert–Schmidt norm of this marked-position block is the contraction of their Gram with $\mathsf{G}_{\ell_{\mathrm{act}}}^{-1}$. Thus, with $\Gamma_{\alpha\beta} = q\langle v_\alpha, v_\beta \rangle$,

$$q\|\mathrm{block}\|_{\mathrm{HS}}^2 = c_0^{\ell_{\mathrm{act}}} \sum_{\alpha,\beta \in \mathcal{S}} (-\xi)^{|\alpha \triangle \beta|} \Gamma_{\alpha\beta}.$$

Cauchy–Schwarz gives $|\Gamma_{\alpha\beta}| \le (\zeta_{k,d_\alpha} + \zeta_{k,d_\beta})/2$. The coefficient matrix is symmetric, so taking absolute values yields

$$q\|\mathrm{block}\|_{\mathrm{HS}}^2 \le c_0^{\ell_{\mathrm{act}}} \sum_{\alpha \in \mathcal{S}} \zeta_{k,d_\alpha} \sum_{\beta \in \mathcal{S}} \xi^{|\alpha \triangle \beta|}. \tag{57}$$

The inner sums are elementary. For $k = 2$, the admissible subsets are the twenty triples in six positions; a fixed triple has $1, 9, 9, 1$ neighbors at distances $0, 2, 4, 6$. For $k = 4$, every subset of four positions except the empty and full sets is admissible; a row of weight t has sum $(1+\xi)^4 - \xi^t - \xi^{4-t}$. For $k \geq 5$ all subsets are admissible, and the row sum is $(1+\xi)^{\ell_{\mathrm{act}}}$. Since $c_0(1+\xi) = (1-\xi)^{-1}$, these facts give

$$E_2 = 20c_0^6\zeta_{2,1}(1 + 9\xi^2 + 9\xi^4 + \xi^6),$$
$$E_4 = c_0^4\big[8\zeta_{4,1}((1+\xi)^4 - \xi - \xi^3) + 6\zeta_{4,2}((1+\xi)^4 - 2\xi^2)\big].$$

For $5 \leq k \leq 8$, group the surviving activity subsets by $t = |\alpha|$. Every subset is allowed, and there are $\binom{8-k}{t}$ of that size. Thus all four degrees use

$$E_k = (1-\xi)^{-(8-k)} \sum_{t=0}^{8-k} \binom{8-k}{t} \zeta_{k,k-4+t}. \tag{58}$$

For example, symmetry of ζ reduces the degree-five sum to $2\zeta_{5,1} + 6\zeta_{5,2}$. At degree eight no activity positions remain, and the sum is just $\zeta_{8,4}$.

These formulas give the following strict rational bounds:

k	2	4	5	6	7	8
$E_k <$	23.11	20.5	11.76	5.8	2.84	1.412.

(59)

By symmetry, each bound uses at most two distinct ζ values, so each comparison is a rational substitution in (56). The mixed-map code bound applies to every marked-position choice; input positions use adjoint byte maps. This proves (59). □

Together with the activity estimate (37) and the signed degree-three estimate of Lemma 20, these bounds cover every fixed residual-position pattern. Section 7.4 combines them with the block operator norms to obtain the word Gram moment and total energy.

7 Word Norms, Moments, and Sources

The word estimates combine the two byte layers with the dependence created by MixColumns. For a fixed input, we follow the resulting source law through both layers. For an arbitrary input vector, we assemble the byte-degree blocks of Section 6. These arguments give the column energies, operator norm, and Gram moment needed by the code and fiber estimates. One-sided Fourier maps require an additional entry-moment calculation, because they can occur when two adjacent diffusion graphs use different Fourier representations (Section 8.1).

7.1 Required bounds and orientations

We now bound the local quantities of Section 2.4 for the AES word residual $R = K - P$. The bounds must hold in each orientation used by the code inequalities.

For AES, use the bytewise trace pairing and write $\mathcal{F} = \mathcal{F}_b^{\otimes 4}$, where $\mathcal{F}_b$ is defined in Section 6. Write

$$\mathcal{O} = \{R, R^*, \mathcal{F}R\mathcal{F}, \mathcal{F}R^*\mathcal{F}\}, \qquad R_{ij} = F_i R F_j.$$

Here the F_i are the byte-degree projections of Section 6.2.

Lemma 21 (Local AES residual bounds). *Let $q = 2^{32}$ and put*

$$\nu = \frac{52}{5\sqrt{q}}, \qquad \delta = \frac{2}{15q}, \qquad \tau = \frac{300}{q}, \tag{60}$$

$$h = \frac{4251}{q}, \qquad p_{\mathrm{ent}} = \frac{27}{q}. \tag{61}$$

For every $A \in \mathcal{O}$,

$$\begin{gathered} \|A\| < \nu, \quad \max_x \|Ae_x\|_2^2 < \delta, \quad \mathrm{Tr}((A^*A)^2) < \tau^2, \\ \|A\|_{\mathrm{HS}}^2 < h, \quad \max_{x,y} |A_{xy}| < p_{\mathrm{ent}}. \end{gathered} \tag{62}$$

Section 7.2 proves the column estimate, and Lemma 22 proves the entry estimate. Lemma 24 gives the operator norm; Lemma 25 gives the Gram moment and total energy.

Adjoints and Fourier conjugation preserve the norm and Gram moment, but can change column energies. The proof of δ therefore checks the byte-channel representation of all four operators. The Fourier forms permit code duality in Proposition 2; the adjoint forms permit reversed transfers.

For the one-sided Fourier maps $A\mathcal{F}$, the fourth-moment constants are

$$r_4 = \frac{2100}{q}, \qquad s_4 = \frac{1200}{q}. \tag{63}$$

Lemma 27 uses these constants for the row, column, and total fourth moments. Simultaneous Fourier conjugation gives another nonnegative full channel, whereas a one-sided Fourier map can have signed entries. Section 7.5 estimates their fourth moments after establishing the corresponding energy bounds.

7.2 Columns, entries, and ordinary sources

For a nonzero input difference, we prove the word-column collision bound $\mathrm{Col}(Ke_a) < 17/(15q)$ and its counterparts in the other three orientations. Subtracting the uniform nonzero-word law then gives the residual column energy. The proof uses the restricted-source estimate of Section 5.3 to retain both the dependencies introduced by MixColumns and the probabilities of its zero-coordinate events.

Individual columns: following a fixed input. A column of K starts from a fixed word difference. After the first SubBytes layer, its four byte differences are independent: a zero input byte stays zero, and a nonzero input byte produces a law with collision at most r_b and largest probability at most p_b. MixColumns then makes the output bytes dependent. To bound the final column, we must follow this dependence through the second SubBytes layer.

We separate the output according to its zero positions and residual positions. For example, suppose two input bytes are active and one output byte is prescribed to be zero. The MDS property leaves one free input symbol; the other input is determined by the zero equation. We keep the probability of this event in the restricted law, so its mass may be below one. The byte collision and norm estimates then control its residual response. Conditioning on the zero event would divide by its probability; retaining the unnormalized law lets the proof charge the event and its residual response together. Lemma 6 gives the needed restricted-source bound. We apply it with the byte channel or one of its three other orientations.

The lemma bounds the residual coordinates. To recover an entire activity component, we must also account for its constant-nonzero coordinates. On an exact support of size $\ell = 4 - z$, each such output factor contributes $1/(m-1)$ in squared norm. Thus the response in (18) is multiplied by $(m-1)^{j-\ell}$. Orthogonality permits summing over the $\binom{4}{\ell}$ supports and $\binom{\ell}{j}$ marked sets. We use this form in the byte calculation below.

The same restriction argument applies to all four orientations because $D, D^*, \widehat{D}, \widehat{D}^*$ have the common parameters in (27); each uses one of the mixing maps $M, M^{-1}, M^{\top}, M^{-\top}$, all MDS. Here M is MixColumns, and $K = D^{\otimes 4} \Pi_M D^{\otimes 4}$. The bytewise trace pairing gives

$$\mathcal{F} \Pi_M \mathcal{F} = \Pi_{M^{-\top}}, \qquad \mathcal{F} K \mathcal{F} = \widehat{D}^{\otimes 4} \Pi_{M^{-\top}} \widehat{D}^{\otimes 4}.$$

Taking adjoints gives M^{-1} and $M^{\top}$ with the transposed byte channels. Thus each orientation has the same independent-byte structure and the same collision, point-probability and residual-norm parameters. The byte and word activity projections are Fourier invariant: their ranges are spanned by the zero coordinate vector and the constant vector, which Fourier transformation interchanges up to normalization. Thus subtracting the corresponding activity projection gives the residual orientations in $\mathcal{O}$. Write K_A for the full channel associated with a residual orientation A.

A single active byte. Put $Q_b = 255$, $Q = q - 1$, and let λ be the nonzero-byte law produced by the first substitution layer. Its collision is r_b. MixColumns sends this one source bijectively to each of the four output bytes. For each position, pull its residual Gram back to the source labels and call it G_i. Lemma 13 gives the same diagonal, off-diagonal and norm bounds for every G_i, even when the four relabelings are different.

Each G_i kills constants. Centering λ therefore gives $\lambda^* G_i \lambda \le s^2 \Delta_b$, where $\|\lambda - Q_b^{-1}\mathbf{1}\|_2^2 = \Delta_b$. For powers $j \ge 2$, separate equal from distinct source labels:

$$\sum_{x,y} \lambda_x \lambda_y (G_i)_{xy}^j \le M_j, \qquad M_1 = s^2 \Delta_b,$$

$$M_j = r_b \Delta_b^j + (1 - r_b) c_j^j \quad (j = 2, 3, 4), \qquad c_2 = c_4 = \gamma_{b,-}, \quad c_3 = \gamma_{b,+}.$$

For even powers use the absolute off-diagonal bound; for the cubic power use its upper endpoint, since $\lambda_x \lambda_y \ge 0$.

The full column Grams are $C_i = Q_b^{-1}\mathbf{1}\mathbf{1}^* + G_i$ and are entrywise nonnegative. The word-column collision is a product of four responses to the same source. Weighted Hölder bounds it by

$$\sum_{x,y} \lambda_x \lambda_y \prod_{i=1}^{4} (C_i)_{xy} \le \prod_{i=1}^{4} \left(\sum_{x,y} \lambda_x \lambda_y (C_i)_{xy}^4 \right)^{1/4}.$$

Expanding the fourth power bounds each factor before taking its root by

$$Q_b^{-4} + 4Q_b^{-3} M_1 + 6Q_b^{-2} M_2 + 4Q_b^{-1} M_3 + M_4 < \frac{1133}{1000q} < \frac{17}{15q}. \tag{64}$$

The first comparison follows by substituting $r_b = 65/8192$, $s = 9/128$ and the rational Gram data of Lemma 13. This proves the required collision bound for one active input byte in all four orientations.

Two or more active bytes. Suppose the input byte weight is $t \ge 2$. For each nonempty exact output zero set $\mathcal{Z}$, let $p_{\mathcal{Z}}$ be its probability. Lemma 6 and $r_b \le p_b$ give the common bounds

$$0 \le p_{\mathcal{Z}} \le b_z := p_b^{z-1} r_b, \qquad z = |\mathcal{Z}| \in \{1, 2, 3\}.$$

All zero-set sums below range over $1 \le |\mathcal{Z}| \le 3$; the full zero set is impossible. Other impossible sets have probability zero. Put $B_0 = 4r_b + 6p_b r_b + 4p_b^2 r_b < 1$. Write $\sigma = \sum_{\mathcal{Z} \ne \emptyset} p_{\mathcal{Z}} \le B_0$.

First consider the activity projection, which replaces the law on each exact support by a uniform law of the same mass. Its collision therefore depends only on the support probabilities. Multiplying by Q_b^4 gives the quadratic on the left below. The bounds $\sigma^2 \le B_0 \sigma$ and $p_{\mathcal{Z}}^2 \le b_{|\mathcal{Z}|} p_{\mathcal{Z}}$ give

$$\begin{aligned}(1 - \sigma)^2 + \sum_{\mathcal{Z} \ne \emptyset} Q_b^{|\mathcal{Z}|} p_{\mathcal{Z}}^2 &\le 1 + \sum_{\mathcal{Z} \ne \emptyset} \left(B_0 - 2 + Q_b^{|\mathcal{Z}|} b_{|\mathcal{Z}|} \right) p_{\mathcal{Z}} \\ &\le (1 - B_0)^2 + \sum_{\mathcal{Z} \ne \emptyset} Q_b^{|\mathcal{Z}|} b_{|\mathcal{Z}|}^2 =: A_0.\end{aligned}$$

The coefficients in the first sum are positive: for $z \in \{1, 2, 3\}$,

$$Q_b^z b_z = Q_b r_b (Q_b p_b)^{z-1} > 2,$$

since $255(65/8192) > 2$ and $255/64 > 1$. Thus replacing each event probability by its cap is valid after passing to the displayed affine upper bound. For a short rational estimate, use $B_0 \geq 4r_b$, $1/128 < r_b < 1/125$, $p_b = 1/64$, and $Q_b < 256$:

$$\begin{aligned} A_0 &\leq 1 - 8r_b + r_b^2(16 + 4Q_b + 6Q_b^2 p_b^2 + 4Q_b^3 p_b^4) \\ &< 1 - \frac{1}{16} + \frac{1136}{15625} < \frac{51}{50}. \end{aligned}$$

The use of $B_0 < 1$ ensures $(1 - B_0)^2 \leq (1 - 4r_b)^2$.

It remains to sum the positive residual degrees. Write

$$x = Q_b r_b s^2 < 1/96, \qquad y = Q_b \Delta_b < 9/8, \qquad c = Q_b p_b^2 < 1/16, \qquad x < y.$$

For z prescribed zeros and $j \geq 1$ residual coordinates, Lemma 6, including the constant-nonzero factors, gives the squared response bound

$$Q_b^{-4} c^z x^{\min(t-z,j)} y^{j-\min(t-z,j)}.$$

When $z = 0$, at least two input symbols remain free: use x at $j = 1$ and $x^2 y^{j-2}$ at $j \geq 2$. When $z \geq 1$, every possible support leaves at least one free input, so xy^{j-1} suffices.

The ranges for distinct supports and residual-position sets are orthogonal, so their squared responses add. We bound the sum by marking positions, which avoids enumerating the possible support sizes. With no zero positions, the degree-one terms contribute at most $4x$. For higher degrees, mark a pair of residual positions. There are six choices, and each remaining position contributes either 1 or y. Every configuration is counted at least once, giving $6x^2(1 + y)^2$.

With both a zero and a residual position, mark one of each. The twelve ordered choices contribute cx; each remaining position contributes 1, c, or y. A configuration with z zeros and j residuals is counted $zj \geq 1$ times. Thus the total residual energy, multiplied by Q_b^4, is at most

$$\begin{aligned} S_0 &\leq 4x + 6x^2(1 + y)^2 + 12cx(1 + c + y)^2 \\ &< \frac{1}{24} + \frac{289}{98304} + \frac{1225}{32768} = \frac{2015}{24576} < \frac{1}{12}. \end{aligned}$$

Since $(256/255)^4 < 51/50$, activity and residual energy together give

$$q \operatorname{Col}(K_A e_a) \leq \frac{q}{Q_b^4}(A_0 + S_0) < \frac{51}{50}\left(\frac{51}{50} + \frac{1}{12}\right) = \frac{5627}{5000} < \frac{17}{15} \quad (t \geq 2). \tag{65}$$

Together with (64), this covers every nonzero input and every full-channel orientation K_A.

A nonzero column has mass one on the Q nonzero words, so subtracting its uniform part leaves squared norm equal to its collision minus $1/Q$. Consequently every residual orientation A satisfies

$$\|Ae_a\|_2^2 < \frac{17}{15q} - \frac{1}{q-1} < \frac{2}{15q} \quad (a \neq 0),$$

and its zero column vanishes. This proves the column assertion of Lemma 21. For every $A \in \mathcal{O}$ and its associated full channel K_A, we have

$$\mathrm{Col}(K_A e_a) < \frac{17}{15q} \quad (a \neq 0), \qquad \max_x \|Ae_x\|_2^2 < \delta. \tag{66}$$

Since $\mathcal{O}$ is closed under adjoints, the same energy bound holds for rows. □

Entries and the ordinary source. The entry estimate uses the differential and squared-correlation maxima of Keliher and Sui [37]. Their two distinct maxima also explain why the sharper source bound is used only for ordinary forward source laws.

Keliher and Sui [37] analyze the independent-round-key experiment and reduce its two-round maxima to the same superbox. For nonzero word labels their results give

$$\max_{a,b} K(b,a) \leq \frac{53}{4q}, \qquad \max_{a,b} (\mathcal{F}K\mathcal{F})(b,a) \leq \frac{109953193}{2^{22}q}. \tag{67}$$

The final invertible linear layer and byte permutation merely relabel the differences or masks. The second expression follows from the identity relating the Fourier transform of a difference channel to squared linear correlation. Transposition preserves both maxima.

Lemma 22 (Entries and source probabilities). *Every $A \in \mathcal{O}$ has maximum absolute entry below $27/q$. For $\mu = Ke_a$, $a \neq 0$,*

$$\mathrm{Col}(\mu) < \frac{17}{15q}, \qquad \|\mu\|_\infty < \frac{14}{q}. \tag{68}$$

Proof. The maxima in (67) are below $27/q$. Indeed, $53 < 108$ and $109953193 < 27 \cdot 2^{22}$. They cover the ordinary and simultaneous Fourier full channels; transposition covers their adjoints. At nonzero indices, subtracting P changes an entry $z \geq 0$ to $z - 1/Q$, whose absolute value is at most $\max\{z, 1/Q\} < 27/q$. The zero row and column of each residual vanish.

The ordinary forward source satisfies $\|\mu\|_\infty \leq 53/(4q) < 14/q$. Its collision bound is (66). The squared-correlation maximum is a different estimate and is not needed for this source law. □

7.3 The word operator norm

The byte-degree decomposition is orthogonal, so a matrix of block norm bounds controls the whole operator. We collect the estimates of Section 6 in a five-by-five nonnegative matrix and bound its norm by a positive-weight row test.

Lemma 23 (Byte-degree block comparison). *For $0 \le i, j \le 4$, the AES word residual satisfies $\sqrt{q}\,\|F_i R F_j\| \le O_{ij}$, where*

$$O = \begin{pmatrix} 6.3 & 0 & 4.601 & 3.3 & 1.61 \\ 0 & 1.16 & 2 & 1.61 & .113 \\ 4.601 & 2 & 1.45 & .113 & .008 \\ 3.3 & 1.61 & .113 & .008 & .00056 \\ 1.61 & .113 & .008 & .00056 & .00004 \end{pmatrix}. \tag{69}$$

Every decimal in this matrix is an exact rational upper bound.

Proof. The activity block gives 6.3 by (37). For the blocks between activity and degree two, (48) gives 4.601; for those between activity and degree three, Lemma 19 gives 3.3. Section 6.6 supplies all remaining entries from the activity-array estimates and the byte contraction factors. Those calculations also permit the mixed-map and Fourier orientations. □

Lemma 24 (Word norm). *Every $A \in \mathcal{O}$ satisfies $\|A\| < 52/(5\sqrt{q})$.*

Proof. Use the nonnegative symmetric matrix O in Lemma 23. For a word vector x, put $y_j = \|F_j x\|_2$. Orthogonality gives $\|y\|_2 = \|x\|_2$. The triangle inequality on each output block then gives

$$\sqrt{q}\,\|Rx\|_2 \le \|Oy\|_2 \le \|O\|\,\|x\|_2 .$$

For $v = (80, 15, 45, 29, 13)^T$, exact multiplication yields

$$\left(\frac{52}{5} I - O\right) v = (4.325, .441, 1.289, 8.12572, 4.32824)^T > 0.$$

The matrix $\operatorname{diag}(v)^{-1} O \operatorname{diag}(v)$ has every row sum below $52/5$. Its spectral radius is therefore below $52/5$. Similarity preserves the spectral radius, and for the symmetric nonnegative matrix O that radius equals its Euclidean operator norm. This proves the claim for R. Adjoint and unitary Fourier conjugation give the other three cases. □

7.4 The word Gram moment and total energy

The Gram moment is the squared Hilbert–Schmidt norm of $R^* R$. We estimate its blocks using the operator bounds above together with the block energies from Section 6. In each product, either factor can carry the Hilbert–Schmidt norm. A common choice handles most terms; changing that choice in two products gives the required constant.

Lemma 25 (Word Gram and total energy). *Every $A \in \mathcal{O}$ satisfies*

$$\operatorname{Tr}((A^* A)^2) < 90000/q^2, \qquad \|A\|_{\mathrm{HS}}^2 < 4251/q.$$

Proof. Use the operator comparison O of Lemma 23. The block energies (59), the activity energy, and Lemma 20 give a second comparison matrix H, now for Hilbert–Schmidt norms:

$$H = \frac{1}{10}\begin{pmatrix} 85 & 0 & 118 & 94 & 46 \\ 0 & 193 & 230 & 182 & 69 \\ 118 & 230 & 272 & 168 & 59 \\ 94 & 182 & 168 & 97 & 34 \\ 46 & 69 & 59 & 34 & 12 \end{pmatrix}, \qquad \sqrt{q}\,\|R_{ij}\| \le O_{ij}, \quad \sqrt{q}\,\|R_{ij}\|_{\mathrm{HS}} \le H_{ij}. \tag{70}$$

For $i+j > 0$, the squared block energy is at most $\binom{4}{i}\binom{4}{j}E_{i+j}/q$, by orthogonality of the marked positions. In particular, the degree-three entries use $4E_3 < 88 < 9.4^2$ and $24E_3 < 528 < 23^2$. The $(0,0)$ entry uses its activity-block energy below $71/q$.

The (i,j) Gram block is a sum over the common output degree k of the two factors. For $i \le j$, use the Hilbert–Schmidt bound on the factor with input degree i and the operator bound on the other. The triangle and ideal inequalities give

$$C = H^T O, \qquad q\,\|(R^*R)_{ij}\|_{\mathrm{HS}} \le C_{ij} = \sum_{k=0}^{4} H_{ki}O_{kj} \quad (i \le j). \tag{71}$$

The lower Gram blocks are adjoints of the upper ones. Hence, on putting

$$t_i = C_{ii}^2 + 2\sum_{j>i} C_{ij}^2,$$

their squared Hilbert–Schmidt norms sum to at most $\sum_i t_i/q^2$.

Five rational quadratic checks give

$$(t_0, t_1, t_2, t_3, t_4) < (33163, 22966, 33565, 4482, 68),$$

whose integer upper bounds sum to $94244 < 94250$. This first sum is slightly above the target 90000; we lower it by choosing the other factor for two Hilbert–Schmidt estimates below. Each entry of C is a dot product with at most five terms. For example,

$$t_0 = 146.2678^2 + 2\big(39.2538^2 + 57.3175^2 + 29.461176^2 + 13.784848^2\big) < 33163.$$

The block with input and output degrees both one has a small operator bound relative to its Hilbert–Schmidt bound. We use this in the Gram blocks $(1,2)$ and $(1,3)$: for $j \in \{2,3\}$, change the factor assignment only at $k = 1$ in the Gram block $(1,j)$. This lowers its upper bound by

$$g_j = H_{11}O_{1j} - O_{11}H_{1j}, \qquad g_2 = 11.92, \quad g_3 = 9.961.$$

The alternative ideal inequality uses

$$\|R_{11}\|\,\|R_{1j}\|_{\mathrm{HS}} \quad \text{in place of} \quad \|R_{11}\|_{\mathrm{HS}}\,\|R_{1j}\|\,.$$

Lowering a Gram-block bound from C_{1j} to $C_{1j} - g_j$ subtracts $2g_j(2C_{1j} - g_j)$ from the squared sum, including its adjoint block. Here

$$C_{12} = 19.3(2) + 23(1.45) + 18.2(.113) + 6.9(.008) = 74.0618 > 74,$$
$$C_{13} = 19.3(1.61) + 23(.113) + 18.2(.008) + 6.9(.00056) = 33.821464 > 33.8.$$

The two subtractions therefore exceed, respectively,

$$2(11.92)(148 - 11.92) = 3244.1472 > 3240,$$
$$2(9.961)(67.6 - 9.961) = 1148.284158 > 1140.$$

Combining the uniform assignment and these two changes proves

$$\begin{aligned} q^2 \operatorname{Tr}((R^* R)^2) &\le \sum_i t_i - 2\sum_{j=2}^{3} g_j(2C_{1j} - g_j) \\ &< 94250 - 3240 - 1140 = 89870 < 90000. \end{aligned}$$

For the total energy, the five row sums of squared entries of $10H$ are

$$32101, 128034, 172513, 80749, 11658.$$

Hence

$$q \, \|R\|_{\mathrm{HS}}^2 \le \sum_{i,j} H_{ij}^2 = \frac{425055}{100} = 4250.55 < 4251.$$

Adjoints and simultaneous Fourier conjugation preserve the Hilbert–Schmidt norm and the Gram moment, giving all $A \in \mathcal{O}$. □

7.5 One-sided Fourier moments

When only one side of a channel is Fourier transformed, its entries need not be nonnegative. The energy bounds still follow from the four orientations already treated. For fourth moments, the byte autocorrelation estimate (32) of Charpin, Helleseth, and Zinoviev [12], together with the byte estimates from Section 6 and Nyberg [48], supplies the necessary entry information. We first establish the energy bounds, then use this entry information in Lemma 27.

Lemma 26 (Energies after a one-sided transform). *For $A \in \mathcal{O}$ and $J = A\mathcal{F}$,*

$$\|J\| < \nu, \qquad \|J\|_{\mathrm{HS}}^2 < h, \qquad \max_x \sum_y |J_{xy}|^2 < \delta, \qquad \max_y \sum_x |J_{xy}|^2 < \delta.$$

Proof. Unitary multiplication preserves the operator and Hilbert–Schmidt norms. It also preserves every row norm when applied on the right. For a column, left unitarity gives $\|A\mathcal{F}e_y\|_2 = \|\mathcal{F}A\mathcal{F}e_y\|_2$. The latter matrix belongs to $\mathcal{O}$, so (66) applies. Adjoint column bounds give the row statement. □

Lemma 27 (Mixed fourth moments). *For every $A \in \mathcal{O}$, the map $J = A\mathcal{F}$ satisfies*

$$\max_x \sum_y |J_{xy}|^4 < r_4^2\delta, \qquad \max_y \sum_x |J_{xy}|^4 < r_4^2\delta, \qquad \sum_{x,y} |J_{xy}|^4 < s_4^2 h. \tag{72}$$

Proof. It suffices to treat $R\mathcal{F}$ and $R^*\mathcal{F}$. The other two maps are their adjoints, which exchange rows and columns and preserve the global fourth sum. We first bound entries when at least one word label has at least two active bytes. The remaining one-byte blocks are estimated through their row and column fourth sums. Combining the two estimates with the energy bounds proves the claim.

Put $r_b = 65/8192$, $p_b = 1/64$, and, for a nonzero byte mask v, define

$$g_v(x) = 16(D\mathcal{F}_b)_{xv}.$$

By (32), $g_v(0) = 1$ and $|g_v(x)| \le 1/8$ for $x \neq 0$. Fourier inversion expresses g_v as a nonnegative mixture of nontrivial additive characters: its weights are the entries of the nonzero column v of $\mathcal{F}_b D\mathcal{F}_b$, and they sum to one.

Fix a nonzero row b of K. Draw independent byte variables Y_i with laws $D(b_i, \cdot)$ and put $X = M^{-1}Y$. For a word mask a with active positions I, expansion of the channel gives

$$(K\mathcal{F})_{ba} = q^{-1/2}\mathbb{E}\prod_{i\in I} g_{a_i}(X_i). \tag{73}$$

The nonzero byte laws have collision at most r_b and point probabilities at most p_b. Since M^{-1} is MDS and $r_b \le p_b$, Lemma 6(a) gives, for distinct positions i, j,

$$\Pr[X_i = 0] \le r_b, \qquad \Pr[X_i = X_j = 0] \le p_b r_b. \tag{74}$$

The same lemma includes impossible zero events when too few active inputs remain. All probabilities retain their original mass.

If $|I| \ge 2$, retain two factors of (73) and bound the others by one. Writing $|g_v(x)| \le 1/8 + (7/8)\mathbf{1}_{x=0}$ gives

$$q|(K\mathcal{F})_{ba}| \le 65536\left\{\frac{1}{64} + 2\frac{1}{8}\frac{7}{8}r_b + \left(\frac{7}{8}\right)^2 p_b r_b\right\} = \frac{585713}{512} < 1144.$$

If $|I| = 1$ and b has at least two active bytes, expand g_v in its character mixture. Every coefficient of the appropriate row of M^{-1} is nonzero. Independence therefore gives at least two nontrivial byte biases, each of magnitude at most $1/8$, so $q|(K\mathcal{F})_{ba}| \le 1024$. For $a, b \neq 0$, $q(P\mathcal{F})_{ba} = -65536/Q$. Centering now gives

$$|(R\mathcal{F})_{ba}| < 1152/q \quad \text{unless both } a, b \text{ have one active byte.} \tag{75}$$

The same proof for $R^*\mathcal{F}$ uses the adjoint byte channel and M^{-1} as diffusion; the required autocorrelation and MDS properties hold in that case too. Zero rows and columns vanish.

It remains to estimate the blocks whose input and output labels each have one active byte. Here the shared intermediate byte gives a product of two byte

channels, with an invertible additive relabeling between them. We will use both the zero sums and the energy of this product to bound each row and column fourth sum. Let D_1, D_2 be the applicable ordinary or adjoint byte channels, $B_i = 256(D_i - P_b)$, $W = 16\mathcal{F}_b$, and let Π_L be an invertible additive relabeling. The byte norm and Gram diagonal give

$$\|B_i\| \le 18, \qquad \text{each row or column of } B_i \text{ has squared norm below } 264.$$

The same column-energy bound holds for $B_i\mathcal{F}_b$, because its Gram is a Fourier-conjugated allowed byte Gram. For $Z = B_1\Pi_L B_2 W$, every nonzero row or column $(z_a)_{a\neq 0}$ satisfies

$$|z_a| < 16 \cdot 264 = 4224, \qquad \sum_a z_a^2 < 256 \cdot 264 \cdot 18^2, \qquad \sum_a z_a = 0. \tag{76}$$

The entry bound is Cauchy–Schwarz on the two byte factors. For a row energy use $\|B_2 W\| \le 16 \cdot 18$; for a column energy use $\|B_1\| \le 18$ and the Fourier column bound for $B_2 W$. The zero sums follow because the residuals kill both the zero and constant directions.

Set $u_a = z_a/128$. Then $|u_a| < 33$, $\sum_a u_a^2 < 1344 < 33^2 + 16^2$, and $\sum_a u_a = 0$.

On a block where the input and output each have one active byte, write $N_i = 256 D_i$. The scaled full-channel entries are those of $N_1\Pi_L N_2 W$. Since $N_i = B_i + 256 P_b$, the cross terms vanish, and the remaining activity term is $65536 P_b W$. At nonzero indices it equals $-65536/255$. Word centering adds $65536/Q$. Thus the entries of $qR\mathcal{F}$ on each such block have the form

$$128(u_a - c), \qquad c = 512(1/255 - 1/Q), \qquad 0 < c < 33/16.$$

The centering shift contributes to the fourth moment even though the unshifted row sums to zero. We retain that cancellation and estimate the remaining quadratic, cubic, and quartic terms together. Define the increasing convex function

$$\phi_c(t) = t^2 + 4ct^{3/2} + 6c^2 t \qquad (t \ge 0).$$

Its value at zero is zero. For numbers in $[0, 33^2]$ with total at most $33^2 + 16^2$, their summed values under ϕ_c are at most $\phi_c(33^2) + \phi_c(16^2)$. To see this, transfer mass from a smaller positive coordinate to a larger one until one is zero or the other reaches 33^2. Convexity never decreases the sum. Repeating leaves at most one full coordinate and one remainder; monotonicity gives the two stated endpoints.

In the fourth-power expansion, the zero sum of the u_a removes the linear term. Bounding the cubic term by its absolute value and applying the preceding inequality to u_a^2 gives

$$\begin{aligned} \sum_a (u_a - c)^4 &\le \phi_c(33^2) + \phi_c(16^2) + 255c^4 \\ &= 33^4 + 16^4 + 4c(33^3 + 16^3) + 6c^2(33^2 + 16^2) + 255c^4 \\ &< 1252000 + 331000 + 35000 + 5000 \\ &= 1623000 < 25 \cdot 2^{16}. \end{aligned} \tag{77}$$

Each comparison follows by inserting $c < 33/16$; only integer powers of 33 and 16 remain. Therefore every exceptional byte-position block has scaled row and column fourth sums below $128^4 \cdot 25 \cdot 2^{16} = 25 \cdot 2^{44}$. The argument permits every intermediate additive relabeling L.

We now combine the pointwise estimate outside these blocks with their fourth-sum estimate. There are $4 \cdot 255 = 1020$ one-active-byte word labels. Each row or column meets at most four exceptional byte-position blocks. Outside these blocks, bound two of the four entry factors by (75) and sum the remaining squared entries using Lemma 26. Since $4 \cdot 25 \cdot 2^{44}/q = 409600$, each row or column fourth sum is less than

$$\frac{(2/15)1152^2 + 409600}{q^3} = \frac{586547.2}{q^3} < \frac{588000}{q^3} = \left(\frac{2100}{q}\right)^2 \delta.$$

For the global sum there are at most 1020 exceptional rows. The total energy is below $4251/q$, so

$$\sum_{x,y} |J_{xy}|^4 < \frac{1152^2 \cdot 4251 + 1020 \cdot 409600}{q^3} = \frac{6059311104}{q^3}$$
$$< \frac{6121440000}{q^3} = \left(\frac{1200}{q}\right)^2 h.$$

This proves all three fourth-moment inequalities. □

Equations (66), (62) and (72) give the local inputs used in Section 8 to apply the graph estimates of Section 5.4.

8 AES Diffusion Graph Bounds

We now apply the word estimates to the diffusion graphs in the one-step and two-step operators of Section 4.2. Fixing residual positions and activity vectors shortens each graph to an MDS code. For one step, an encoder estimate controls the resulting block. For two steps, the fiber estimates of Section 5.4 retain the middle constraints of both graphs. We first estimate each pair of activity choices, then combine them through the shared activity projection.

8.1 Graph coordinates and dual representations

Write $\xi = q^{-1/2}$ and $\eta = q/(q-1)$. We use the local estimates (62) and (72), proved in Section 7, in their stated orientations. In particular, one-sided Fourier maps require their own row and column energies and fourth moments. For the graph calculations we use the slightly larger quantities

$$\bar{\delta} = \frac{(183/500)^2}{q}, \qquad \bar{h} = \frac{(326/5)^2}{q}. \tag{78}$$

These round the column and total energies upward, so that $\delta < \bar{\delta}$ and $h < \bar{h}$. Their square roots are rational because $\sqrt{q} = 2^{16}$. The two comparisons reduce to

$$100000 < 100467, \qquad 106275 < 106276.$$

The parameters ν and τ retain their values in (60). Operator norms and second Gram moments are invariant under left or right unitary Fourier multiplication; the energy and entry estimates are used only in the orientations covered by Section 7.

In a composition of two diffusion graphs, the middle coordinates join the graphs, and the outside coordinates belong to the input or output of the composition. Fix one graph and mark i outside coordinates and ℓ middle coordinates for residual variation, with $N = i + \ell > 0$. On every unmarked coordinate use one of the unit activity vectors $e_0, q^{-1/2}\mathbf{1}_V$. Suppose a outside and μ middle activity coordinates use the latter vector. The other $8 - i - \ell - a - \mu$ coordinates are fixed to zero. The original graph has dimension four, so define the remaining dimension count by

$$d = 4 - (8 - i - \ell - a - \mu) = a + \mu + i + \ell - 4. \tag{79}$$

MDS shortening leaves dimension $\max(0, d)$. If $d \le 0$, the shortened graph contains only the zero point, which residual projection removes. If $d \ge N$, the marked marginal is uniform and residual projection again gives zero. We call the remaining condition $0 < d < N$ *strict shortening*. The projection onto the marked coordinates is then an additive $[N, d]$ MDS code. Its raw indicator has coefficient

$$\kappa_N(d) = q^{-(d+4-N)/2} = q^{-(a+\mu)/2}. \tag{80}$$

Indeed, each unrestricted activity coordinate contributes $q^{-1/2}$, and each marked graph point has one preimage after the other coordinates have been fixed.

Lemma 28 (The coefficient under duality). *For $0 < d < N$, either the marked code or its dual can be used in the same graph block. Replacing dimension d by $N - d$ replaces $\kappa_N(d)$ by $\kappa_N(N - d)$, with no additional norm factor.*

Proof. The normalized Fourier transform takes the raw indicator of an additive $[N, d]$ code to $q^{d-N/2}$ times the raw indicator of its dual. Therefore

$$\kappa_N(d) q^{d-N/2} = q^{(d-4)/2} = \kappa_N(N - d).$$

The transforms on the outside coordinates are unitary. Moreover, $R\mathcal{F} = \mathcal{F}(\mathcal{F}R\mathcal{F})$, so an outside residual map can be replaced by the permitted simultaneous Fourier orientation. □

Two adjacent graphs may choose their representations independently. If their choices are $f, g \in \{0, 1\}$, where one denotes the dual representation, their middle local map becomes $\mathcal{F}^f R \mathcal{F}^g$. When $f = g$ this is ordinary or simultaneous Fourier; when $f \ne g$ it is one-sided. Thus independent duality choices require the one-sided estimates of Section 7.5. The projection P_e is invariant under $\mathcal{F}$, since its orthogonal complement is spanned by e_0 and $\mathbf{1}_V$. Hence these changes preserve all middle residual projections.

8.2 One-step bounds

For a one-step block, let i and j count the output and input residual coordinates, and let a and b count their unrestricted activity coordinates. Put $N = i + j > 0$ and $d = a + b + N - 4$, where $0 \le a \le 4 - i$ and $0 \le b \le 4 - j$. If $0 < d < N$, use the smaller dimension

$$s = \min(d, N - d), \qquad I = \min(i, s), \qquad \ell = \min(s, i, j).$$

Here I is the rank of the output image. Its I information coordinates use the operator norm estimate; the remaining output coordinates use the column-energy estimate. The quantity ℓ describes the net power of q after the two fiber sizes and the activity coefficient are combined. Define the scaled step array by

$$F_{ij}^{(1)}(a, b) = q^{(i+j-2-\ell)/2} \nu^I \bar{\delta}^{(i-I)/2}. \tag{81}$$

If strict shortening fails, set this value to zero. In particular, every step with $i + j = 1$ is zero.

Lemma 29 (A fixed-position step bound). *The quantity in* (81) *bounds q times the indicated one-step block, including when exactly one of i, j is zero.*

Proof. For either code dimension $u \in \{d, N - d\}$, the row and column fibers of the marked incidence matrix have sizes $q^{\max(0,u-i)}$ and $q^{\max(0,u-j)}$. Its norm is at most the square root of their product. On the output image, Lemma 5 bounds the residual maps by $\nu^{\min(i,u)} \bar{\delta}^{(i-\min(i,u))/2}$. Multiply by $q\kappa_N(u)$. The resulting power of q is independent of the choice of representation, since

$$-u + \max(0, u - i) + \max(0, u - j) = -\min(u, N - u, i, j) = -\ell.$$

The remaining factor increases with $\min(i, u)$, since

$$\frac{\nu}{\sqrt{\bar{\delta}}} = \frac{52/5}{183/500} > 1.$$

Thus the smaller dimension s gives the smaller bound, namely (81). Lemma 28 permits either representation without a loss. A side of degree zero has a singleton image, so the same calculation covers it. If strict shortening fails, residual projection kills the block; for $N = 1$ no integer satisfies the strict condition. □

For an input in the full activity space, we can combine the activity choices before estimating the norm. The overlaps between their row supports control the norm of the whole activity-input block. On Ω let $P_A = P^{\otimes 4} - U$ be its centered projection, and use the degree projections P_i from Section 4.1.

Lemma 30 (An activity-input estimate). *With $b_1 = 0$, $b_2 = .0389$, $b_3 = .0134$, and $b_4 = .001728$,*

$$q^2 \, \|P_i T P_A\| \le b_i \qquad (1 \le i \le 4). \tag{82}$$

Proof. Put $c = (1-\xi)^{-1}$. For fixed output residual positions, a surviving activity choice produces a code of length i and dimension $0 < d < i$. By Proposition 2 its response has norm at most

$$q^{(i-4)/2}(\tau^2/q)^{d_*/2}\delta^{(i-2d_*)/2}, \qquad d_* = \min(d, i-d). \tag{83}$$

Here $q^{(i-4)/2}$ converts the activity coefficient times the raw indicator to its unit indicator. There is no surviving dimension for $i = 1$.

For degrees two and three, the nonzero block norms have a common upper bound. Factoring out this bound leaves a zero–one support matrix for comparison. Its row Gram counts shared input choices; the largest row sum of that Gram bounds the squared norm of the support matrix.

For $i = 2$, each nonzero entry is at most $\tau q^{-3/2}$. There are two output activity coordinates and four input activity coordinates; their unrestricted counts must sum to three. The four output rows have support sizes $4, 6, 6, 4$. The two rows of size six coincide, and the other supports are disjoint. The row Gram therefore has largest row sum 12. There are six orthogonal output residual-position choices. The activity frame has least Gram eigenvalue $(1-\xi)^6$, giving

$$\|P_2 T P_A\|^2 \le 72c^6\tau^2/q^3.$$

For $i = 3$, each entry is at most $\tau\sqrt{\delta}/q$. The two output rows have ten input choices each, with six in common. Their row Gram has largest row sum 16. Four residual-position choices and five activity coordinates give

$$\|P_3 T P_A\|^2 \le 64c^5\tau^2\delta/q^2.$$

For $i = 4$ there is one row. Input weights one and three give eight line choices; weight two gives six plane choices. Consequently

$$\|P_4 T P_A\|^2 \le c^4\big(8\tau^2\delta^2/q + 6\tau^4/q^2\big).$$

These comparisons use the scalar matrix of frame-block norms, so they hold for arbitrary activity coefficients.

Multiplying by q^4 and inserting $\tau = 300/q$ and $\delta = 2/(15q)$ gives, respectively,

$$\frac{6480000c^6}{q}, \qquad \frac{768000c^5}{q}, \qquad c^4\left(\frac{12800}{q} + \frac{48600000000}{q^2}\right).$$

To check the numerical constants, Bernoulli's inequality gives

$$c^6 \le \frac{1}{1-6\xi} = \frac{65536}{65530} < 1.0001.$$

Since $1 < c$, the same bound holds for its fourth and fifth powers. At $q = 2^{32}$ the three expressions are therefore below

$$.001509 < .0389^2, \qquad .000178832 < .0134^2, \qquad .000002984 < .001728^2.$$

For example, the first upper bound follows from

$$6480000(1.0001)/q = 810081/536870912 < .001509.$$

This proves (82). □

8.3 Two-step bounds with fixed activity choices

For two steps, we factor each graph into an outside map and a normalized sum over compatible middle labels. The two middle sums, with their local word maps between them, form the overlap operator (19). We can then apply the estimates of Section 5.4 to this common middle part.

The outside maps on the two sides differ because $T = H\Pi$. In $TP_{\geq 3}T$, one word layer acts at the output and the other between the diffusion graphs; there is no word layer at the input. Thus the left outside factor includes residual word maps, while the right outside factor consists only of the graph incidence and activity coefficient.

Normalized incidence and fiber averaging. Consider a marked $[i + \ell, s]$ representation with $s \leq \ell$. Its projection C onto the middle coordinates has dimension s, and its outside image O_C has dimension $I = \min(i, s)$. Every admissible outside value has q^{s-I} middle preimages. Let U_C and U_{O_C} be the coordinate inclusions, and let J_C be the normalized fiber map of Section 5.4, with fiber dimension $a_C = s - I$.

Lemma 31 (An identity for arbitrary input vectors). *Let $\mathcal{I}_C$ be the raw incidence matrix, with outside values indexing rows and middle assignments indexing columns. Then*

$$\mathcal{I}_C = q^{a_C/2} U_{O_C} J_C^* U_C^*. \tag{84}$$

After the outside residual maps and the activity coefficient have been included, its norm is at most

$$a_L(i, \ell, s) = q^{(i+\ell-4-I)/2} \nu^I \bar{\delta}^{(i-I)/2}, \qquad I = \min(i, s). \tag{85}$$

For a right graph with j outside residual coordinates and dimension $t \leq \ell$, the corresponding factor, without an outside word layer, is

$$a_R(j, \ell, t) = q^{(j+\ell-4-J)/2}, \qquad J = \min(j, t). \tag{86}$$

Proof. For an admissible outside value, J_C^* sums its q^{a_C} middle assignments with coefficient $q^{-a_C/2}$. Multiplication by $q^{a_C/2}$ gives exactly the entries of the raw incidence matrix. This proves (84) entry by entry, including for signed or complex vectors.

The image O_C has an information set of size I. Lemma 5, applied with $\bar{\delta}$, gives

$$\left\| R^{\otimes i} U_{O_C} \right\| \leq \nu^I \bar{\delta}^{(i-I)/2}.$$

Multiply this bound by $\kappa_{i+\ell}(s) q^{(s-I)/2}$ to obtain (85). On the right there is no outside word layer in $TP_{\geq 3}T$; the same calculation without this encoder gives (86). □

After applying this factorization on both sides, the middle operator is precisely (19). Its norm is multiplied by the outside factors (85) and (86). The identity holds on arbitrary input vectors, so these factors also apply to the signed intermediate errors in the recurrence.

Reduced codes and overlap estimates. We choose the smaller of each code and its dual so that the middle coordinates determine every graph point. This makes the incidence factorization applicable and restricts the possible fiber dimensions.

Let i, j be the outside residual degrees, with unrestricted activity counts a, b. The middle residual degree is ℓ; its two activity choices have unrestricted counts μ, λ. Thus

$$\begin{gathered} i, j \in \{0, \dots, 4\}, \quad 0 \le a \le 4 - i, \quad 0 \le b \le 4 - j, \\ \ell \in \{3, 4\}, \qquad \mu, \lambda \in \{0, \dots, 4 - \ell\}, \end{gathered}$$

and set

$$d_L = a + \mu + i + \ell - 4, \qquad d_R = b + \lambda + j + \ell - 4.$$

If $0 < d_L < i + \ell$ and $0 < d_R < j + \ell$, use the smaller of each code and its dual, retaining the original code on a tie:

$$\begin{aligned} s &= \min(d_L, i + \ell - d_L), \quad & f &= \mathbf{1}_{2d_L > i+\ell}, \\ t &= \min(d_R, j + \ell - d_R), \quad & g &= \mathbf{1}_{2d_R > j+\ell}. \end{aligned} \tag{87}$$

Both dimensions are positive and at most ℓ: for example, $s \le \lfloor (i + \ell)/2 \rfloor \le \ell$, since $i \le 4$ and $\ell \ge 3$. Thus the middle projection is injective and the fiber bounds apply. Lemma 28 preserves the graph operator under either choice.

Their outside fibers have dimensions $\alpha = s - \min(i, s)$ and $\beta = t - \min(j, t)$. In particular,

$$0 \le \alpha \le \ell - s, \qquad 0 \le \beta \le \ell - t. \tag{88}$$

Indeed, the first fiber dimension is zero if $s \le i$; otherwise $\alpha = s - i \le \ell - s$ by $2s \le i + \ell$. The right side is identical. Consequently each fiber dimension is at most the number of noninformation coordinates on that side. The first branch of Lemma 7 therefore applies to the information-set overlap below.

Let $C, D \subseteq V^\ell$ be the resulting middle codes. Put $u = \max(s, t)$ and $v = \min(s, t)$. For these fixed dimensions, denote by $\mathcal{E}_{fg}(\ell, s, t, \alpha, \beta)$ the smallest applicable upper bound below for (19). All include the normalized fiber maps on both sides. When the dimensions sum to at least ℓ, the information-set estimate applies. Equal Fourier choices also permit an entry-based Hilbert–Schmidt estimate. These give, respectively,

$$\nu^{s+t-\ell} \bar{\delta}^{\ell - (s+t+\alpha+\beta)/2} (\bar{h}/q)^{(\alpha+\beta)/2} \qquad (s + t \ge \ell), \tag{89}$$

$$p_{\mathrm{ent}}^{\ell-u} \bar{\delta}^{(u-v)/2} \bar{h}^{v/2} \qquad (f = g). \tag{90}$$

The powers in the first expression are nonnegative by (88).

When the dimension sum is smaller than ℓ, the thin-code estimates use fourth moments. For equal Fourier choices, the entry cap supplies these moments; for unequal choices, use the one-sided bounds (72). In the formulas below the moment parameters are therefore

$$(r_4, s_4) = \begin{cases} (p_{\mathrm{ent}}, p_{\mathrm{ent}}), & f = g, \\ (2100/q, 1200/q), & f \ne g. \end{cases} \tag{91}$$

For two lines, $s = t = 1$, $\ell \in \{3, 4\}$ and $\alpha, \beta \in \{0, 1\}$, also use, with $e = \ell - 2$,

$$\sqrt{\frac{\rho_\alpha \rho_\beta}{q^{\alpha+\beta}}}, \qquad (\rho_0, \rho_1) = (\bar{\delta} r_4^e, \bar{h} s_4^e). \tag{92}$$

For a line and a plane in four coordinates, use

$$\sqrt{\frac{\bar{\delta}^{2-\alpha} \bar{h}^\alpha \kappa_\beta}{q^{\alpha+\beta}}}, \qquad \alpha \in \{0, 1\}, \quad \beta \in \{0, 1, 2\}, \tag{93}$$
$$(\kappa_0, \kappa_1, \kappa_2) = (\bar{\delta} r_4^2, \bar{h} s_4^2, \bar{\delta}\bar{h}).$$

Its adjoint is included when the plane is on the left. These cases cover every pair: if $s + t < \ell$, positive dimensions leave only $(1, 1)$ for $\ell = 3$, and $(1, 1), (1, 2), (2, 1)$ for $\ell = 4$. These are the partial-sum formulas of Lemmas 10 and 11 with the enlarged resources. They include every fiber dimension through the common normalization.

Lemma 32 (Validity of the overlap choices). *For the reduced representations above, every applicable expression in* (89)–(93) *is an upper bound for the indicated normalized overlap, for arbitrary signed or complex inputs.*

Proof. For the first bound, (88) reduces the left normalized collision factor in Lemma 8 to $\bar{\delta}^{\ell-s-\alpha}(\bar{h}/q)^\alpha$, and similarly on the right. Multiplying and taking a square root gives (89). The two line and line–plane formulas are Lemmas 10 and 11, using (91). An entry cap supplies the necessary fourth moments when $f = g$; the other case uses the one-sided moments directly.

For (90), suppose $s \le t$. Omit the fiber isometries and let A be the raw overlap kernel. Choose t information coordinates of D. Bound the other squared entries by p_{ent}^2; summing over the right code leaves a product of t row energies. In the left code, sum s of these factors over an information set and bound the remaining $t - s$ by $\bar{\delta}$. Thus

$$\|A\|_{\mathrm{HS}}^2 \le p_{\mathrm{ent}}^{2(\ell-t)} \bar{\delta}^{t-s} \bar{h}^s.$$

Take a square root and reinsert the isometries, which cannot increase the norm. For $t \le s$, apply the same argument to the adjoint. The entry cap requires equal Fourier bits. All these are operator norm inequalities, so the sign of the input does not affect them. □

A whole-graph bound. The left graph ties the outside and middle labels to the same code point. Applying the Gram-moment estimate to all of its local maps together retains this relation before taking a norm. This gives a second bound for the composed block, alongside the product of the outside factors and the middle overlap. For the reduced representations, put

$$\mathcal{G} = q^{(i+j+2\ell-8-s-J)/2} \tau^s \bar{\delta}^{(i+\ell-2s)/2}, \qquad J = \min(j, t). \tag{94}$$

Lemma 33 (Whole-graph Gram bound). *For the reduced graph representations above, $\mathcal{G}$ bounds the same composed block as the incidence estimate.*

Proof. The left code has dimension $s \le (i+\ell)/2$. Proposition 2, including its mixed-map form, bounds the response of its unit indicator after the outside and middle local maps by

$$\tau^s q^{-s/2} \bar{\delta}^{(i+\ell-2s)/2}.$$

Converting the unit indicator to the graph matrix multiplies by $\kappa_{i+\ell}(s) q^{s/2} = q^{(i+\ell-4)/2}$. The resulting Hilbert–Schmidt bound also bounds its operator norm. Multiply by the right incidence factor $a_R = q^{(j+\ell-4-J)/2}$ from (86) to obtain (94). Every middle local map has been included exactly once, on the left. □

8.4 Combining the shared activity projection

It remains to combine the activity choices on the shared middle layer. For degree three, the one activity coordinate contributes a single projection whose matrix couples the choices of the two graphs. The one-word activity synthesis map $B = [e_0, q^{-1/2}\mathbf{1}_V]$ has inverse Gram matrix

$$(B^*B)^{-1} = \eta \begin{pmatrix} 1 & -\xi \\ -\xi & 1 \end{pmatrix}, \qquad P = B(B^*B)^{-1}B^*. \tag{95}$$

The two unit frame vectors have inner product ξ. The inverse Gram matrix accounts for this nonorthogonality: its two diagonal coefficients are η and its two cross coefficients are $-\eta\xi$. For middle degree three there is one such projection, so its absolute coefficient is $\eta\xi^{|\mu-\lambda|}$ for $\mu, \lambda \in \{0, 1\}$. Middle degree four has no activity coordinate.

For fixed outside counts a, b and middle choices ℓ, μ, λ, use (87) and the fiber dimensions α, β above. The whole-graph estimate and the incidence–overlap product bound the same block in this representation. Take the smaller of these two estimates:

$$\begin{aligned} &E_{ij;\ell}(a, b, \mu, \lambda) \\ &\quad = \min\left\{\mathcal{G},\ a_L a_R \mathcal{E}_{fg}(\ell, s, t, \alpha, \beta)\right\}. \end{aligned} \tag{96}$$

The incidence factors are (85) and (86). If either strict shortening condition fails, set $E_{ij;\ell} = 0$.

Proposition 3 (A fixed-position two-step bound). *Define*

$$G^{(2)}_{ij}(a, b) = \sum_{\ell=3}^{4} \binom{4}{\ell} \eta^{4-\ell} \sum_{\mu,\lambda=0}^{4-\ell} \xi^{|\mu-\lambda|} E_{ij;\ell}(a, b, \mu, \lambda). \tag{97}$$

It bounds the composition through $P_{\ge 3}$ between the indicated fixed outside residual positions and activity frame choices.

Proof. Expand $P_{\geq 3}$ by its middle residual positions. For each choice, the remaining activity projections give exactly (95). A pair of activity frame choices then produces the shortened dimensions above. Its incidence factorization is (84) on both graphs, so Lemmas 31 and 32 give the incidence–overlap product. The whole-graph bound of Lemma 33 estimates the same operator. Lemma 28 preserves the block norm under the chosen representations, so both estimates apply to the original block. Their minimum gives (96).

The triangle inequality sums the $\binom{4}{\ell}$ middle residual choices and the absolute activity metric coefficients. This is (97). The inverse metric occurs once because there is one shared projection. Finally, $SP_{\geq 3}S = TP_{\geq 3}T$: uniform vectors are killed by the middle projection and preserved by T. □

For fixed i, j, a, b, abbreviate $E_\ell^{\mu\lambda} = E_{ij;\ell}(a, b, \mu, \lambda)$. The scaled two-step array is

$$\begin{aligned} F_{ij}^{(2)}(a,b) &:= q^3 G_{ij}^{(2)}(a,b) \\ &= q^3\big[4\eta(E_3^{00} + \xi E_3^{01} + \xi E_3^{10} + E_3^{11}) + E_4^{00}\big]. \end{aligned} \tag{98}$$

The factor four counts the middle residual supports of size three. Their single activity projection contributes the four weights $\eta(1, \xi, \xi, 1)$; degree four contributes the final term. All coefficients are nonnegative. Thus any applicable overlap or whole-graph bound may replace the minimum when estimating a term; one need not determine which bound is smallest.

The outside activity vectors are still unit indicators at zero or unrestricted averages. Section 9.1 expands the nonzero activity vectors in this frame and sums the fixed-position bounds over the supports of each profile component.

9 Recurrence Coefficients and Initialization

The recurrence needs two propagation estimates and two initial estimates. For propagation, we sum the fixed-position bounds of Section 8 over the supports in each profile component. This gives one-step and two-step maps that apply to arbitrary signed errors. Initialization also uses the independent word laws produced by a fixed input difference: we estimate the four-round profile and the two-step term that contributes to the six-round error. Proposition 5 collects the four estimates, ready for the normalized recurrence of Section 10.2.

9.1 From fixed positions to profile components

Recall that a type (i, t) records the residual degree and the number of active words, as defined in Section 4.1. We use the order

$$\begin{aligned} &(0,1), (0,2), (0,3), (0,4), \quad (1,1), (1,2), (1,3), (1,4), \\ &(2,2), (2,3), (2,4), \quad (3,3), (3,4), (4,4). \end{aligned} \tag{99}$$

An index $\alpha \in \{1, \ldots, 14\}$ refers to its type in this list.

The nonzero activity vector has the expansion

$$e_+ = Q^{-1/2}\mathbf{1}_{V\setminus\{0\}} = \sqrt{\eta}\,(q^{-1/2}\mathbf{1}_V - \xi e_0), \qquad \xi = q^{-1/2}, \quad \eta = q/Q. \tag{100}$$

This gives a direct way to pass from graph bounds with unrestricted activity averages to blocks with prescribed nonzero supports. We separate the expansion of one fixed support from the number of possible supports. For a nonnegative graph-bound array F, define

$$\mathcal{K}_{it,jw}(F) = \sum_{a=0}^{t-i}\sum_{b=0}^{w-j}\binom{t-i}{a}\binom{w-j}{b}\xi^{t-i+w-j-a-b}F_{ij}(a,b), \tag{101}$$

$$\mathcal{B}_{it,jw}(F) = \sqrt{n_{i,t}n_{j,w}}\,\eta^{(t-i+w-j)/2}\mathcal{K}_{it,jw}(F). \tag{102}$$

Here $\mathcal{K}$ is the binomial expansion for fixed positions; $\mathcal{B}$ includes the activity normalization and the multiplicities $n_{i,t} = \binom{4}{t}\binom{t}{i}$.

Lemma 34 (Exact-support lifting). *Suppose $F_{ij}(a,b)$ bounds an operator block with fixed residual positions of degrees i,j, with a unrestricted activity coordinates at the output, b at the input, and all other activity coordinates zero. The bound is to hold for every such choice of positions. Then the block between types (i,t) and (j,w) has norm at most $\mathcal{B}_{it,jw}(F)$. If the input support is fixed, replace $n_{j,w}$ by $\binom{w}{j}$.*

Proof. Expand every constant-nonzero activity factor by (100). For fixed residual and support positions, the triangle inequality gives $\mathcal{K}$ in (101), together with the factor $\eta^{(t-i+w-j)/2}$. Its binomial coefficients count the choices of unrestricted activity factors.

There are $n_{i,t}$ orthogonal output blocks and $n_{j,w}$ orthogonal input blocks. If every block has norm at most c, the entire block matrix has norm at most $c\sqrt{n_{i,t}n_{j,w}}$, by Cauchy–Schwarz on the input block lengths. Fixing the input support leaves only the $\binom{w}{j}$ choices of residual positions. □

For example, suppose there is one nonzero activity coordinate on each side. Expanding those two coordinates by (100) and taking absolute coefficients gives

$$\eta\big(F_{ij}(1,1) + \xi F_{ij}(1,0) + \xi F_{ij}(0,1) + \xi^2 F_{ij}(0,0)\big).$$

Here the four terms are the special case $t-i = w-j = 1$ of $\mathcal{K}$, before the support multiplicity is applied.

The middle activity coordinate in a two-step composition through degree three is shared by both graphs. It must first be contracted with the inverse activity metric, as derived in (95). Only then do we apply the outside-support expansion above. Degree four has no middle activity coordinate.

9.2 Rational propagation maps

Let $\mathsf{D}_{14} = \mathrm{diag}(d_1,\dots,d_{14})$, where the positive scales d_α are in Table 2. Dividing each component length by its scale balances the different sizes of the components. Set $\omega = 2^{-41}$.

Table 2. Scales and initial two-step bounds. The type order is (99); $\mathsf{D}_{14} = \mathrm{diag}(d_\alpha)$. The four-round vector is $u_0 = \mathbf{1} + 9e_8 + 2e_{13} + 129e_{14}$.

α	Type	d_α	g_α
1	$(0,1)$	.815	1
2	$(0,2)$	1	3
3	$(0,3)$	.815	450
4	$(0,4)$	.815/32768	450
5	$(1,1)$	.00247	1
6	$(1,2)$	.00138	25
7	$(1,3)$	.00138	900
8	$(1,4)$	.00247	15000
9	$(2,2)$	.000243	1
10	$(2,3)$	.000178	25
11	$(2,4)$	.000243	450
12	$(3,3)$	.0000189	2
13	$(3,4)$	.0000189	10
14	$(4,4)$	.0000000777	5

The scale ratio in the estimates below comes from measuring component α in units of d_α. For one input component,

$$\frac{\|P_\alpha S P_\beta f\|_2}{d_\alpha} \le \left(\frac{d_\beta}{d_\alpha}\,\|P_\alpha S P_\beta\|\right)\frac{\|P_\beta f\|_2}{d_\beta}.$$

Thus the coefficient in parentheses is the transition bound in scaled coordinates. The maps L and M below also remove the target decay factors ω and ω^2 from the one-step and two-step bounds, respectively.

Table 2 also specifies the two initial upper vectors. The vector u_0 controls the four-round error, and g controls the initial two-step term at six rounds. Section 9.3 proves their bounds. Their normalization, including the common factor 1000, is stated in Proposition 5.

Use the scaled graph arrays from (81) and (98). The step array is rational. To obtain a rational upper bound for the two-step array, replace each occurrence of $\sqrt{(2100/q)(1200/q)}$ in an overlap candidate by $1600/q$ before taking the minimum. Denote the resulting array by $\widehat{F}^{(2)}$. The inequality $2100 \cdot 1200 < 1600^2$ increases each affected candidate; minimum and positive summation preserve this order. Hence $F^{(2)} \le \widehat{F}^{(2)}$. All other graph square roots are rational by (78) and $\sqrt{q} = 2^{16}$. Applying Lemma 34 to these arrays gives

$$\begin{aligned} q\,\|P_{i,t} S P_{j,w}\| &\le \mathcal{B}_{it,jw}(F^{(1)}) \quad (i+j>0),\\ q^3\,\|P_{i,t} S P_{\ge 3} S P_{j,w}\| &\le \mathcal{B}_{it,jw}(\widehat{F}^{(2)}). \end{aligned} \tag{103}$$

For a fixed input support replace $n_{j,w}$ by $\binom{w}{j}$ in (102); denote the result by $\mathcal{B}^{[w]}_{it,jw}(F)$. This retains only the residual-position choices inside that support.

We use one common rational rounding to define the recurrence maps. For the four possible support counts, put

$$\big(r(1), r(4), r(6), r(12)\big) = (1, 2, 2.4495, 3.4642). \tag{104}$$

The four inequalities $\sqrt{m} \le r(m)$ follow by squaring. Also $\sqrt{\eta} \le \eta$, since $\eta > 1$. Consequently

$$\begin{aligned} &\mathcal{B}_{it,jw}(F) \le \widehat{\mathcal{B}}_{it,jw}(F), \\ &\widehat{\mathcal{B}}_{it,jw}(F) := r(n_{i,t})r(n_{j,w})\eta^{t-i+w-j}\mathcal{K}_{it,jw}(F). \end{aligned} \tag{105}$$

This substitution applies to every type pair and preserves the nonnegative support sum.

For two activity components, use the exact-support counts $N_r(q)$ from (34) at alphabet size q. They count graph points on each specified support of size r, independently of the positions, because the word-scale diffusion graph is again an $[8, 4]$ MDS code. For each pair of support sizes, the normalized centered entry is

$$Q^{-(t+w)/2}N_{t+w}(q) - Q^{(t+w)/2}/(q^4 - 1).$$

It is independent of the particular supports. Hence the squared, q-scaled activity block norm is

$$q^2\binom{4}{t}\binom{4}{w}Q^{-t-w}\left(N_{t+w}(q) - \frac{Q^{t+w}}{q^4-1}\right)^2. \tag{106}$$

Taking the nonnegative square root and making the same two roundings gives the following rational upper bound on the q-scaled activity block norm:

$$\begin{aligned} \widehat{B}^{\mathrm{act}}_{t,w} := r\left(\binom{4}{t}\right) r\left(\binom{4}{w}\right) \eta^{t+w} q^{1-(t+w)/2} \\ \cdot \left| N_{t+w}(q) - \frac{Q^{t+w}}{q^4-1} \right|. \end{aligned} \tag{107}$$

For input degree zero and positive output degree, we also have the bound b_i/q from Lemma 30. The centered projection can be inserted on the input because $P_{i,t}U = 0$.

For output type (i, t) at index α and input type (j, w) at index β, define

$$\begin{aligned} \widehat{B}^{\mathrm{step}}_{\alpha\beta} &:= \begin{cases} \widehat{B}^{\mathrm{act}}_{t,w}, & i = j = 0, \\ \min\left\{\widehat{\mathcal{B}}_{it,0w}(F^{(1)}), b_i/q\right\}, & i > 0,\ j = 0, \\ \widehat{\mathcal{B}}_{it,jw}(F^{(1)}), & j > 0, \end{cases} \\ \widehat{B}^{\mathrm{ret}}_{\alpha\beta} &:= \widehat{\mathcal{B}}_{it,jw}(\widehat{F}^{(2)}). \end{aligned} \tag{108}$$

The support and activity estimates prove directly that

$$\begin{aligned} q\,\|P_\alpha S P_\beta\| &\le \widehat{B}^{\mathrm{step}}_{\alpha\beta}, \\ q^3\,\|P_\alpha S P_{\ge 3} S P_\beta\| &\le \widehat{B}^{\mathrm{ret}}_{\alpha\beta}. \end{aligned} \tag{109}$$

We now express these component bounds in the scaled coordinates and remove the factors ω and ω^2. In the one-step map, only columns $1, \dots, 11$ contribute: they are the input components of residual degree at most two in (13). Degrees three and four are handled through the two-step composition. Define the two nonnegative linear maps by

$$\begin{aligned} \mathsf{L}_{\alpha\beta} &= 2^9 \frac{d_\beta}{d_\alpha} \widehat{B}^{\mathrm{step}}_{\alpha\beta} && (1 \le \beta \le 11), \\ \mathsf{L}_{\alpha\beta} &= 0 && (12 \le \beta \le 14), \\ \mathsf{M}_{\alpha\beta} &= 2^{-14} \frac{d_\beta}{d_\alpha} \widehat{B}^{\mathrm{ret}}_{\alpha\beta} && (1 \le \beta \le 14). \end{aligned} \tag{110}$$

Here $2^9 = (q\omega)^{-1}$ and $2^{-14} = (q^3\omega^2)^{-1}$. Thus the definitions remove exactly the required decay factors from the two support bounds. Every row is evaluated from its own input and output types by the same formulas. All terminating decimals in those formulas and in Table 2 are exact rational numbers.

Worked propagation coefficients. The following examples show how the graph dimensions and activity choices determine a coefficient. The first expands two outside activity coordinates. The second includes a nontrivial shared middle projection, and the third isolates its degree-three and degree-four contributions. Section 10.3 applies the resulting maps to the complete profiles.

Put $c = 52/5$ and $e = 183/500$, so $\nu = c/\sqrt{q}$ and $\sqrt{\delta} = e/\sqrt{q}$. The fixed-position step formula (81) and five-term two-step formula (98) were proved in Lemma 29 and Proposition 3. We now apply them with the support lift.

A one-step example with two activity coordinates. Take output type $(4, 4)$ and input type $(2, 4)$, which have indices 14 and 11. There is no output activity coordinate, and the input has two. The three values $b = 0, 1, 2$ in (81) give

$$F^{(1)}_{42}(0, 0) = F^{(1)}_{42}(0, 2) = A/q, \qquad F^{(1)}_{42}(0, 1) = B/q,$$

where

$$A = c^2e^2 = 14.48868096, \qquad B = c^3e = 411.700224.$$

Hence the support sum is h/q, where

$$\mathsf{h} = A(1 + q^{-1}) + 2Bq^{-1/2} = \frac{24329052059229177}{1677721600000000} < 14.501246.$$

The support multiplicity is $n_{4,4}n_{2,4} = 6$. Using the common rational lift from (105) gives

$$q\widehat{B}^{\mathrm{step}}_{14,11} = 2.4495\,\eta^2\mathsf{h} < 35.522.$$

Here $2.4495^2 > 6$ and $\eta < 1.000000001$. Thus this pair contributes exactly $2^9(d_{11}/d_{14})\widehat{B}^{\mathrm{step}}_{14,11}$ to the step map.

A two-step coefficient with an outside-support sum. The pair $(\alpha,\beta)=(6,3)$ has types $(1,2)$ and $(0,3)$. It has one output activity coordinate and three input activity coordinates. Define the eight nonnegative terms in its outside-support sum by

$$W_{ab}=\binom{3}{b}\xi^{4-a-b}\widehat{F}_{10}^{(2)}(a,b),\qquad 0\le a\le 1,\quad 0\le b\le 3.$$

Using the dimension rule gives the rational upper bounds

	$b=0$	$b=1$	$b=2$	$b=3$
$10^9W_{0b}\le$	0	1	1	2222
$10^9W_{1b}\le$	0	1	13	4890777.

For example, write $h_0=326/5$, so $\sqrt{\overline{h}}=h_0/\sqrt{q}$. For $(a,b)=(1,3)$, the middle choices (ℓ,μ,λ) equal to $(3,0,0)$, $(3,1,0)$, and $(4,0,0)$ use the line bound, the large-code overlap bound, and the adjoint line–plane bound, respectively. The fixed dimension rule gives $(s,t,f,g)=(1,1,0,1)$ for the first and $(2,1,0,1)$ for the other two. Including their metric weights gives

$$W_{13}\le\frac{ce}{q}\left[4\eta(1600h_0+\xi h_0^2)+1200h_0^2\right]<.004890777.$$

Here the 1600 uses $\sqrt{2100\cdot 1200}<1600$; the other substitutions are rational. Summing the eight caps gives $\sum_{a,b}W_{ab}\le .004893015$. The support multiplicity is $n_{1,2}n_{0,3}=48$. The rational lift therefore gives

$$\widehat{B}_{6,3}^{\mathrm{ret}}\le 2(3.4642)\,\eta^4(.004893015)<.033901.$$

The resulting two-step coefficient is $2^{-14}(d_3/d_6)\widehat{B}_{6,3}^{\mathrm{ret}}$. This example retains all eight outside terms and the shared middle metric; their positive coefficients permit the same summation for any nonnegative profile.

A two-step coefficient with both middle degrees. Keep $c=52/5$ and $e=183/500$ and use the five-term array (98). Take $i=j=4$ and $a=b=0$. At middle degree three, each shortened dimension is either 3 or 4. Choose dimension 3 on both sides, using the dual representation precisely when the original dimension is 4. Then $I=J=3$, the fiber dimensions are zero, and the large-code overlap bound is ν^3. Its product with the incidence factors is $\nu^6\sqrt{\delta}$ for all four activity choices. Their total metric coefficient is $8\eta(1+\xi)$. At middle degree four, choose $s=t=4$; the same overlap bound gives ν^8 and there is no activity metric. Therefore

$$\begin{aligned}F_{44}^{(2)}(0,0)&\le 8\eta(1+\xi)c^6e\xi+c^8/q\\&=\frac{2412014379008}{42666015625}+\frac{815730721}{25600000000}\\&<56.533+.032<57.\end{aligned}$$

The selected large-overlap bounds contain no mixed fourth-moment root. Rounding therefore leaves them unchanged, so the same upper bound also holds for $\widehat{F}_{44}^{(2)}(0,0)$. There are no outside activity coordinates or support multiplicities in this case. The corresponding map entry is consequently below $57/2^{14}$ by (110). This example shows where the four degree-three metric entries and the separate degree-four term enter the same two-step coefficient.

9.3 Initial profile and two-step source estimate

For initialization we can use the law produced by a fixed input difference. Its active words become independent after the first word layer. Each has a full-column law $\mu = Ke_a$, $a \neq 0$, with

$$\mathrm{Col}(\mu) \leq \rho := \frac{17}{15q}, \qquad \|\mu\|_\infty \leq p_{\mathrm{src}} := \frac{14}{q}. \tag{111}$$

The collision estimate is the full-column bound (66), proved for every nonzero input word in Section 7.2. The required point estimate is the superbox specialization [20] of the two-round differential bound of Keliher and Sui [37], namely $\|\mu\|_\infty \leq 53/(4q) < 14/q$. The full-column and residual-column estimates serve different parts of the calculation. The source law is a probability distribution, with $\mathrm{Col}(\mu) = Q^{-1} + \|Re_a\|_2^2$. Thus the stated collision constant uses the full-column proof; adding Q^{-1} to the residual upper bound δ would give a larger value. The point bound above applies to ordinary forward sources; the Fourier entry bound p_{ent} remains separate.

We apply Lemma 6, whose proof in Section 5.3 retains the mass of each restriction. It follows the marginal approach of Dinur [21]. Feed w independent nonzero laws, each with collision at most ρ and maximum probability at most p_{src}, through an additive four-coordinate MDS map. For disjoint output coordinate sets Z, I, with $|Z| = z < w$ and $|I| = i > 0$, put

$$h_Z(y) = \Pr[Y_I = y,\ Y_Z = 0,\ \mathcal{A}],$$

where $\mathcal{A}$ is any further indicator event on the input tuple. With $k = \min(w - z, i)$, its next residual response satisfies

$$\left\|R^{\otimes i} h_Z\right\|_2^2 \leq p_{\mathrm{src}}^{2z} (\rho\nu^2)^k \delta^{i-k}. \tag{112}$$

Lemma 6 uses the residual column parameter $\rho - Q^{-1} < \delta$, so replacing it by δ gives exactly (112). The specified zero event is impossible for $z \geq w$. For one, two, or three specified zeros, its probability is at most $\rho, p_{\mathrm{src}}\rho, p_{\mathrm{src}}^2\rho$, respectively. These are upper bounds that cover every possible w; some cases vanish or admit a smaller bound. No probability in this statement is renormalized after a restriction.

The four-round profile. At four rounds, a second word layer acts on the dependent outputs of diffusion. For a positive residual degree we use the restricted-source estimate above, with a common-source refinement when all four outputs come from one active word. The activity components depend only on the probabilities of their exact supports.

Proposition 4 (Four-round component lengths). *For every nonzero input difference, the four-round AES error satisfies*

$$q^{5/2}\mathsf{D}_{14}^{-1}\mathbf{v}(f_4) \le 1000u_0, \qquad u_0 = \mathbf{1} + 9e_8 + 2e_{13} + 129e_{14}. \tag{113}$$

Proof. We prove the stronger bound with $900u_0$ in place of $1000u_0$. Use the norm, column and Gram estimates of Section 7 and the ordinary-source bounds (111). Let the input have w active words. After the first word layer these are independent source laws of the form Ke_a, each supported on nonzero words. Write $L_{it} = q^{5/2}\left\|P_{i,t}f_4\right\|_2$.

For type (i,t) with $i > 0$, set $z = 4 - t$. Fix its zero positions and residual positions. The remaining coordinates must be nonzero; this is an allowed additional indicator in (112), because K preserves whether a word is zero. Each constant-nonzero coordinate contributes Q^{-1} in squared norm. Distinct support and residual-position choices are orthogonal. Thus, if $d = w - z > 0$ and $k = \min(d, i)$,

$$q^5\left\|P_{i,t}f_4\right\|_2^2 \le n_{i,t}q^5Q^{-(t-i)}p_{\mathrm{src}}^{2z}(\rho\nu^2)^k\delta^{i-k}. \tag{114}$$

If $d \le 0$, this source has no contribution to the type. Subtracting u does not change a positive-residual component.

Put

$$\theta = q/Q < 1.000001, \qquad A = 45968/375 < 123, \qquad B = 2/15.$$

The right side of (114) is

$$n_{i,t}\theta^{t-i}q^{t-3-k}14^{2(4-t)}A^kB^{i-k}. \tag{115}$$

Increasing k multiplies this expression by

$$\frac{\rho\nu^2}{\delta} = \frac{A}{Bq} = \frac{22984}{25q} < 1.$$

The ratio is below one, so the smallest possible k gives the largest estimate. For every proper support with $1 \le i \le t < 4$, use $k \ge 1$, $n_{i,t} \le 12$ and $B < 1$ to obtain

$$\begin{aligned} L_{it}^2 &\le n_{i,t}\theta^{t-i}AB^{i-1}(196/q)^{4-t} \\ &\le \frac{12\theta^3A\cdot 196}{q} < \frac{300000}{q} < .01^2. \end{aligned}$$

Here $\theta^3 < 1.000004$ and $q > 4.2\cdot 10^9$ suffice. The smallest target $1000d_\alpha(u_0)_\alpha$ among these types is .0189, and $.01 < .9(.0189)$.

For full support with one residual coordinate, (115) gives

$$L_{14}^2 \le 4\theta^3 A < 4(1.000004)(123) < 22.2^2.$$

Its target 24.7 satisfies $22.2 < .9(24.7)$.

For full support with $i \ge 2$, a single active input supplies only one independent source label. We retain that common label in the Gram estimate. First suppose $w = 1$. All four words entering the last layer are bijective images of one source label. Apply Corollary 2 to the selected residual maps, absorbing these bijections into their input labels. Their column energies and second Gram moments are unchanged. The other $4 - i$ positions contribute $Q^{-(4-i)}$ in squared norm, giving

$$\begin{aligned} L_{i4}^2 &\le \binom{4}{i} q^5 Q^{i-4} p_{\mathrm{src}}^2 \tau^2 \delta^{i-2} \\ &= \binom{4}{i} \theta^{4-i} B^{i-2} \frac{17640000}{q}. \end{aligned} \tag{116}$$

With at least two active input words, the restricted-source estimate already gives the required bound: use $k \ge 2$ in (115). The resulting bound replaces 17640000 by A^2, which is smaller. Thus (116) holds for every input support.

Define

$$h_i = \sqrt{\binom{4}{i}} B^{(i-2)/2} \qquad (2 \le i \le 4).$$

Since $\theta^2 < 1.000003$ and $q > 4.2 \cdot 10^9$,

$$\frac{17640000\theta^2}{q} < \left(\frac{13}{200}\right)^2,$$

$$17640000(1.000003) < 17745000 = 4.2 \cdot 10^9 \left(\frac{13}{200}\right)^2.$$

Hence $L_{i4} < (13/200)h_i$. For α the index of $(i, 4)$, the full-support targets of degrees two, three and four satisfy the common lower bound

$$1000 d_\alpha(u_0)_\alpha > \frac{3}{40} h_i,$$

because its three squared comparisons are

$$.243^2 > \frac{27}{800}, \qquad .0567^2 > \frac{3}{1000}, \qquad .010101^2 > \frac{1}{10000}.$$

Their normalized lengths are therefore below $(13/200)/(3/40) = 13/15 < .9$.

It remains to treat residual degree zero. The activity projection is constant on each exact support, so its error is the difference between the actual and uniform support probabilities. For a proper support, the actual exact-support probability is at most $\rho p_{\mathrm{src}}^{3-t}$. The ideal probability $Q^t/(q^4 - 1)$ is smaller than q^{t-4} and

hence below the same cap. The absolute difference is therefore at most that cap. Orthogonality of support patterns gives

$$\begin{aligned} L_{0t} &\le \sqrt{\binom{4}{t}}\theta^{t/2}\frac{17}{15}\left(\frac{14}{\sqrt{q}}\right)^{3-t} \\ &\le \sqrt{6}\theta^{3/2}\frac{17}{15} < 3 \qquad (1 \le t < 4). \end{aligned} \tag{117}$$

For the last step, use $\sqrt{6} < 5/2$ and $\theta^{3/2} < \theta^2 < 1.000003$. These activity targets are at least 815, with $3/815 < .9$.

For full support, each actual single-zero probability is at most ρ; the ideal single-zero probability is $(q^3-1)/(q^4-1) < 1/q < \rho$. Both all-active probabilities lie in $[1-4\rho, 1]$, so

$$L_{04} \le \frac{68\theta^2}{15\sqrt{q}} < \frac{5}{\sqrt{q}}.$$

Its target is $815/32768 = 1630/\sqrt{q}$, and $5/1630 < .9$. Every component is thus below .9 times its target, proving the stronger initialization bound and hence (113). □

The initial two-step term. The six-round decomposition (13) also contains a two-step term acting on the two-round source $\mu_2 = He_a$. This source has the fixed input support and independent active words. Its residual-degree lengths can therefore be estimated before the composed operator is applied.

Lemma 35 (Source lengths on a fixed support). *For a two-round source with w active input words, a componentwise upper bound on its $q^{5/2}$-scaled residual-degree profile is*

$$(b_2^{[w]})_j = q^{5/2}\sqrt{\binom{w}{j}\delta^j Q^{-(w-j)}}, \qquad 0 \le j \le w. \tag{118}$$

Proof. Each active column is the sum of its uniform-nonzero part, of squared norm Q^{-1}, and its residual column, of squared norm at most δ. The degree-j expansion has $\binom{w}{j}$ orthogonal residual-position choices. Multiplying the corresponding squared norms and adding these choices proves the claim. □

The degree-zero term includes the uniform part of the source. Since $SP_{\ge 3}S$ annihilates u, its action on μ_2 equals its action on f_2. We may therefore use the full-source lengths in the next estimate.

Let $C_{\alpha j}^{[w]}$ be the fixed-input-support, q^3-scaled two-step bound $\mathcal{B}_{it,jw}^{[w]}(\widehat{F}^{(2)})$ above. The triangle inequality on the actual source components gives

$$q^{5/2}\omega^{-1}\mathsf{D}_{14}^{-1}\mathbf{v}(SP_{\ge 3}Sf_2) \le \frac{\mathsf{D}_{14}^{-1}C^{[w]}b_2^{[w]}}{q^3\omega}. \tag{119}$$

Table 3. Common bounds for the two source summands in (120), after division by g_α. The row condition refers to the output type (i, t). The first class consists of the partially active cells in which every active word carries a residual component. The source roots are bounded by 1.000001 and .366.

Type	$j = 0$	$j = 1$	Sum
$i = t < 4$	.036	.951	.987
Otherwise	.007	.988	.995

The normalization is relative to the four-round starting profile. This term contributes at six rounds, one two-round increment later, so it carries one factor of ω^{-1}. Section 10.2 uses the same normalization for the full error.

For $w = 1$, the remaining comparisons with g are the following fourteen explicit inequalities:

$$\frac{64}{125 d_\alpha}\left(C^{[1]}_{\alpha 0}\sqrt{q/Q} + C^{[1]}_{\alpha 1}\sqrt{2/15}\right) \le g_\alpha, \qquad 1 \le \alpha \le 14. \tag{120}$$

Both square roots can be replaced by rational upper bounds: $\sqrt{q/Q} < 1.000001$ and $\sqrt{2/15} < .366$. The two source-root roundings are the rational comparisons

$$\frac{q}{q-1} < (1.000001)^2, \qquad \frac{2}{15} < .366^2;$$

the latter is again $100000 < 100467$.

Table 3 groups the output types into two classes. The first has residual variation on every active word of a proper support; the second contains the remaining types. Each row bounds the two source summands after division by the target component.

Lemma 36 (One-source comparison). *The one-source comparisons* (120) *hold for every output type.*

Proof. Use Proposition 3 for the five middle terms and (102) with fixed input support for the outside sum. Inserting the two source-root bounds above reduces the comparison to the table entries $c_{\alpha 0}, c_{\alpha 1}$:

$$\left(\frac{64 \cdot 1.000001}{125 d_\alpha g_\alpha}\right)^2 \left(C^{[1]}_{\alpha 0}\right)^2 \le c^2_{\alpha 0}, \quad \left(\frac{64 \cdot .366}{125 d_\alpha g_\alpha}\right)^2 \left(C^{[1]}_{\alpha 1}\right)^2 \le c^2_{\alpha 1}.$$

These are rational inequalities: the graph candidates use $\widehat{F}^{(2)}$, and squaring removes the support-count square roots in the fixed-support lift. Substitution gives the caps in Table 3. Taking nonnegative square roots and adding the two summands gives at most .987 for the first class and .995 for the second. Both are below one, proving (120). □

For a complete example of the first class, take output type $(1,1)$, at index 5. Write $c = 52/5$, $e = 183/500$, and $h_0 = 326/5$ as in the coefficient examples above. There are no outside output activity coordinates. For input degree zero, the outside expansion has two terms, but $\widehat{F}^{(2)}_{10}(0,0) = 0$ by strict shortening. In the remaining term, with outside counts $(0,1)$, only the middle choices $(3,1,1)$ and $(4,0,0)$ survive. Both use two lines with equal Fourier choices and fiber dimensions $(0,1)$. The line estimate (92), multiplied by the outside factors and by q^3, gives respectively $27ceh_0/q$ and $27^2ceh_0/q$. Their middle weights are 4η and 1, so

$$\widehat{F}^{(2)}_{10}(0,1) \le \frac{ceh_0}{q}(108\eta + 729).$$

For input degree one, the outside counts are $(0,0)$. The same two middle choices survive, now with fiber dimensions $(0,0)$. Using the equal-Fourier Hilbert–Schmidt estimate (90) gives the two scaled terms $27^2ch_0/q$ and $27^3ch_0/q$. Hence

$$\widehat{F}^{(2)}_{11}(0,0) \le \frac{ch_0}{q}(2916\eta + 19683).$$

The output multiplicity is $n_{1,1} = 4$, while each fixed-input multiplicity is $\binom{1}{j} = 1$. Thus the support lift is

$$C^{[1]}_{50} = 2\sqrt{\eta}\,\widehat{F}^{(2)}_{10}(0,1), \qquad C^{[1]}_{51} = 2\widehat{F}^{(2)}_{11}(0,0).$$

Insert these bounds into (120), using $d_5 = .00247$ and $g_5 = 1$. The two source roots, together with the factor $\sqrt{\eta}$ in the first lift, give

$$\frac{64C^{[1]}_{50}\sqrt{q/Q}}{125(.00247)} \le \frac{128(52/5)(183/500)(326/5)}{125(.00247)2^{32}}\eta(108\eta + 729) < .0201 < .036,$$
$$\frac{64C^{[1]}_{51}\sqrt{2/15}}{125(.00247)} < \frac{128(.366)(52/5)(326/5)}{125(.00247)2^{32}}(2916\eta + 19683) < .5414 < .951.$$

Both final comparisons follow already from $\eta < 1.000001$. This verifies the two caps in the first row of Table 3 for this output type, including its fixed-support multiplicity and source normalization.

With at least two active words, the common two-step map suffices. Its support multiplicities include all input supports, so it also bounds the contribution from the fixed support of the source. We use one upper vector for the three possible source sizes.

Lemma 37 (Initial two-step terms with several active sources). *The estimate* (126) *holds for input differences with* $w = 2,3,4$ *active words.*

Proof. Embed $b^{[w]}_2$ in the coordinates (j,w) and put

$$a = .01\mathbf{1} + .13e_2 + 48e_6 + 71e_9.$$

Directly from (118),

$$\frac{\omega}{1000}\mathsf{D}_{14}^{-1}b_2^{[w]} \le a \qquad (w = 2, 3, 4). \tag{121}$$

For $w = 2$ the three source coordinates, at indices $2, 6, 9$, are below $.129, 48, 71$, respectively. For $w = 3$ every coordinate is below .006; the largest is the degree-three coordinate. For $w = 4$ every coordinate is below .00001. Each statement follows by squaring (118) and using $q/Q < 1.000001$.

Substitute (105) into (110) and sum against this one positive source vector. The resulting coordinate bounds, rounded upwards to multiples of $1/1000$, are

$$\begin{aligned}\mathsf{M}a \le (&.003, .021, 1.085, 1.085, .015, .419, 2.590, 26.707,\\ &.294, .223, .628, .429, .031, 1.166)^\top.\end{aligned}$$

In particular,

$$\mathsf{M}a \le (1, 1, 2, 2, 1, 1, 3, 28, 1, 1, 1, 1, 1, 2)^\top \le g. \tag{122}$$

Thus (109) and (121) imply the desired initial two-step estimate for $w = 2, 3, 4$: divide (119) by 1000, apply the two-step map (110), and then use $\omega\mathsf{M}\mathsf{D}_{14}^{-1}b_2^{[w]}/1000 \le \mathsf{M}a \le g$. □

9.4 The four AES profile estimates

We can now collect the estimates used by the general contraction argument. The first two control propagation of arbitrary signed errors: one step for residual degrees at most two, and two steps through the shared high-degree projection for every input type. The last two initialize the recurrence uniformly over nonzero input differences.

Proposition 5 (AES profile estimates). *For the AES operators and projections just defined, the following statements hold with the rational maps in* (110)*.*

1. *For each output type α and each input type β of residual degree at most two,*

$$\frac{d_\beta}{d_\alpha}\,\|P_\alpha S P_\beta\| \le \omega\mathsf{L}_{\alpha\beta}. \tag{123}$$

2. *For all input and output types,*

$$\frac{d_\beta}{d_\alpha}\,\|P_\alpha S P_{\ge 3} S P_\beta\| \le \omega^2\mathsf{M}_{\alpha\beta}. \tag{124}$$

3. *Uniformly over every nonzero input difference,*

$$q^{5/2}\mathsf{D}_{14}^{-1}\mathbf{v}(f_4) \le 1000u_0, \tag{125}$$

$$q^{5/2}\omega^{-1}\mathsf{D}_{14}^{-1}\mathbf{v}(SP_{\ge 3}Sf_2) \le 1000g. \tag{126}$$

The starting vectors u_0, g are given in Table 2.

Proof (of Proposition 5). The local word estimates and the graph bounds have already been combined into (109), including the activity-only blocks and the shared middle projection. For the one-step estimate, divide its first inequality by q, multiply by d_β/d_α, and use the definition of L in (110). This proves (123).

For the two-step estimate, divide the second inequality by q^3, multiply by the same scale ratio, and use the definition of M in (110). This proves (124).

Proposition 4 proves (125) using the actual ordinary source laws.

For the initial two-step term, Lemma 36 handles $w = 1$ through (120). Lemma 37 handles $w = 2, 3, 4$ using the two-step map applied to the common source upper vector. Together with (119), these give (126) for every nonzero input difference. □

10 From the Recurrence to Pairwise Mixing

Proposition 5 reduces the remaining mixing argument to a nonnegative recurrence. After removing the proposed geometric decay, we show that the error profiles stay below two fixed vectors, one for each parity of the index. A common distance estimate for these vectors then gives Theorem 1 at every even round count from eight onward. All terminating decimals are exact rational numbers.

10.1 A support-sensitive distance bound

We specialize Lemma 2 to four words and round its support weights upward. The resulting functional retains the separate activity contribution and the Euclidean combination of residual degrees on each support size:

$$\Psi(v) = 2\sum_{t=1}^{3}\gamma_t v_{0,t} + \sum_{t=1}^{4}\gamma_t\sqrt{\sum_{i=1}^{t} v_{i,t}^2}, \qquad (\gamma_1,\gamma_2,\gamma_3,\gamma_4) = (2^{-47}, 5\,2^{-33}, 2^{-15}, 1), \qquad \Phi(v) = \Psi(\mathsf{D}_{14}v). \tag{127}$$

The map Φ expresses this functional in the scaled coordinates of Section 9. Both Ψ and Φ are componentwise monotone and positively homogeneous on nonnegative profiles. Thus an upper bound on a profile also bounds its output, with the same overall scale factor.

Lemma 38 (Distance conversion). *For any centered error f on Ω,*

$$\tfrac{1}{2}\,\|f\|_1 \le \frac{q^2}{2}\Psi(\mathbf{v}(f)). \tag{128}$$

Proof. Apply Lemma 2 with $b = 4$. Since $Q < q$, its four support weights are at most

$$(2q^{-3/2}, \sqrt{6}\,q^{-1}, 2q^{-1/2}, 1) \le (2^{-47}, 5\,2^{-33}, 2^{-15}, 1).$$

Here $q = 2^{32}$ and $\sqrt{6} < 5/2$. Replacing weights by these larger values preserves the upper bound. □

10.2 The normalized error recurrence

Recall the proposed two-round rate $\omega = 2^{-41}$ and the positive diagonal scale matrix D_{14} from Section 9.2. The maps L and M there have these factors removed. Restoring them gives the matrices used in Proposition 1:

$$A = \omega \mathsf{D}_{14} \mathsf{L} \mathsf{D}_{14}^{-1}, \qquad B = \omega^2 \mathsf{D}_{14} \mathsf{M} \mathsf{D}_{14}^{-1}.$$

Fix a nonzero input difference. After j two-round steps beyond four rounds, normalize its error profile by

$$x_j = q^{5/2} \omega^{-j} \mathsf{D}_{14}^{-1} \mathbf{v}(f_{4+2j}), \qquad j \geq 0. \tag{129}$$

Here x_0 is the four-round profile and x_1 is the six-round profile. The factors $q^{5/2}$ and ω^{-j} remove the common initial size and proposed decay; D_{14}^{-1} measures each component in units of its scale. A uniform bound on these normalized profiles will therefore give geometric decay of the original error.

Proposition 6 (Normalized error bounds). *For every nonzero input difference, the profiles in* (129) *satisfy*

$$\begin{aligned} x_0 &\leq 1000 u_0, \\ x_1 &\leq \mathsf{L} x_0 + 1000 g, \\ x_{j+1} &\leq \mathsf{L} x_j + \mathsf{M} x_{j-1} \qquad (j \geq 1). \end{aligned} \tag{130}$$

The corresponding error satisfies

$$\tfrac{1}{2} \left\| f_{4+2j} \right\|_1 \leq 2^{-17} \omega^j \Phi(x_j). \tag{131}$$

Proof. The bound on x_0 is (125). For x_1, split the six-round error using the two-step identity (13) at $r = 2$. The step bound (123) applies to the low residual degrees; the fixed-source bound (126) controls the initial two-step term and gives $1000g$ after normalization.

For subsequent steps, both input errors already have profiles in the sequence. Apply the same decomposition with the two-step operator bound (124). The one-step and two-step bounds carry ω and ω^2, respectively. These factors cancel the difference in powers of ω between the output and its two input profiles in (129). This proves (130).

Undoing the normalization in Lemma 38 gives

$$\tfrac{1}{2} \left\| f_{4+2j} \right\|_1 \leq \frac{q^2}{2} q^{-5/2} \omega^j \Psi(\mathsf{D}_{14} x_j) = 2^{-17} \omega^j \Phi(x_j),$$

since $q = 2^{32}$. □

10.3 Two invariant profiles

The two-step recurrence needs bounds at two consecutive indices to start an induction. We use the eight- and ten-round profiles, at indices 2 and 3. The available upper bounds have large coordinates in different positions, so we choose separate upper vectors for even and odd indices. Starting from the four-round vector, define

$$u_1 = \mathsf{L}u_0 + g, \qquad u_{j+1} = \mathsf{L}u_j + \mathsf{M}u_{j-1} \quad (j = 1, 2). \tag{132}$$

Positivity in (130) gives $x_j \le 1000u_j$ for $j = 0, 1, 2, 3$. Only u_2 and u_3 are needed to start the alternating bound; later indices will use the induction of Proposition 1.

Let $\mathbf{e}, \mathbf{o}$ be the positive integer vectors in Table 4. They satisfy

$$\begin{aligned} u_2 &\le \mathbf{e}, \quad \mathsf{Lo} + \mathsf{Me} \le \mathbf{e}, \\ u_3 &\le \mathbf{o}, \quad \mathsf{Le} + \mathsf{Mo} \le \mathbf{o}. \end{aligned} \tag{133}$$

The left column gives the two starting bounds. The right column preserves them: $\mathsf{Lo} + \mathsf{Me}$ controls the next even profile, and $\mathsf{Le} + \mathsf{Mo}$ controls the next odd profile.

To check all four inequalities by the same rational calculation, set

$$\vartheta_\alpha = \max\left\{ \frac{u_{2,\alpha}}{\mathbf{e}_\alpha}, \ \frac{u_{3,\alpha}}{\mathbf{o}_\alpha}, \ \frac{(\mathsf{Lo} + \mathsf{Me})_\alpha}{\mathbf{e}_\alpha}, \ \frac{(\mathsf{Le} + \mathsf{Mo})_\alpha}{\mathbf{o}_\alpha} \right\}. \tag{134}$$

Table 4 gives a common rational upper bound for these four ratios in each row. In particular, $\vartheta_\alpha < 99/100$ for all α. This proves (133) with a uniform margin.

The table is obtained from the rational map formulas (110), using the support sums (105). We expand coordinate 5, which has the largest upper ratio in the table. Its output type is $(1, 1)$, with $\mathbf{e}_5 = 133$ and $\mathbf{o}_5 = 528$. For this type, the only nonzero one-step entries are

$$\begin{aligned} \mathsf{L}_{5,8} &= 2^9\, 4\eta^3 \frac{52/5}{2^{32}} < .000005, \\ \mathsf{L}_{5,11} &= 2^9 \frac{.000243}{.00247}\, 2(2.4495)\eta^2 \frac{52/5}{2^{16}} < .03916. \end{aligned}$$

Indeed, the step formula requires all input activity coordinates to be unrestricted here: three for type $(1, 4)$ and two for type $(2, 4)$. The degree-zero input bound vanishes, and columns 12–14 are zero by definition.

For the two-step row, substitute the same output type into (110) and round each entry upward to a multiple of 10^{-6}. Writing $\boldsymbol{a}$ for the resulting integer row gives

$$\begin{aligned} 10^6(\mathsf{M}_{5,1}, \ldots, \mathsf{M}_{5,14}) &\le \boldsymbol{a}, \\ \boldsymbol{a} &= (4, 107, 11731, 1, 1, 11, 436, 22888, \\ &\qquad 94, 1076, 375747, 4691, 272550, 15920). \end{aligned}$$

Table 4. Two upper profiles and one bound for all four tests in (134). The last column is in units of 1/1000; its largest entry is $989 < 990$.

α	$\mathbf{e}_\alpha$	$\mathbf{o}_\alpha$	$1000\vartheta_\alpha \le$
1	342	76	985
2	45	130	983
3	39	30	978
4	39	30	978
5	133	528	989
6	74	64	949
7	81	49	949
8	39	44	962
9	37	48	975
10	163	51	973
11	97	56	986
12	37	29	978
13	316	73	979
14	179	90	982

For example, its eighth entry is obtained as $2^{-14}\widehat{B}^{\mathrm{ret}}_{5,8} < .022887304 < 22888\cdot 10^{-6}$; the scale ratio is one because $d_8 = d_5$.

The initial recurrence (132) gives the following integer upper vector and two further component bounds:

$$u_1 \le \boldsymbol{v}, \qquad \boldsymbol{v} = (2, 4, 451, 451, 2, 26, 901, 15001, 2, 26, 451, 3, 11, 6)^{\top},$$
$$u_{2,8} < 2.815, \qquad u_{2,11} < 14.228.$$

These roundings suffice for the delicate starting comparison. The two largest products in $\boldsymbol{av}$ have indices 8 and 11; the other twelve sum to 8820499. Therefore

$$\begin{aligned}(\mathsf{M}u_1)_5 &\le 10^{-6}\boldsymbol{av}\\ &= 10^{-6}(22888\cdot 15001 + 375747\cdot 451 + 8820499)\\ &= 521.625284 < 521.626,\\ (\mathsf{L}u_2)_5 &< .000005(2.815) + .03916(14.228)\\ &= .557182555 < .558.\end{aligned}$$

Adding the two contributions and dividing by $\mathbf{o}_5$ gives

$$\frac{u_{3,5}}{528} < \frac{521.626 + .558}{528} = \frac{522.184}{528} < \frac{989}{1000}, \qquad 528\cdot\frac{989}{1000} = 522.192.$$

The other three tests at this coordinate have more room. Direct substitution into the same row formulas gives

$$u_{2,5} < 22, \qquad (\mathsf{Lo} + \mathsf{Me})_5 < 130, \qquad (\mathsf{Le} + \mathsf{Mo})_5 < 49.$$

Their ratios are below 22/133, 130/133, and 49/528, respectively, each less than 989/1000. This proves the complete coordinate-5 row of Table 4.

Lemma 39 (Alternating profile bounds). *For every nonzero input difference and every integer $k \geq 1$,*

$$x_{2k} \leq 1000\mathbf{e}, \qquad x_{2k+1} \leq 1000\mathbf{o}.$$

Proof. Apply part (ii) of Proposition 1 to the shifted sequence $\big(x_{n+2}\big)_{n\geq 0}$ with

$$A = \mathsf{L}, \quad B = \mathsf{M}, \quad \rho = 1, \quad c = 1000, \quad w_0 = \mathbf{e}, \quad w_1 = \mathbf{o}.$$

The decay parameter is 1 because the powers of ω have already been removed in (129). The recurrence hypothesis is (130). The first two bounds follow from $x_2 \leq 1000u_2$ and $x_3 \leq 1000u_3$; all four weight and starting conditions then follow from (133). □

10.4 The final distance bound

It remains to estimate the distance functional on 1000**e** and 1000**o**. The support weights in (127) make the contributions from fewer than four active words small. Collect these terms as

$$\begin{aligned} b_t(v) &= 2(\mathsf{D}_{14}v)_{0,t} + \sqrt{\sum_{i=1}^{t} (\mathsf{D}_{14}v)_{i,t}^2}, \qquad t = 1, 2, 3, \\ \Phi_{\mathrm{part}}(v) &= 2^{-47} b_1(v) + 5\,2^{-33} b_2(v) + 2^{-15} b_3(v). \end{aligned} \tag{135}$$

The remaining term is the Euclidean norm of the four all-active residual lengths, at coordinates $8, 11, 13, 14$ in (99). We estimate the partial-support sum and this norm separately.

Lemma 40 (Uniform output bound). *For every nonzero input difference and every integer $j \geq 2$, $\Phi(x_j) < 112$.*

Proof. First consider the partial-support terms. For either $v = 1000\mathbf{e}$ or $v = 1000\mathbf{o}$, bound each square root in b_t by the sum of its nonnegative components. Substitution of the scales then gives

$$(b_1(v), b_2(v), b_3(v)) < (2^{20}, 2^{18}, 64000).$$

Consequently

$$\Phi_{\mathrm{part}}(v) < 2^{-27} + 5\,2^{-15} + 125/64 < 2.$$

For the all-active terms, retain the Euclidean norm. The four lengths satisfy

$$\begin{aligned} 1000(d_8\mathbf{e}_8, d_{11}\mathbf{e}_{11}, d_{13}\mathbf{e}_{13}, d_{14}\mathbf{e}_{14}) &< (97, 24, 6, 1), \\ 1000(d_8\mathbf{o}_8, d_{11}\mathbf{o}_{11}, d_{13}\mathbf{o}_{13}, d_{14}\mathbf{o}_{14}) &< (109, 14, 2, 1). \end{aligned}$$

Both norms are below 110, since

$$97^2 + 24^2 + 6^2 + 1 = 10022 < 110^2, \qquad 109^2 + 14^2 + 2^2 + 1 = 12082 < 110^2.$$

Both profile outputs are therefore below $2 + 110 = 112$. Lemma 39 and monotonicity of Φ give the claim. □

Proof of the mixing theorem. Lemma 40 supplies one output constant for both parities. Substituting it into (131) restores the factor 2^{-41} for every two additional rounds.

Proof (of Theorem 1). Fix a nonzero input difference and put $j = k + 2$ for an integer $k \geq 0$. Proposition 6 and Lemma 40 give

$$\begin{aligned} \tfrac{1}{2} \left\| f_{8+2k} \right\|_1 &\leq 2^{-17}\, 2^{-41(k+2)} \Phi(x_{k+2}) \\ &< 112\, 2^{-99-41k} = 7\, 2^{-95-41k} \\ &< 2^{-92.19-41k}. \end{aligned}$$

The inequality $7^{100} < 2^{281}$ justifies the downward decimal rounding in the final line. The common bound applies to every nonzero input difference. Taking the maximum gives the claimed bound on ε_{8+2k} by (4). □

The starting index. The starting comparisons in (133) place the eight- and ten-round profiles under the two upper vectors. The same output constant cannot be obtained at six rounds from the starting vector used here: it has $1000u_{1,8} \geq 1000g_8 = 15000000$. The degree-one all-active contribution to $\Phi(1000u_1)$ alone is at least $.00247 \cdot 15000000 = 37050$, already above 112. This explains the eight-round starting point of this argument. The calculation concerns an upper vector and gives no lower bound on the actual error profile or AES's distance.

11 Conclusions

The bound $\varepsilon_{8+2k} < 2^{-92.19-41k}$ gives 10, 14, and 16 sufficient rounds for the three target accuracies. Under independent retained round keys, whitening transfers the pairwise bound to at most two adaptive classical forward queries. The proof separates estimates of the actual input distribution from operator bounds that propagate arbitrary signed errors.

The method is useful when local responses share inputs or when an intermediate error retains information about the transition that produced it. In the first case, the common-source Gram estimates sum overlaps before taking a worst case. In the second, estimating two transitions together keeps their shared middle projection inside the operator. For additive MDS diffusion, fixing the outside coordinates leaves affine fibers of compatible middle values. Summing over these fibers preserves constraints that can be lost when the two transitions are estimated separately.

To apply the argument to another cipher with additive MDS diffusion, one must supply local operator norms, row and column energies, second Gram moments, and fourth moments for the one-sided Fourier transforms used by its diffusion graphs. These quantities give the one-step and two-step bounds. One must also estimate two initial profiles from the actual input law and convert component lengths to the desired distance. Proposition 1 then reduces propagation to two inequalities for positive weight vectors, one for each parity. Lemma 39 applies this criterion to AES starting at eight rounds. Standard AES key schedules and higher-query guarantees require separate arguments.

References

1. Baignères, T., Vaudenay, S.: Proving the security of AES substitution-permutation network. In: Selected Areas in Cryptography—SAC 2005. LNCS, vol. 3897, pp. 65–81. Springer (2006). https://doi.org/10.1007/11693383_5
2. Beierle, C., Felke, P., Leander, G., Neumann, P., Stennes, L.: On perfect linear approximations and differentials over two-round SPNs. In: Advances in Cryptology—CRYPTO 2023. LNCS, vol. 14083, pp. 209–239. Springer (2023). https://doi.org/10.1007/978-3-031-38548-3_8
3. Beyne, T.: A geometric approach to linear cryptanalysis. In: Advances in Cryptology—ASIACRYPT 2021. LNCS, vol. 13090, pp. 36–66. Springer (2021). https://doi.org/10.1007/978-3-030-92062-3_2
4. Beyne, T., Leander, G., Schütt, I.: Pairwise independence of AES-like block ciphers. In: Daemen, J., Thomé, E. (eds.) Advances in Cryptology—EUROCRYPT 2026. LNCS, vol. 16546, pp. 181–210. Springer (2026). https://doi.org/10.1007/978-3-032-25333-0_7, corrected full version: ePrint 2025/1495, 27 September 2026
5. Beyne, T., Rijmen, V.: Differential cryptanalysis in the fixed-key model. In: Advances in Cryptology—CRYPTO 2022. LNCS, vol. 13509, pp. 687–716. Springer (2022). https://doi.org/10.1007/978-3-031-15982-4_23
6. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. Journal of Cryptology **4**(1), 3–72 (1991). https://doi.org/10.1007/BF00630563
7. Brodsky, A., Hoory, S.: Simple permutations mix even better. Random Structures & Algorithms **32**(3), 274–289 (2008). https://doi.org/10.1002/rsa.20194
8. Canteaut, A., Roué, J.: Differential attacks against SPN: A thorough analysis. In: Codes, Cryptology, and Information Security—C2SI 2015. LNCS, vol. 9084, pp. 45–62. Springer (2015). https://doi.org/10.1007/978-3-319-18681-8_4
9. Canteaut, A., Roué, J.: On the behaviors of affine equivalent sboxes regarding differential and linear attacks. In: Advances in Cryptology—EUROCRYPT 2015. LNCS, vol. 9056, pp. 45–74. Springer (2015). https://doi.org/10.1007/978-3-662-46800-5_3
10. Carlitz, L., Uchiyama, S.: Bounds for exponential sums. Duke Mathematical Journal **24**(1), 37–41 (1957). https://doi.org/10.1215/S0012-7094-57-02406-7
11. Chabaud, F., Vaudenay, S.: Links between differential and linear cryptanalysis. In: Advances in Cryptology—EUROCRYPT '94. LNCS, vol. 950, pp. 356–365. Springer (1995). https://doi.org/10.1007/BFb0053450
12. Charpin, P., Helleseth, T., Zinoviev, V.: Propagation characteristics of $x \mapsto x^{-1}$ and Kloosterman sums. Finite Fields and Their Applications **13**(2), 366–381 (2007). https://doi.org/10.1016/j.ffa.2005.08.007
13. Cho, H.S., Sung, S.H., Kwon, D., Lee, J.K., Song, J.H., Lim, J.: New method for bounding the maximum differential probability for SPNs and ARIA. In: Information Security and Cryptology—ICISC 2004. LNCS, vol. 3506, pp. 21–32. Springer (2005). https://doi.org/10.1007/11496618_4
14. Christ, M.: The optimal constants in Hölder–brascamp–lieb inequalities for discrete abelian groups (2013), https://arxiv.org/abs/1307.8442
15. Chun, K., Kim, S., Lee, S., Sung, S.H., Yoon, S.: Differential and linear cryptanalysis for 2-round SPNs. Information Processing Letters **87**(5), 277–282 (2003). https://doi.org/10.1016/S0020-0190(03)00333-8
16. Cogliati, B., Dodis, Y., Katz, J., Lee, J., Steinberger, J., Thiruvengadam, A., Zhang, Z.: Provable security of (tweakable) block ciphers based on

substitution-permutation networks. In: Advances in Cryptology—CRYPTO 2018. LNCS, vol. 10991, pp. 722–753. Springer (2018). https://doi.org/10.1007/978-3-319-96884-1_24
17. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In: Fast Software Encryption—FSE 1994. LNCS, vol. 1008, pp. 275–285. Springer (1995). https://doi.org/10.1007/3-540-60590-8_21
18. Daemen, J., Lamberger, M., Pramstaller, N., Rijmen, V., Vercauteren, F.: Computational aspects of the expected differential probability of 4-round AES and AES-like ciphers. Computing **85**(1–2), 85–104 (2009). https://doi.org/10.1007/s00607-009-0034-y
19. Daemen, J., Rijmen, V.: The wide trail design strategy. In: Cryptography and Coding. LNCS, vol. 2260, pp. 222–238. Springer (2001). https://doi.org/10.1007/3-540-45325-3_20
20. Daemen, J., Rijmen, V.: Understanding two-round differentials in AES. In: Security and Cryptography for Networks—SCN 2006. LNCS, vol. 4116, pp. 78–94. Springer (2006). https://doi.org/10.1007/11832072_6, full report: ePrint 2006/039, titled *Two-Round AES Differentials?*
21. Dinur, I.: New techniques for analyzing differentials with application to AES. In: Daemen, J., Thomé, E. (eds.) Advances in Cryptology—EUROCRYPT 2026. LNCS, vol. 16546, pp. 211–242. Springer (2026). https://doi.org/10.1007/978-3-032-25333-0_8, full version: ePrint 2025/1326, 5 March 2026
22. Dodis, Y., Katz, J., Steinberger, J.P., Thiruvengadam, A., Zhang, Z.: Provable security of substitution-permutation networks. Cryptology ePrint Archive, Paper 2017/016 (2017), https://eprint.iacr.org/2017/016
23. Dodis, Y., Stam, M., Steinberger, J., Liu, T.: Indifferentiability of confusion-diffusion networks. In: Advances in Cryptology—EUROCRYPT 2016. LNCS, vol. 9666, pp. 679–704. Springer (2016). https://doi.org/10.1007/978-3-662-49896-5_24
24. Dujmovic, J., Pelecanos, A., Tessaro, S.: When simple permutations mix poorly: Limited independence does not imply pseudorandomness. In: Advances in Cryptology—EUROCRYPT 2026. LNCS, vol. 16546, pp. 424–450. Springer (2026). https://doi.org/10.1007/978-3-032-25333-0_15
25. Dunkelman, O., Kumar, A., Lambooij, E., Sanadhya, S.K.: Counting active S-boxes is not enough. In: Progress in Cryptology—INDOCRYPT 2020. LNCS, vol. 12578, pp. 332–344. Springer (2020). https://doi.org/10.1007/978-3-030-65277-7_15
26. Gowers, W.T.: An almost m-wise independent random permutation of the cube. Combinatorics, Probability and Computing **5**(2), 119–130 (1996). https://doi.org/10.1017/S0963548300001917
27. Hong, S., Lee, S., Lim, J., Sung, J., Cheon, D., Cho, I.: Provable security against differential and linear cryptanalysis for the SPN structure. In: Fast Software Encryption—FSE 2000. LNCS, vol. 1978, pp. 273–283. Springer (2001). https://doi.org/10.1007/3-540-44706-7_19
28. Hoory, S., Magen, A., Myers, S.A., Rackoff, C.: Simple permutations mix well. Theoretical Computer Science **348**(2–3), 251–261 (2005). https://doi.org/10.1016/j.tcs.2005.09.016
29. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press (1991). https://doi.org/10.1017/CBO9780511840371, chapter 5: The Hadamard product

30. Hurst, N.: Improving pairwise independence bounds for AES via numerical spectral computation. COSIC seminar, KU Leuven (August 2026), 25 August 2026. Public seminar abstract
31. Jain, V., Liu, T., Mizgerd, C., Pelecanos, A., Tessaro, S., Vaikuntanathan, V.: How fast does the inverse walk approximate a random permutation? In: Advances in Cryptology—CRYPTO 2026. LNCS, vol. 16805, pp. 257–285. Springer (2026). https://doi.org/10.1007/978-3-032-35412-9_9
32. Kang, J.S., Hong, S., Lee, S., Yi, O., Park, C., Lim, J.: Practical and provable security against differential and linear cryptanalysis for substitution-permutation networks. ETRI Journal **23**(4), 158–167 (2001). https://doi.org/10.4218/etrij.01.0101.0402
33. Kaplan, E., Naor, M., Reingold, O.: Derandomized constructions of k-wise (almost) independent permutations. Algorithmica **55**(1), 113–133 (2009). https://doi.org/10.1007/s00453-008-9267-y
34. Keliher, L.: Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In: Advanced Encryption Standard—AES 2004. LNCS, vol. 3373, pp. 42–57. Springer (2005). https://doi.org/10.1007/11506447_5
35. Keliher, L., Meijer, H., Tavares, S.: Improving the upper bound on the maximum average linear hull probability for Rijndael. In: Selected Areas in Cryptography—SAC 2001. LNCS, vol. 2259, pp. 112–128. Springer (2001). https://doi.org/10.1007/3-540-45537-X_9
36. Keliher, L., Meijer, H., Tavares, S.: New method for upper bounding the maximum average linear hull probability for SPNs. In: Advances in Cryptology—EUROCRYPT 2001. LNCS, vol. 2045, pp. 420–436. Springer (2001). https://doi.org/10.1007/3-540-44987-6_26
37. Keliher, L., Sui, J.: 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 report: ePrint 2005/321
38. Kranz, T., Leander, G., Wiemer, F.: Linear cryptanalysis: Key schedules and tweakable block ciphers. IACR Transactions on Symmetric Cryptology **2017**(1), 474–505 (2017). https://doi.org/10.46586/tosc.v2017.i1.474-505, full version: ePrint 2017/154
39. Lai, X., Massey, J.L., Murphy, S.: Markov ciphers and differential cryptanalysis. In: Advances in Cryptology—EUROCRYPT '91. LNCS, vol. 547, pp. 17–38. Springer (1991). https://doi.org/10.1007/3-540-46416-6_2
40. Liu, T., Pelecanos, A., Tessaro, S., Vaikuntanathan, V.: Layout graphs, random walks and the t-wise independence of SPN block ciphers. In: Advances in Cryptology—CRYPTO 2023. LNCS, vol. 14083, pp. 694–726. Springer (2023). https://doi.org/10.1007/978-3-031-38548-3_23
41. Liu, T., Tessaro, S., Vaikuntanathan, V.: The t-wise independence of substitution-permutation networks. In: Advances in Cryptology—CRYPTO 2021. LNCS, vol. 12828, pp. 454–483. Springer (2021). https://doi.org/10.1007/978-3-030-84259-8_16
42. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error-Correcting Codes, North-Holland Mathematical Library, vol. 16. North-Holland (1977)
43. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Advances in Cryptology—EUROCRYPT '93. LNCS, vol. 765, pp. 386–397. Springer (1994). https://doi.org/10.1007/3-540-48285-7_33
44. Maurer, U., Pietrzak, K., Renner, R.: Indistinguishability amplification. In: Advances in Cryptology—CRYPTO 2007. LNCS, vol. 4622, pp. 130–149. Springer (2007). https://doi.org/10.1007/978-3-540-74143-5_8

45. Maurer, U., Tessaro, S.: Computational indistinguishability amplification: Tight product theorems for system composition. In: Advances in Cryptology—CRYPTO 2009. LNCS, vol. 5677, pp. 355–373. Springer (2009). https://doi.org/10.1007/978-3-642-03356-8_21
46. Miles, E., Viola, E.: Substitution-permutation networks, pseudorandom functions, and natural proofs. Journal of the ACM **62**(6), 46:1–46:29 (2015). https://doi.org/10.1145/2792978
47. National Institute of Standards and Technology: Advanced encryption standard (AES). Federal Information Processing Standards Publication 197-upd1, U.S. Department of Commerce (May 2023). https://doi.org/10.6028/NIST.FIPS.197-upd1, updated 9 May 2023
48. Nyberg, K.: Differentially uniform mappings for cryptography. In: Advances in Cryptology—EUROCRYPT '93. LNCS, vol. 765, pp. 55–64. Springer (1994). https://doi.org/10.1007/3-540-48285-7_6
49. Nyberg, K., Knudsen, L.R.: Provable security against a differential attack. Journal of Cryptology **8**(1), 27–37 (1995). https://doi.org/10.1007/BF00204800
50. Park, S., Sung, S.H., Chee, S., Yoon, E.J., Lim, J.: On the security of Rijndael-like structures against differential and linear cryptanalysis. In: Advances in Cryptology—ASIACRYPT 2002. LNCS, vol. 2501, pp. 176–191. Springer (2002). https://doi.org/10.1007/3-540-36178-2_11
51. Park, S., Sung, S.H., Lee, S., Lim, J.: Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES. In: Fast Software Encryption—FSE 2003. LNCS, vol. 2887, pp. 247–260. Springer (2003). https://doi.org/10.1007/978-3-540-39887-5_19
52. Rønjom, S.: A short note on a weight probability distribution related to SPNs. Cryptology ePrint Archive, Paper 2019/750 (2019), https://eprint.iacr.org/2019/750
53. Sano, F., Ohkuma, K., Shimizu, H., Kawamura, S.: On the security of nested SPN cipher against the differential and linear cryptanalysis. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **E86-A**(1), 37–46 (2003), https://globals.ieice.org/en_transactions/fundamentals/10.1587/e86-a_1_37/_p
54. Vaudenay, S.: Decorrelation: A theory for block cipher security. Journal of Cryptology **16**(4), 249–286 (2003). https://doi.org/10.1007/s00145-003-0220-6