Sharp Minimum-Distance Lower Tails for RAA Codes

Majid Khabbazian
University of Alberta, Canada

Abstract

Repeat–accumulate–accumulate (RAA) codes are sparse random linear codes with linear-time encoders. Although RAA ensembles can have linear minimum distance with high probability, this does not determine how often a randomly sampled code falls below a prescribed relative-distance threshold. We determine this lower-tail probability and identify the rare mechanisms that dominate it.

Let n be the message length, $r \geq 4$ a fixed repetition factor, and $N = rn$ the block length. For the randomly scaled ensemble

$$G = R\Pi_1 V_1 A \Pi_2 V_2 A,$$

where R is repetition, A is the prefix-sum accumulator, Π_1, Π_2 are independent uniform interleavers, and V_1, V_2 have independent uniform nonzero diagonal entries in $\mathbb{F}_q$, we prove that, for every fixed $0 < \delta \leq 1/2$ and $\gamma > 0$,

$$q - 1 \geq \gamma N \quad \Longrightarrow \quad \mathbb{P}[d_{\min}(G) \leq \delta N] = \Theta_{r,\delta,\gamma}(N^{1-r}).$$

For the unscaled ensemble

$$G_0 = R\Pi_1 A \Pi_2 A,$$

if $\mathrm{char}(\mathbb{F}_q) > r$ and $0 < \delta \leq 1/10$, then

$$\mathbb{P}[d_{\min}(G_0) \leq \delta N] = \Theta_{r,\delta}(N^{2-r}).$$

Thus random scaling improves the failure probability by one power of N. We also prove

$$\mathbb{P}[d_{\min}(G) \leq \delta N] = \Omega_{r,\delta}\Big(N^{1-\lceil r/2 \rceil}(q-1)^{-\lfloor r/2 \rfloor}\Big),$$

showing that linear field growth is necessary, in asymptotic order, for the constant-factor N^{1-r} law.

1 Introduction

Repeat–accumulate (RA), repeat–accumulate–accumulate (RAA), and more generally repeat–multiple–accumulate codes are sparse random-interleaver ensembles with linear-time encoders. An RA encoder consists of repetition, an interleaver, and a prefix-sum accumulator; an RAA encoder adds a second interleaver–accumulator stage. This additional stage changes the distance behavior fundamentally: standard random-interleaver RA codes are not asymptotically good, whereas RAA codes can have minimum distance linear in the block length with probability tending to one [4, 3, 6].

For an asymptotically good ensemble, a finer question remains: *how quickly does the probability of drawing a code below a prescribed relative-distance threshold decay?* For fixed $\delta > 0$, we study

$$\mathbb{P}[d_{\min}(G) \leq \delta N].$$

Determining its sharp order quantifies the reliability of sampling from the ensemble and identifies the algebraic or geometric mechanisms that dominate failure.

We compare two finite-field RAA ensembles. The standard encoder

$$G_0 = R\Pi_1 A\Pi_2 A$$

uses repetition, permutation, and field additions. The scaled encoder

$$G = R\Pi_1 V_1 A\Pi_2 V_2 A$$

additionally multiplies every coordinate by an independent random nonzero field element before each accumulator. These multiplications preserve linear-time encoding but add concrete arithmetic cost. In return, they randomize prefix-sum cancellations that may occur deterministically without scaling. Our results determine precisely how this additional algebraic randomness changes the minimum-distance failure probability.

1.1 Main results

Theorem 1.1 (Sharp scaled tail)**.** *Fix $r \geq 4$, $0 < \delta \leq 1/2$, and $\gamma > 0$. Let $N \to \infty$ through multiples of r, and let $q = q(N)$ be prime powers satisfying $q - 1 \geq \gamma N$. Then*

$$\mathbb{P}[d_{\min}(G) \leq \delta N] = \Theta_{r,\delta,\gamma}(N^{1-r}).$$

The lower bound is generated by a weight-one message whose r repeated copies are sent to the terminal r coordinates by the first interleaver. Its first accumulated word then has bounded support, and the second interleaver places that support inside the last δN coordinates with constant probability. The upper bound proves that, under a linear field, every other message weight is asymptotically no more likely.

Theorem 1.2 (Sharp unscaled tail)**.** *Fix $r \geq 4$ and $0 < \delta \leq 1/10$. Let $q = q(N)$ be prime powers satisfying $\mathrm{char}(\mathbb{F}_q) > r$. Then*

$$\mathbb{P}[d_{\min}(G_0) \leq \delta N] = \Theta_{r,\delta}(N^{2-r}).$$

For two message coordinates $u \neq v$, consider $e_u - e_v$. If the $2r$ repeated copies are interleaved into r adjacent mixed pairs, each pair cancels exactly, and the first accumulator output has weight r. There are $\Theta(N^2)$ choices of $\{u, v\}$, while each pairing event has probability $\Theta(N^{-r})$, explaining the exponent $2 - r$. Random scaling destroys this deterministic identity; over a sufficiently large field, the weight-one geometric event becomes dominant instead.

Corollary 1.3 (Effect of random scaling)**.** *In the common range $r \geq 4$, $0 < \delta \leq 1/10$, $\mathrm{char}(\mathbb{F}_q) > r$, and $q - 1 \geq \gamma N$,*

$$\mathbb{P}[d_{\min}(G_0) \leq \delta N] = \Theta_{r,\delta,\gamma}\left(N\,\mathbb{P}[d_{\min}(G) \leq \delta N]\right).$$

Put

$$Q := q - 1, \qquad s := \lfloor r/2 \rfloor, \qquad t := \lceil r/2 \rceil.$$

A weight-one message can be interleaved into s adjacent pairs, together with one terminal singleton when r is odd, and each pair cancels with probability Q^{-1}.

Theorem 1.4 (Field-dependent paired-cancellation obstruction)**.** *For every fixed $r \geq 4$ and $0 < \delta \leq 1/2$,*

$$\mathbb{P}[d_{\min}(G) \leq \delta N] = \Omega_{r,\delta}\left(N^{1-t}Q^{-s}\right).$$

Table 1: Proved low-distance mechanisms. The unscaled upper theorem assumes $\mathrm{char}(\mathbb{F}_q) > r$ and $\delta \le 1/10$; its lower mechanism is characteristic-independent.

Regime	Dominant or obstructing witness	Probability scale
No scaling	mixed adjacent pairs for $e_u - e_v$	$\Theta(N^{2-r})$
Random scaling, $Q \ge \gamma N$	weight-one boundary concentration	$\Theta(N^{1-r})$
Random scaling, arbitrary Q	probabilistic scalar-pair cancellation	$\Omega(N^{1-\lceil r/2 \rceil} Q^{-\lfloor r/2 \rfloor})$

Corollary 1.5 (Necessity of linear field growth)**.** *If $Q = o(N)$, then*

$$\frac{\mathbb{P}[d_{\min}(G) \le \delta N]}{N^{1-r}} \longrightarrow \infty.$$

In particular, a constant-factor $O(N^{1-r})$ upper bound is impossible for sublinear Q.

Thus $Q = \Omega(N)$ is necessary and sufficient, in asymptotic order, for the sharp constant-factor N^{1-r} law. We do not claim a complete interpolation formula for every sublinear Q; for example, a tail with additional polylogarithmic factors is not excluded.

1.2 Relation to previous distance analyses

Classical analyses of RA and RAA codes established asymptotic distance properties using input–output weight enumerators and random-interleaver arguments [4, 3, 6]. In particular, random RAA ensembles can have minimum distance linear in the block length with probability tending to one, and explicit or partially explicit interleavers have been studied as a route to constructive variants [5]. These results address asymptotic goodness, but not the sharp probability of falling below a fixed relative-distance threshold.

Akhiani and Zhang studied a *randomly scaled* large-field RAA ensemble, in which independent random nonzero coordinate multipliers are inserted into the encoder [2]. For this scaled ensemble, their fixed-threshold failure estimate has the form

$$\mathbb{P}[d_{\min}(G) \le \delta N] \le \widetilde{O}_r(N^{2-r}) + O(Q^{-1}), \qquad Q := q - 1.$$

This gives a qualitative high-probability distance guarantee, but does not determine the sharp failure probability, the dominant low-distance witness, or the field-size order required to attain the optimal tail. It also does not analyze the standard unscaled RAA encoder and therefore does not quantify what is gained by paying for the additional coordinate multiplications.

We resolve these questions. At the linear field scale $Q = \Theta(N)$, the preceding estimate gives only $\widetilde{O}_r(N^{-1})$, whereas our result gives the exact order N^{1-r}. For $r = 4$, this is the difference between an N^{-1} upper bound and a sharp N^{-3} law; for general fixed $r \ge 4$, the improvement is by N^{r-2}, up to logarithmic factors. Even when Q is large enough to make the Q^{-1} term negligible, the preceding N^{2-r} term remains one polynomial power above the sharp N^{1-r} tail.

We also identify the order-optimal field threshold and prove the distinct sharp unscaled law $\Theta(N^{2-r})$. The matching bounds explain the one-power separation: without scaling, deterministic weight-two cancellations dominate; with random scaling over a linear-size field, those cancellations are suppressed and a rarer weight-one support-concentration event dominates.

These advances require a different proof framework rather than a refinement of the previous weight-enumerator estimate. We count adaptive cancellation traces instead of individual messages, charge both accumulators against one shared projective-dimension budget, and combine this algebraic

count with exact random-gap and hypergeometric transition laws. The shared-rank argument extends to arbitrary adaptive linear stages and is not specific to RAA codes.

1.3 Proof overview

At an accumulator, call an input coordinate *active* when its value is nonzero. Fix a message support of size w. Once the active-coordinate order is exposed, the accumulator prefix sums are linear forms in the w nonzero message values. We group messages by the exact sets of prefix forms that vanish, called *cancellation traces*, rather than union bounding over all value assignments.

The central shared-rank theorem shows that the two accumulators consume one common projective-dimension budget. If d_i is the new rank contributed by the vanishing-prefix equations at accumulator i, then

$$d_1 + d_2 \leq w - 1.$$

Together with the fact that zero prefixes cannot be adjacent, this sharply limits the number of trace pairs. In the master bound, canonical basis labels are exposed before the unordered support gaps on which their output weights depend. This exposure order prevents an existential witness from selecting a support after observing the gap randomness.

A uniform interleaver decomposes into an independent active order and a uniform unordered support. The trace depends on the order and algebraic values, while the accumulator output weight depends on the gaps between the support positions. Conditional on a fixed trace, these gaps obey exact composition and hypergeometric laws. In the scaled ensemble, zero equations beyond the shared projective budget incur the additional penalty

$$Q^{-(a+b-(w-1))_+}.$$

Summing trace counts together with this field penalty before estimating the gaps is what makes a linear field size sufficient.

Sparse message weights are controlled through generating functions. For the unscaled ensemble, the high-weight exponent is bounded analytically using a relative-entropy decomposition with a quantitative quadratic remainder. For the scaled ensemble, the small-surplus range is controlled analytically by complementary mutual-information estimates: a KL-duality bound at lower normalized weight and a fourth-moment bound at higher normalized weight. Matching lower bounds identify the dominant witnesses in all three regimes of Table 1.

1.4 Organization

Section 2 defines the two ensembles and fixes notation. Section 3 develops cancellation traces and exact random-gap laws. Section 4 derives unified master bounds for both ensembles with a complete conditioning ledger. The unscaled and scaled upper bounds appear in sections 5 and 6. Matching lower bounds and the field-size obstruction are proved in section 7. We conclude with the mechanism classification, the scope of the method, and its current limitations.

2 Models and notation

For a positive integer M, write $[M] := \{1, \ldots, M\}$. For a vector z, let $\mathrm{supp}(z)$ and $\mathrm{wt}(z)$ denote its support and Hamming weight. We write e_i for the i-th standard basis row vector and $\mathbf{1}_S$ for the indicator row vector of a set S. For a real number u, put $u_+ := \max\{u, 0\}$. We use the convention

$$\binom{a}{b} = 0 \qquad \text{if } b \notin \{0, \ldots, a\},$$

and all logarithms are natural unless stated otherwise.

Let q be a prime power, and write

$$\mathbb{F}_q^* := \mathbb{F}_q \setminus \{0\}, \qquad Q := q-1, \qquad p := \mathrm{char}(\mathbb{F}_q).$$

The field size $q = q(N)$ may vary with the block length.

Fix a repetition factor $r \geq 2$ and a message length $n \geq 1$, and put

$$N := rn.$$

All vectors are row vectors. The repetition matrix $R \in \mathbb{F}_q^{n \times N}$ maps

$$(x_1, \ldots, x_n) \longmapsto (\underbrace{x_1, \ldots, x_1}_{r}, \ldots, \underbrace{x_n, \ldots, x_n}_{r}).$$

Thus a message of weight w produces exactly $m := rw$ nonzero coordinates before the first interleaver.

The accumulator matrix $A \in \mathbb{F}_q^{N \times N}$ is defined by

$$(zA)_j = \sum_{i=1}^{j} z_i, \qquad 1 \leq j \leq N.$$

It is triangular with ones on the diagonal and is therefore invertible.

Let Π_1, Π_2 be independent uniform permutation matrices on $[N]$. For $S \subseteq [N]$ and a permutation matrix Π, define

$$\Pi(S) := \mathrm{supp}(\mathbf{1}_S \Pi).$$

Let V_1, V_2 be independent random diagonal matrices whose diagonal entries are independent and uniform in $\mathbb{F}_q^*$. The four random matrices Π_1, Π_2, V_1, V_2 are mutually independent.

We compare the unscaled RAA ensemble

$$G_0 := R\Pi_1 A \Pi_2 A \tag{1}$$

with the randomly scaled RAA ensemble

$$G := R\Pi_1 V_1 A \Pi_2 V_2 A. \tag{2}$$

The diagonal matrices preserve support but independently multiply each coordinate by a uniformly random nonzero field element before the corresponding accumulator. Since R has rank n and every subsequent factor is invertible, both G_0 and G have rank n.

For an $n \times N$ full-rank generator matrix B, define

$$d_{\min}(B) := \min_{x \in \mathbb{F}_q^n \setminus \{0\}} \mathrm{wt}(xB).$$

For $0 < \delta \leq 1$, put $L_\delta := \lfloor \delta N \rfloor$ and define

$$\mathcal{E}_0(\delta) := \{d_{\min}(G_0) \leq L_\delta\}, \qquad \mathcal{E}_{\mathrm{sc}}(\delta) := \{d_{\min}(G) \leq L_\delta\}.$$

For each $1 \leq w \leq n$, define the existential weight-w events

$$\mathcal{E}_{0,w}(\delta) := \big\{\exists x \in \mathbb{F}_q^n : \mathrm{wt}(x) = w,\ \mathrm{wt}(xG_0) \leq L_\delta\big\}, \tag{3}$$

$$\mathcal{E}_{\mathrm{sc},w}(\delta) := \big\{\exists x \in \mathbb{F}_q^n : \mathrm{wt}(x) = w,\ \mathrm{wt}(xG) \leq L_\delta\big\}. \tag{4}$$

Probabilities involving $\mathcal{E}_0$ are over Π_1, Π_2, whereas those involving $\mathcal{E}_{\mathrm{sc}}$ are over Π_1, Π_2, V_1, V_2.

For an accumulator input $z \in \mathbb{F}_q^N$, a coordinate is called *active* if it is nonzero. Thus the active support is $\mathrm{supp}(z)$. If

$$\mathrm{supp}(z) = \{p_1 < \cdots < p_M\},$$

then $p_1, \ldots, p_M$ are the active positions and $z_{p_1}, \ldots, z_{p_M}$ are the active values in left-to-right order.

Unless stated otherwise, asymptotics are taken as $N \to \infty$ through multiples of the fixed repetition factor r, with parameters such as δ and γ fixed. Subscripts on O, Ω, and Θ indicate the parameters on which the implicit constants may depend. Whenever $N/2$ is used as an integer threshold, it means $\lfloor N/2 \rfloor$ unless an ideal real half is explicitly used only inside an entropy upper bound.

3 Cancellation traces and random gaps

This section isolates the parts of the argument common to the scaled and unscaled ensembles. The guiding separation is

$$\text{algebraic cancellation trace} \qquad \text{versus} \qquad \text{random support geometry.}$$

3.1 A random interleaver as order plus support

Let $B \subseteq [N]$ be a fixed set of M active input coordinates, and let π be the permutation associated with a uniform random permutation matrix Π, so that

$$e_i \Pi = e_{\pi(i)}.$$

Write

$$\pi(B) = \{p_1 < \cdots < p_M\}, \qquad \lambda_j := \pi^{-1}(p_j).$$

The set $\{p_1, \ldots, p_M\}$ is the *active support*; the sequence $(\lambda_1, \ldots, \lambda_M)$ is the *active order*.

Lemma 3.1 (Order–support decomposition)**.** *The active support $\pi(B)$ is a uniform M-subset of $[N]$. Conditional on this support, the active order is a uniform ordering of the elements of B. Consequently, the active support and active order are independent.*

Proof. The restriction of a uniform permutation to B is a uniform injection from B into $[N]$. Every M-subset is the range of exactly $M!$ such injections, and conditional on the range all $M!$ assignments of the elements of B to its positions are equally likely. □

When an independent random diagonal matrix follows the interleaver, the sequence of diagonal entries encountered at the active positions, read from left to right, is an independent uniform sequence in $(\mathbb{F}_q^*)^M$, and is independent of both the active order and the unordered active support. We use this equivalent deferred-decision representation throughout the master bounds.

3.2 Accumulator traces

Consider an accumulator input $z \in \mathbb{F}_q^N$ with active support

$$\{p_1 < \cdots < p_M\}, \qquad c_i := z_{p_i} \neq 0.$$

Define the active prefix sums

$$\sigma_i := \sum_{j=1}^{i} c_j, \qquad 1 \le i \le M,$$

and put $p_{M+1} := N + 1$ and

$$\Delta_i := p_{i+1} - p_i.$$

The exact zero set

$$Z(z) := \{i \in [M] : \sigma_i = 0\}$$

is the *cancellation trace* of this accumulator.

Lemma 3.2 (Nonadjacency and the gap identity)**.** *For every nonzero active sequence $c_1, \dots, c_M$:*

(i) $1 \notin Z(z)$*;*

(ii) $Z(z)$ *contains no two consecutive indices;*

(iii)

$$\mathrm{wt}(zA) = \sum_{i \in [M] \setminus Z(z)} \Delta_i. \tag{3.2.1}$$

Consequently, the number of possible size-a traces is at most

$$D(M, a) := \binom{M - a}{a}, \qquad 0 \le a \le \lfloor M/2 \rfloor . \tag{5}$$

Proof. The first prefix is $c_1 \neq 0$. If $\sigma_i = \sigma_{i+1} = 0$, then $c_{i+1} = \sigma_{i+1} - \sigma_i = 0$, a contradiction. Between p_i and $p_{i+1} - 1$, the ordinary coordinate prefix sum is the constant σ_i, proving the gap identity. Finally, write a nonadjacent a-subset as

$$2 \le i_1 < \cdots < i_a \le M, \qquad i_{k+1} \ge i_k + 2.$$

The map $j_k = i_k - k$ is a bijection with the a-subsets of $[M - a]$, whose number is $\binom{M-a}{a}$. □

3.3 One shared rank budget for two adaptive stages

Fix a message support of size w, and collect its w nonzero values into a vector x in a w-dimensional vector space V. Once the relevant orders, support geometry, and any scaling coefficients are fixed, every accumulator prefix sum is a linear function of x.

Let $\ell_1, \dots, \ell_m \in V^*$ be the first-stage forms and define

$$Z_1(x) := \{i \in [m] : \ell_i(x) = 0\}.$$

The second-stage forms may depend on the complete first trace. Thus, for every possible $Z \subseteq [m]$, let $g_1^Z, \dots, g_h^Z \in V^*$, and define

$$Z_2(x) := \{j \in [h] : g_j^{Z_1(x)}(x) = 0\}.$$

For $M \ge 0$, define

$$\mathcal{A}(M, w) := \sum_{d=0}^{w-1} \binom{M}{d}. \tag{6}$$

Theorem 3.3 (Two-stage shared-rank trace bound)**.** *The number of trace pairs* $(Z_1(x), Z_2(x))$ *realized by nonzero* $x \in V$ *is at most* $\mathcal{A}(m+h, w)$*.*

Proof. Fix a realized trace pair and put

$$U_1 := \operatorname{span}\{\ell_i : i \in Z_1\}.$$

Choose canonically a set $B_1 \subseteq Z_1$ whose forms form a basis of U_1. Next put

$$U_2 := \operatorname{span}\Big(U_1 \cup \{g_j^{Z_1} : j \in Z_2\}\Big),$$

and choose canonically $B_2 \subseteq Z_2$ whose forms extend the basis of U_1 to a basis of U_2. Every form in U_2 vanishes at x, so

$$|B_1| + |B_2| = \dim U_2 \le \dim x^{\perp} = w - 1.$$

The label (B_1, B_2) determines the complete trace. Indeed,

$$U_1 = \operatorname{span}\{\ell_i : i \in B_1\}, \qquad Z_1 = \{i \in [m] : \ell_i \in U_1\}.$$

The reverse inclusion in the second identity includes every dependent zero equation: every form in U_1 vanishes at x. Once Z_1 is known, the adaptive second family is known; then

$$U_2 = \operatorname{span}\Big(U_1 \cup \{g_j^{Z_1} : j \in B_2\}\Big), \qquad Z_2 = \{j \in [h] : g_j^{Z_1} \in U_2\}.$$

Thus distinct traces have distinct canonical labels. If $d_i := |B_i|$, then the number of labels of sizes (d_1, d_2) is at most $\binom{m}{d_1}\binom{h}{d_2}$. Summing over $d_1 + d_2 \le w - 1$ and using Vandermonde's identity gives

$$\sum_{d_1+d_2 \le w-1} \binom{m}{d_1}\binom{h}{d_2} = \sum_{d=0}^{w-1} \binom{m+h}{d} = \mathcal{A}(m+h, w).$$

□

For an RAA encoder, a message of weight w gives $m := rw$ first-stage prefix forms. Conditional on first-output weight h, define

$$J(m, h, w) := \mathcal{A}(m+h, w) = \sum_{d=0}^{w-1} \binom{m+h}{d}. \tag{7}$$

If the traces have exactly a and b zeros, nonadjacency gives the separate count

$$D(m, a)D(h, b) = \binom{m-a}{a}\binom{h-b}{b}.$$

Remark 3.4 (More than two stages)*.* The argument extends directly to any fixed number of adaptive linear stages. If stage i contains at most M_i forms, then the number of realized layered traces is at most

$$\sum_{d=0}^{w-1} \binom{M_1 + \cdots + M_L}{d}.$$

3.4 Exact random-gap laws

If $1 \le p_1 < \cdots < p_M \le N$, set $p_{M+1} := N+1$ and

$$g_0 := p_1 - 1, \qquad g_i := p_{i+1} - p_i - 1 \quad (1 \le i \le M).$$

Then $(g_0, \ldots, g_M)$ is a weak composition of $N - M$ into $M+1$ parts. Conversely, every such weak composition reconstructs a unique M-subset by

$$p_1 = g_0 + 1, \qquad p_i = i + \sum_{j=0}^{i-1} g_j \quad (2 \le i \le M).$$

Thus the weak-composition correspondence is a bijection, and a uniform M-subset induces the uniform distribution on these compositions.

Lemma 3.5 (Exact prescribed-gap sum)**.** *Let $\{p_1 < \cdots < p_M\}$ be a uniform M-subset of $[N]$, put $p_{M+1} := N+1$, and define $\Delta_i := p_{i+1} - p_i$. For every fixed nonempty $S \subseteq [M]$,*

$$\mathbb{P}\left[\sum_{i \in S} \Delta_i = h\right] = \frac{\binom{h-1}{|S|-1}\binom{N-h}{M-|S|}}{\binom{N}{M}}. \tag{8}$$

In particular, the distribution depends on S only through $|S|$.

Proof. Put $j := |S|$. Since $\Delta_i = 1 + g_i$, the event is equivalent to

$$\sum_{i \in S} g_i = h - j.$$

The selected j coordinates have $\binom{h-1}{j-1}$ weak compositions. The remaining $M - j + 1$ coordinates have total $N - M - h + j$, giving $\binom{N-h}{M-j}$ choices. Divide by the total number $\binom{N}{M}$ of supports. This proves equality in distribution for all prescribed j-sets of gaps; it does not identify their events pointwise. □

For the first accumulator, a trace of size a leaves $m - a$ nonzero active prefixes. Define

$$g_{1,N}(m, a, h) := \frac{\binom{h-1}{m-a-1}\binom{N-h}{a}}{\binom{N}{m}}. \tag{9}$$

By lemmas 3.2 and 3.5, this is the exact conditional law of the first output weight for every fixed exact trace of size a.

For the second accumulator, let $L := \lfloor \delta N \rfloor$. If its input has weight h and its trace has b zeros, then $h - b$ gaps contribute to the final output. Define

$$g_{2,N,\delta}(h, b) := \frac{1}{\binom{N}{h}} \sum_{\ell=h-b}^{L} \binom{\ell - 1}{h - b - 1}\binom{N - \ell}{b}. \tag{10}$$

Lemma 3.6 (Hypergeometric form of the second transition)**.** *Let $H = \{N - L + 1, \ldots, N\}$, and let X be the number of points of a uniform h-subset of $[N]$ that lie in H. Then*

$$X \sim \mathrm{Hyp}(N, L, h)$$

and

$$g_{2,N,\delta}(h, b) = \mathbb{P}[X \ge h - b].$$

In particular, the low-output probability depends on the exact second trace only through b.

Proof. Put $k := h - b$. For a fixed second-stage trace, the final output weight is the sum of a prescribed set of k gaps. By lemma 3.5, the distribution of this sum depends only on the number k of selected gaps. We may therefore replace it in distribution by the last k gaps,

$$\Delta_{b+1}, \dots, \Delta_h.$$

Since $p_{h+1} = N + 1$, their sum is $N + 1 - p_{b+1}$. Hence

$$N + 1 - p_{b+1} \le L \quad \Longleftrightarrow \quad p_{b+1} \ge N - L + 1.$$

The latter event says that at most b selected positions lie before the terminal interval H, or equivalently that at least $h - b$ lie in H. This is exactly $\mathbb{P}[X \ge h - b]$. □

The following transform is used repeatedly in both upper bounds.

Lemma 3.7 (First-gap generating transform)**.** *Let $1 \le m \le N$, $0 \le a \le m - 1$, and $0 < \rho < 1$. Put $k := m - a$. Then*

$$\sum_{h=1}^{N} g_{1,N}(m, a, h)\rho^h \le \left(\frac{m}{N}\frac{\rho}{1-\rho}\right)^k. \tag{11}$$

Proof. Using $h \ge k$ on every nonzero term and $\binom{N-h}{a} \le \binom{N-k}{a}$,

$$\begin{aligned}\sum_{h} g_{1,N}(m, a, h)\rho^h &\le \frac{\binom{N-k}{a}}{\binom{N}{m}} \sum_{h \ge k} \binom{h-1}{k-1}\rho^h \\ &= \frac{\binom{N-k}{a}}{\binom{N}{m}} \left(\frac{\rho}{1-\rho}\right)^k.\end{aligned}$$

Since $m = a + k$,

$$\frac{\binom{N-k}{a}}{\binom{N}{m}} = \prod_{i=1}^{k} \frac{a+i}{N-k+i}.$$

For $1 \le i \le k$,

$$\frac{a+i}{N-k+i} \le \frac{m}{N},$$

because this inequality is equivalent to $N(i-k) \le m(i-k)$. Multiplying the k factors proves (11). □

4 Unified master bounds

Fix a message support $T \subseteq [n]$ of size w, and put

$$m := rw.$$

For a, h, b, let $\mathcal{E}_{\bullet,T}(a, h, b; \delta)$ be the event that some message supported exactly on T has first trace size a, first-output weight h, second trace size b, and final output weight at most L_δ, in the indicated ensemble. Then

$$\mathcal{E}_{\bullet,w}(\delta) \subseteq \bigcup_{|T|=w} \bigcup_{a,h,b} \mathcal{E}_{\bullet,T}(a, h, b; \delta).$$

4.1 Prescribed scalar anti-cancellation

Multiplying all nonzero message values on a fixed support by a common scalar changes neither trace. Thus there are Q^{w-1} projective assignments of nonzero values on T.

Lemma 4.1 (Prescribed two-stage anti-cancellation)**.** *Fix all of the following data:*

- *a message support T of size w;*
- *the first active order and the exact first unordered active support $P = \{p_1 < \cdots < p_m\}$;*
- *a prescribed set $I_1 \subseteq [m]$ of size a;*
- *an exact target first-output support $S \subseteq [N]$ of size h;*
- *the active order of the elements of S at the second interleaver;*
- *the exact second unordered active support $P' = \{p'_1 < \cdots < p'_h\}$; and*
- *a prescribed set $I_2 \subseteq [h]$ of size b.*

Conditional on these data, the probability over the active diagonal entries of V_1, V_2 that there exists a message x supported exactly on T satisfying

$$I_1 \subseteq Z_1(x), \qquad \operatorname{supp}(xR\Pi_1 V_1 A) = S, \qquad I_2 \subseteq Z_2(xR\Pi_1 V_1 A)$$

is at most

$$Q^{-(a+b-(w-1))_+}. \tag{12}$$

The bound is uniform in the two exact supports P, P'.

Proof. Fix one of the Q^{w-1} projective assignments of nonzero values on T. Expose the active entries of V_1 in increasing active-position order. Process the indices of I_1 in increasing order. At a prescribed first-stage zero index i, the prefix equation has the form

$$c + \xi_i x_{\lambda_i} = 0,$$

where c is determined by previously exposed active scalars and $x_{\lambda_i} \neq 0$. At most one of the Q possible values of the fresh scalar ξ_i satisfies the equation. The first-stage conditions therefore cost at most Q^{-a}.

Condition on a complete first-stage realization satisfying those equations and the exact support requirement. For every element of the exact support S, the corresponding first-output value is now fixed and nonzero. Expose the active entries of V_2 in increasing order on the fixed support P', and process I_2 in increasing order. At a prescribed second-stage zero index j, the equation is

$$c' + \eta_j y_{\mu_j} = 0,$$

where $y_{\mu_j} \neq 0$ because $\mu_j \in S$, and η_j is fresh. Thus the second-stage conditions cost at most Q^{-b}.

For this projective assignment the probability is at most Q^{-a-b}. Union bounding over Q^{w-1} assignments gives $Q^{w-1-a-b}$ when $a + b > w - 1$, and the trivial upper bound 1 otherwise. Nonvanishing conditions required by the exact support S can only decrease the probability. □

4.2 Canonical-label exposure before support gaps

The arrangement branch requires care because the first-output support is witness-dependent. The following proof exposes a canonical first basis label before the first unordered support, and a canonical second basis label before the second unordered support.

Lemma 4.2 (Arrangement branch)**.** *For either ensemble, after averaging over all interleaver orders and, in the scaled case, over all active scalar sequences,*

$$\mathbb{P}[\mathcal{E}_{\bullet,T}(a,h,b;\delta)] \le J(m,h,w)\, g_{1,N}(m,a,h)\, g_{2,N,\delta}(h,b). \tag{13}$$

Proof. Use the deferred-decision representation following lemma 3.1. First condition on the first active order and, in the scaled ensemble, on the first active scalar sequence. The first prefix forms $\ell_1,\dots,\ell_m$ are now fixed, while the first unordered support remains a uniform m-subset.

Every realized trace pair has a canonical first basis label $B_1 \subseteq [m]$. For any proposed B_1, put

$$U_1 := \operatorname{span}\{\ell_i : i \in B_1\}, \qquad Z_1(B_1) := \{i : \ell_i \in U_1\}.$$

If B_1 is the canonical label of an actual trace, this formula recovers its complete first trace, including dependent zero equations. Hence the first trace is fixed before the first unordered support is exposed. For a label producing $|Z_1(B_1)| = a$, the probability that the first output has weight h is exactly $g_{1,N}(m,a,h)$. Once the first unordered support is exposed, its exact output support S is determined by that support and $Z_1(B_1)$.

Conditional on this fixed S, expose the second active order and, in the scaled case, the second active scalar sequence, but not the second unordered support. The second family of forms is now fixed. For a proposed $B_2 \subseteq [h]$, put

$$U_2 := \operatorname{span}\Big(U_1 \cup \{g_j^{Z_1(B_1)} : j \in B_2\}\Big), \qquad Z_2(B_1,B_2) := \{j : g_j^{Z_1(B_1)} \in U_2\}.$$

If (B_1,B_2) is the canonical label of an actual trace pair, this recovers the complete second trace. Thus, for a label pair producing $|Z_2| = b$, the second unordered support is still uniform and gives the exact low-output probability $g_{2,N,\delta}(h,b)$.

If $d_i = |B_i|$, every actual canonical pair satisfies $d_1 + d_2 \le w-1$. Union bounding over all proposed labels therefore costs at most

$$\sum_{d_1+d_2\le w-1} \binom{m}{d_1}\binom{h}{d_2} = J(m,h,w).$$

The bound is uniform in the conditioned orders and scalar sequences, so averaging them proves (13). □

Proposition 4.3 (Unified exact master bounds)**.** *For every $1 \le w \le n$, with $m = rw$,*

$$\mathbb{P}[\mathcal{E}_{0,w}(\delta)] \le \binom{n}{w} \sum_{a=0}^{\lfloor m/2\rfloor} \sum_{h=1}^{N} \sum_{b=0}^{\lfloor h/2\rfloor} g_{1,N}(m,a,h) g_{2,N,\delta}(h,b) \times \min\{J(m,h,w),\ D(m,a)D(h,b)\}, \tag{14}$$

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc},w}(\delta)] \le \binom{n}{w} \sum_{a=0}^{\lfloor m/2\rfloor} \sum_{h=1}^{N} \sum_{b=0}^{\lfloor h/2\rfloor} g_{1,N}(m,a,h) g_{2,N,\delta}(h,b) \times \min\Big\{J(m,h,w),\ D(m,a)D(h,b)Q^{-(a+b-(w-1))_+}\Big\}. \tag{15}$$

Table 2: Conditioning ledger for proposition 4.3. A label is exposed before the unordered support whose gaps determine its output weight.

Step	Fixed data	Randomness still averaged
First algebra	message support, first active order; first active scalar sequence in the scaled arrangement branch	first unordered support
First label	canonical first basis label, hence complete first trace	uniform first support gaps, giving g_1
Second algebra	exact first-output support and its linear-form labels; second active order; second active scalar sequence in the scaled arrangement branch	second unordered support
Second label	canonical second basis label, hence complete second trace	uniform second support gaps, giving g_2
Scalar branch	both active orders, both exact unordered supports, exact first-output support, prescribed trace sets	fresh active entries of V_1, V_2 and projective message values

Proof. Fix T and (a, h, b). The arrangement bound (13) is valid for both ensembles.

For the unscaled prescribed-trace branch, there are at most $D(m, a)$ possible exact first traces and, after the first-output support has size h, at most $D(h, b)$ possible exact second traces. Fix such a pair. The first exact trace is determined before the first unordered support is exposed, so the probability of first-output weight h is exactly g_1. After the resulting exact first-output support is fixed, the second active order is exposed, while the second unordered support remains uniform; the fixed second trace then has low-output probability exactly g_2. Thus the prescribed-trace branch is bounded by $D(m, a)D(h, b)g_1g_2$.

For the scaled prescribed-trace branch, fix exact nonadjacent trace sets I_1, I_2, the first active order, both exact unordered supports, and the second active order. If the first trace is exactly I_1, its exact output support is determined by the first unordered support; require that this support have size h. Conditional on all these exact supports and orders, lemma 4.1 bounds the probability over the active scalars of an existential witness by $Q^{-(a+b-(w-1))_+}$. This bound is uniform in both unordered supports. Averaging the first support gives g_1, and averaging the second support gives g_2. Summing over the prescribed nonadjacent trace sets gives

$$D(m, a)D(h, b)Q^{-(a+b-(w-1))_+}g_1g_2.$$

The arrangement and prescribed-trace expressions are independent upper bounds on the same (T, a, h, b) subevent, so their minimum is valid. Finally sum over a, h, b and union bound over the $\binom{n}{w}$ message supports. □

The distinction between an exact support and its cardinality is essential. Whenever an exact support determines a later family of linear forms, it is fixed before that family is defined. Only after this conditioning do we use the fact that the next random-gap law depends on the support through its cardinality.

5 The unscaled ensemble

We prove the upper bound in theorem 1.2. Throughout this section,

$$0 < \delta \le \delta_0 := \frac{1}{10}, \qquad \rho_0 := 2\sqrt{\delta_0(1-\delta_0)} = \frac{3}{5},$$

and

$$x_0 := \frac{1}{200000}.$$

Because $\mathcal{E}_{0,w}(\delta) \subseteq \mathcal{E}_{0,w}(\delta_0)$, it is enough to prove the upper bound at δ_0.

5.1 A fixed-trace tail bound

Lemma 5.1 (Uniform fixed-trace tail)**.** *For $0 < \delta \le 1/2$, every $1 \le h \le N$, and every $0 \le b \le \lfloor h/2 \rfloor$,*

$$g_{2,N,\delta}(h,b) \le \left(2\sqrt{\delta(1-\delta)}\right)^h. \tag{16}$$

In particular, for $\delta \le 1/10$,

$$g_{2,N,\delta}(h,b) \le (3/5)^h.$$

Proof. Put $L = \lfloor \delta N \rfloor$. If $L = 0$, then the final output cannot have positive contributing-gap sum at most L, so $g_{2,N,\delta}(h,b) = 0$. Assume henceforth that $L \ge 1$, and put $\delta_N = L/N$.

By lemma 3.6, $g_{2,N,\delta}(h,b) = \mathbb{P}[X \ge h-b]$ for $X \sim \mathrm{Hyp}(N,L,h)$. Since $b \le h/2$, this is at most $\mathbb{P}[X \ge h/2]$. For $z \ge 1$, the elementary-symmetric-mean inequality gives

$$\mathbb{E}[z^X] \le (1 - \delta_N + \delta_N z)^h.$$

Indeed, the left side is the normalized h-th elementary symmetric polynomial of L copies of z and $N-L$ copies of 1, and Maclaurin's inequality bounds it by the h-th power of their arithmetic mean. Markov's inequality therefore gives

$$\mathbb{P}[X \ge h/2] \le z^{-h/2}(1 - \delta_N + \delta_N z)^h.$$

Taking $z = (1-\delta_N)/\delta_N \ge 1$ yields $[2\sqrt{\delta_N(1-\delta_N)}]^h$. The base is increasing on $[0, 1/2]$, and $\delta_N \le \delta$, proving the claim. □

Lemma 5.2 (One-layer trace count)**.** *If $M \ge 2w \ge 2$, then*

$$\mathcal{A}(M,w) = \sum_{d=0}^{w-1} \binom{M}{d} \le \left(\frac{4eM}{w}\right)^w. \tag{17}$$

Proof. The binomial coefficients are increasing through $d = w$ because $M \ge 2w$. Thus

$$\mathcal{A}(M,w) \le w\binom{M}{w} \le w\left(\frac{eM}{w}\right)^w \le \left(\frac{4eM}{w}\right)^w.$$

□

5.2 Weight one

Lemma 5.3 (Unscaled weight one)**.** *If* $\mathrm{char}(\mathbb{F}_q) > r$*, then*

$$\mathbb{P}[\mathcal{E}_{0,1}(\delta)] = O_{r,\delta}(N^{1-r}).$$

Proof. Fix a one-symbol message support and let its nonzero value be c. The active first increments are r copies of c, in any order. Their active prefixes are $c, 2c, \dots, rc$, all nonzero because the characteristic exceeds r. Hence the first trace is empty.

The message-value space is one-dimensional. Once the exact first-output support and the second active order are fixed, every second-stage prefix form is a scalar multiple of the single message variable. Its zero/nonzero status is therefore independent of the chosen nonzero message value, so the second exact trace is unique. By lemma 5.1, with $\rho = 2\sqrt{\delta(1-\delta)} < 1$,

$$\mathbb{P}[\text{this support succeeds}] \le \sum_h g_{1,N}(r,0,h)\rho^h \le \left(\frac{r}{N}\frac{\rho}{1-\rho}\right)^r$$

by lemma 3.7. There are $n = N/r$ possible supports. □

5.3 Very sparse messages

Lemma 5.4 (Sparse unscaled weights)**.** *Let* $r \ge 4$*,* $w \ge 2$*,* $m = rw$*, and suppose*

$$x := \frac{m}{N} \le x_0 = \frac{1}{200000}.$$

Then

$$\mathbb{P}[\mathcal{E}_{0,w}(\delta)] \le \left[B_r \left(\frac{w}{N}\right)^{r/2-1}\right]^w, \qquad B_r := \frac{16e^2(4r)^{r/2}}{\log(4/3)}. \tag{18}$$

Consequently,

$$\sum_{w\ge 2:\ rw/N \le x_0} \mathbb{P}[\mathcal{E}_{0,w}(\delta)] = O_r(N^{2-r}). \tag{19}$$

Proof. Fix a message support T. Condition on the first active order. The first prefix forms have at most $\mathcal{A}(m,w)$ exact zero traces. For a fixed first trace of size a, the first output weight has law $g_{1,N}(m,a,h)$. After the exact first-output support is fixed, condition on the second active order. The second prefix forms have at most $\mathcal{A}(h,w)$ exact traces, and each such trace has low-output probability at most $(3/5)^h$ by lemma 5.1. Hence

$$\mathbb{P}[\text{a witness on } T] \le \mathcal{A}(m,w) \max_{0\le a\le m/2} \sum_h g_{1,N}(m,a,h)\mathcal{A}(h,w)(3/5)^h. \tag{20}$$

Nonadjacency gives $h \ge m - a \ge m/2 \ge 2w$, so lemma 5.2 applies to $\mathcal{A}(h,w)$. Write

$$\frac{3}{5} = \frac{3}{4}\cdot\frac{4}{5}, \qquad \ell := \log(4/3).$$

Then

$$\begin{aligned}\mathcal{A}(h,w)(3/5)^h &\le \left(\frac{4eh}{w}\right)^w (3/4)^h(4/5)^h \\ &\le \left(\frac{4}{\ell}\right)^w (4/5)^h,\end{aligned}$$

where

$$\sup_{y\ge 0} y^w (3/4)^y = \left(\frac{w}{e\ell}\right)^w.$$

By lemma 3.7,

$$\sum_h g_{1,N}(m,a,h)(4/5)^h \le (4m/N)^{m-a}.$$

Because $4m/N = 4x \le 1$ and $m - a \ge m/2$, this is at most $(4x)^{m/2}$. Also

$$\mathcal{A}(m,w) \le (4em/w)^w = (4er)^w.$$

Substituting into (20) and union bounding over $\binom{n}{w} \le (e/x)^w$ supports gives

$$\mathbb{P}[\mathcal{E}_{0,w}(\delta)] \le \left[\frac{16e^2 r}{\ell}\, 4^{r/2} x^{r/2-1}\right]^w.$$

Replacing x by rw/N gives (18).

Put $a_r = r/2 - 1$ and let U_w denote the right side of (18). For $w < Nx_0/r$,

$$\frac{U_{w+1}}{U_w} \le B_r e^{a_r} \left(\frac{x_0}{r} + \frac{1}{N}\right)^{a_r}.$$

For each fixed $r \ge 4$, this is at most $1/2$ for sufficiently large N. The limiting expression is largest at $r = 4$: its ratio at consecutive r is

$$2\sqrt{ex_0}\frac{r+1}{r} < 0.01,$$

and the $r = 4$ value is below 0.36. Hence

$$\sum_{w\ge 2:\ rw/N \le x_0} U_w \le 2U_2 = O_r(N^{-2a_r}) = O_r(N^{2-r}).$$

□

5.4 A two-parameter entropy function for one accumulator

For $0 \le u \le 1$, let

$$\mathsf{h}(u) := -u\log u - (1-u)\log(1-u), \qquad 0\log 0 := 0,$$

We use the continuous perspective convention

$$s\,\mathsf{h}(a/s) := 0 \qquad \text{when } s = a = 0.$$

and for a probability vector $\boldsymbol{c}$, let

$$\mathsf{H}(\boldsymbol{c}) := -\sum_i c_i \log c_i.$$

Consider one accumulator with $M = \xi N$ active inputs, $H = \tau N$ nonzero outputs, and $K = \zeta N$ zero prefixes. Its nonadjacency count and exact gap law contribute

$$\binom{M-K}{K} g_{1,N}(M,K,H).$$

The feasibility conditions are

$$\xi - 2\zeta \geq 0, \qquad \tau - \xi + \zeta \geq 0, \qquad 1 - \tau - \zeta \geq 0.$$

Define

$$I(\xi, \tau) := \{\zeta \geq 0 : \xi - 2\zeta \geq 0,\ \tau - \xi + \zeta \geq 0,\ 1 - \tau - \zeta \geq 0\}$$

and

$$\boldsymbol{c}(\xi, \tau, \zeta) := (\zeta, \zeta, \xi - 2\zeta, \tau - \xi + \zeta, 1 - \tau - \zeta).$$

Define

$$\mathcal{M}(\xi, \tau) := \max_{\zeta \in I(\xi,\tau)} \left\{ \mathsf{H}\big(\boldsymbol{c}(\xi, \tau, \zeta)\big) - \mathsf{h}(\xi) - \mathsf{h}(\tau) \right\}, \tag{21}$$

with $\mathcal{M}(\xi, \tau) := -\infty$ if $I(\xi, \tau)$ is empty.

We use the standard finite type bounds

$$\frac{1}{M+1} e^{M\mathsf{h}(K/M)} \leq \binom{M}{K} \leq e^{M\mathsf{h}(K/M)}. \tag{22}$$

For fixed integer $j \geq 0$, the function

$$A \longmapsto A\mathsf{h}(j/A), \qquad A \geq j, \tag{23}$$

is increasing, because its derivative is $-\log(1 - j/A) \geq 0$.

Lemma 5.5 (Finite entropy collapse)**.** *Let $r \geq 4$, $\delta_0 = 1/10$, and let*

$$x = \frac{rw}{N}, \qquad \alpha = \frac{a}{N}, \qquad t = \frac{h}{N}, \qquad \beta = \frac{b}{N}.$$

There is an absolute constant C such that every nonzero summand in the nonadjacency branch of (14), including the factor $\binom{n}{w}$, is at most

$$(N+1)^C \exp\left\{ N \left(\frac{1}{r}\mathsf{h}(x) + \mathcal{M}(x, t) + \mathcal{M}(t, \delta_0) \right) \right\}. \tag{24}$$

Every such nonzero summand satisfies

$$0 < x \leq 4\delta_0 = \frac{2}{5}, \qquad \frac{x}{2} \leq t \leq 2\delta_0 = \frac{1}{5}. \tag{25}$$

For an upper bound uniform over $r \geq 4$, the term $\mathsf{h}(x)/r$ may be replaced by $\mathsf{h}(x)/4$.

Proof. The support count obeys

$$\binom{n}{w} \leq \exp\left\{ \frac{N}{r}\mathsf{h}(x) \right\},$$

because $w/n = x$.

For the first transition, the following identity is exact whenever the term is nonzero:

$$D(m, a) g_{1,N}(m, a, h) = \frac{m - a}{h} \frac{\binom{N}{a, a, m - 2a, h - m + a, N - h - a}}{\binom{N}{m}\binom{N}{h}}. \tag{26}$$

The prefactor is at most one because $h \geq m - a$. Applying the multinomial upper bound and the two binomial lower bounds from (22) gives

$$D(m,a)g_{1,N}(m,a,h) \leq (N+1)^2 \exp\{NF(x,t,\alpha)\},$$

where

$$F(x,t,\alpha) := \mathsf{H}(\boldsymbol{c}(x,t,\alpha)) - \mathsf{h}(x) - \mathsf{h}(t) \leq \mathcal{M}(x,t).$$

This exact identity accounts for both -1 shifts in the gap binomial.

For the second transition, put $L = \lfloor \delta_0 N \rfloor$ and $j = h - b$. If the term is nonzero, then $j \leq L$ and $j \geq h/2$. For $X \sim \mathrm{Hyp}(N, L, h)$, the ratio of consecutive point masses is

$$\frac{\mathbb{P}[X = k+1]}{\mathbb{P}[X = k]} = \frac{(L-k)(h-k)}{(k+1)(N-L-h+k+1)}.$$

For $k \geq h/2$ and $L \leq N/2$, we have

$$L - k \leq N - L - h + k, \qquad h - k \leq k,$$

so the ratio is at most one. Therefore

$$g_{2,N,\delta_0}(h,b) \leq (h+1)\frac{\binom{L}{j}\binom{N-L}{b}}{\binom{N}{h}}. \tag{27}$$

By (23),

$$\binom{L}{j} \leq \exp\left\{\delta_0 N \mathsf{h}\left(\frac{t-\beta}{\delta_0}\right)\right\}.$$

Moreover, $N - L = \lceil (1-\delta_0)N \rceil$. Let $A_0 = \lfloor (1-\delta_0)N \rfloor$. Since every nonzero term has $b \leq h/2 \leq L$, we have $b \leq A_0$ for all sufficiently large N. If $N - L = A_0 + 1$, then

$$\binom{A_0+1}{b} = \frac{A_0+1}{A_0+1-b}\binom{A_0}{b} \leq (N+1)\binom{A_0}{b}.$$

Using (23) again,

$$\binom{N-L}{b} \leq (N+1)\exp\left\{(1-\delta_0)N\mathsf{h}\left(\frac{\beta}{1-\delta_0}\right)\right\}.$$

Finally,

$$D(h,b) = \binom{h-b}{b} \leq \exp\left\{(t-\beta)N\mathsf{h}\left(\frac{\beta}{t-\beta}\right)\right\},$$

and the denominator in (27) contributes $-N\mathsf{h}(t)$ up to another factor $N+1$. Thus

$$D(h,b)g_{2,N,\delta_0}(h,b) \leq (N+1)^3 \exp\{NF(t,\delta_0,\beta)\} \leq (N+1)^3 e^{N\mathcal{M}(t,\delta_0)}.$$

All finitely many small N can be absorbed by increasing C. Multiplying the three bounds proves (24).

For the projected domain, nonzero first-transition factors imply $h \geq m - a \geq m/2$, so $t \geq x/2$. Nonzero second-transition factors imply

$$h/2 \leq h - b \leq L \leq \delta_0 N,$$

so $t \leq 2\delta_0$, and hence $x \leq 2t \leq 4\delta_0$. Finally, $\mathsf{h}(x)/r \leq \mathsf{h}(x)/4$ for $r \geq 4$. □

5.5 An analytic entropy margin

For probability vectors $\boldsymbol{p}, \boldsymbol{q}$, write

$$D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{q}) := \sum_i p_i \log \frac{p_i}{q_i}.$$

Lemma 5.6 (Analytic one-transition bound)**.** *Let $0 < \tau \le 1/5$, $0 < \xi < 1$, and suppose $I(\xi,\tau) \neq \varnothing$. Define*

$$\mathsf{B}(\tau) := \tau + 2\sqrt{\tau(1-\tau)}.$$

Then

$$\mathcal{M}(\xi,\tau) \le \xi \log \mathsf{B}(\tau) - \frac{\xi^2}{2}. \tag{28}$$

At $\tau = 1/10$, the stronger estimate

$$\mathcal{M}\left(\xi, \frac{1}{10}\right) \le \xi \log \frac{7}{10} - \frac{121}{98}\xi^2 \tag{29}$$

holds.

Proof. Fix $\zeta \in I(\xi,\tau)$, and put

$$u := \frac{\zeta}{\xi}, \qquad v := \frac{\tau - \xi + \zeta}{1-\xi}.$$

Then

$$\boldsymbol{c}(\xi,\tau,\zeta) = \big(\xi u, \xi u, \xi(1-2u), (1-\xi)v, (1-\xi)(1-v)\big)$$

and

$$\tau = \xi(1-u) + (1-\xi)v. \tag{30}$$

Writing

$$H_3(u) := -2u\log u - (1-2u)\log(1-2u),$$

we have

$$F(\xi,\tau,\zeta) := \mathsf{H}\big(\boldsymbol{c}(\xi,\tau,\zeta)\big) - \mathsf{h}(\xi) - \mathsf{h}(\tau) = \xi H_3(u) + (1-\xi)\mathsf{h}(v) - \mathsf{h}(\tau). \tag{31}$$

Define

$$u_\tau := \frac{\sqrt{1-\tau}}{2\sqrt{1-\tau} + \sqrt{\tau}}.$$

Using

$$\mathsf{B}(\tau)u_\tau = \sqrt{\tau(1-\tau)}, \qquad \mathsf{B}(\tau)(1-2u_\tau) = \tau,$$

a direct expansion gives

$$\begin{aligned} F(\xi,\tau,\zeta) = \xi \log \mathsf{B}(\tau) &- \xi D_{\mathrm{KL}}((u,u,1-2u)\,\|\,(u_\tau,u_\tau,1-2u_\tau)) \\ &- (1-\xi) D_{\mathrm{KL}}((v,1-v)\,\|\,(\tau,1-\tau)). \end{aligned} \tag{32}$$

The first divergence, as a function of u, has second derivative

$$\frac{2}{u} + \frac{4}{1-2u} \ge 16,$$

so it is at least $8(u-u_\tau)^2$. Feasibility gives $0 \le u \le 1/2$. From (30),

$$v = \frac{\tau - \xi(1-u)}{1-\xi} \le \tau,$$

and feasibility gives $v \ge 0$. On $[0,\tau]$, the second derivative of the binary divergence is at least $1/[\tau(1-\tau)]$, so

$$D_{\mathrm{KL}}((v,1-v)\|(\tau,1-\tau)) \ge \frac{(v-\tau)^2}{2\tau(1-\tau)}.$$

Because

$$v - \tau = \frac{\xi}{1-\xi}\big(u-(1-\tau)\big),$$

we obtain

$$F(\xi,\tau,\zeta) \le \xi \log \mathsf{B}(\tau) - 8\xi(u-u_\tau)^2 - \frac{\xi^2}{2(1-\xi)\tau(1-\tau)}\big(u-(1-\tau)\big)^2.$$

Using

$$A(s-a)^2 + C(s-b)^2 \ge \frac{AC}{A+C}(a-b)^2$$

gives

$$F(\xi,\tau,\zeta) \le \xi \log \mathsf{B}(\tau) - \frac{8\xi^2(1-\tau-u_\tau)^2}{\xi + 16(1-\xi)\tau(1-\tau)}. \tag{33}$$

Put $s := \sqrt{\tau/(1-\tau)} \le 1/2$. Then

$$u_\tau = \frac{1}{2+s}, \qquad 1-\tau = \frac{1}{1+s^2},$$

and

$$1-\tau-u_\tau = \frac{1+s-s^2}{(1+s^2)(2+s)} \ge \frac{2}{5},$$

because

$$5(1+s-s^2) - 2(1+s^2)(2+s) = (1-2s)(s^2+5s+1) \ge 0.$$

Also

$$\xi + 16(1-\xi)\tau(1-\tau) \le \frac{64}{25}.$$

Thus (33) yields

$$F(\xi,\tau,\zeta) \le \xi \log \mathsf{B}(\tau) - \frac{\xi^2}{2}.$$

Maximizing over ζ proves (28).

For $\tau = 1/10$,

$$\mathsf{B}(1/10) = \frac{7}{10}, \qquad u_{1/10} = \frac{3}{7}, \qquad \frac{9}{10} - \frac{3}{7} = \frac{33}{70},$$

and

$$\xi + 16(1-\xi)\frac{1}{10}\frac{9}{10} \le \frac{36}{25}.$$

Substitution into (33) gives (29). Boundary cases follow by continuity. □

Lemma 5.7 (Unscaled entropy margin)**.** *For every*

$$0 < x \le \frac{2}{5}, \qquad \frac{x}{2} \le t \le \frac{1}{5},$$

one has

$$\frac{1}{4}\mathsf{h}(x) + \mathcal{M}(x,t) + \mathcal{M}(t,1/10) \le -\frac{x}{100}. \tag{34}$$

Proof. Put

$$L := \log\frac{7}{10} < 0, \qquad c := \frac{121}{98}.$$

By lemma 5.6,

$$\begin{aligned}&\frac{1}{4}\mathsf{h}(x) + \mathcal{M}(x,t) + \mathcal{M}(t,1/10)\\ &\qquad \le \frac{1}{4}\mathsf{h}(x) + x\log\mathsf{B}(t) - \frac{x^2}{2} + tL - ct^2.\end{aligned} \tag{35}$$

We use

$$\mathsf{B}(t) \le \sqrt{5t}. \tag{36}$$

Indeed, this is equivalent to $4\sqrt{t(1-t)} \le 1 + 3t$, whose squared difference is $(1-5t)^2$. We also use

$$\frac{\mathsf{h}(x)}{x} \le \log\frac{1}{x} + 1 - \frac{x}{2}, \qquad 0 < x < 1. \tag{37}$$

This follows from

$$\frac{1-x}{x}\log\frac{1}{1-x} = 1 - \sum_{k\ge1}\frac{x^k}{k(k+1)} \le 1 - \frac{x}{2}.$$

Add $x/100$ to the right side of (35), divide by x, and put $y := t/x$. Then

$$\begin{aligned}&\frac{1}{x}\left[\frac{1}{4}\mathsf{h}(x) + \mathcal{M}(x,t) + \mathcal{M}(t,1/10) + \frac{x}{100}\right]\\ &\qquad \le \frac{1}{4}\log x + \frac{1}{2}\log(5y) + \frac{13}{50} + yL - x\left(\frac{5}{8} + cy^2\right).\end{aligned} \tag{38}$$

For fixed $y > 0$,

$$\frac{1}{4}\log x - Ax \le -\frac{1}{4}\log(4A) - \frac{1}{4}, \qquad A := \frac{5}{8} + cy^2.$$

Hence the right side of (38) is at most

$$J(y) := \frac{1}{2}\log(5y) - \frac{1}{4}\log\left(\frac{5}{2} + 4cy^2\right) + yL + \frac{1}{100}.$$

By arithmetic–geometric mean,

$$\frac{5}{2} + 4cy^2 \ge 2y\sqrt{10c},$$

so

$$J(y) \le \frac{1}{4}\log\left(\frac{25y}{2\sqrt{10c}}\right) + yL + \frac{1}{100}.$$

The right side is strictly concave and is maximized at $y_* = -1/(4L)$. Its maximum equals

$$\frac{1}{4}\log\left(\frac{175}{8\sqrt{605}\log(10/7)}\right) - \frac{6}{25}. \tag{39}$$

The inequality $\log(1+u) \geq 2u/(2+u)$ gives $\log(10/7) \geq 6/17$, and $\sqrt{605} > 49/2$. Therefore the ratio in (39) is smaller than $425/168$. Moreover,

$$\frac{425}{168} < 1 + \frac{24}{25} + \frac{1}{2}\left(\frac{24}{25}\right)^2 + \frac{1}{6}\left(\frac{24}{25}\right)^3 < e^{24/25}.$$

Thus (39) is negative. □

5.6 High-weight summation

Proposition 5.8 (Unscaled high weights)**.** *For fixed $r \geq 4$,*

$$\sum_{1 \leq w \leq n:\ rw/N \geq x_0} \mathbb{P}[\mathcal{E}_{0,w}(\delta)] = e^{-\Omega_r(N)}. \tag{40}$$

Proof. Use the nonadjacency branch in (14). By lemma 5.5, every nonzero summand is at most a fixed polynomial in N times

$$\exp\left\{N\left(\frac{1}{r}\mathsf{h}(x) + \mathcal{M}(x,t) + \mathcal{M}(t,1/10)\right)\right\}.$$

The finite proof of lemma 5.5 already includes the exact -1 shifts, the complete hypergeometric tail, the floor $L = \lfloor N/10 \rfloor$, and all denominator type factors. By lemma 5.7, the exponent is at most $-xN/100 \leq -x_0N/100$. There are only polynomially many choices of w, a, h, b, so their total is $e^{-\Omega_r(N)}$. □

Theorem 5.9 (Unscaled upper bound)**.** *Under the assumptions of theorem 1.2,*

$$\mathbb{P}[\mathcal{E}_0(\delta)] = O_{r,\delta}(N^{2-r}).$$

Proof. Split by message weight. Weight one contributes $O(N^{1-r})$ by lemma 5.3. Weights $w \geq 2$ with $rw/N \leq x_0$ contribute $O(N^{2-r})$ by lemma 5.4. The remaining weights satisfy $rw/N \geq x_0$ and are exponentially unlikely by proposition 5.8. If equality holds at the splitting point, assigning that weight to either range only enlarges the sum. □

6 The randomly scaled ensemble

We now prove the upper bound in theorem 1.1. Since $\mathcal{E}_{\mathrm{sc}}(\delta) \subseteq \mathcal{E}_{\mathrm{sc}}(1/2)$ for $\delta \leq 1/2$, the upper-bound argument may use the threshold $\lfloor N/2 \rfloor$. Random scaling enters through the scalar branch of (15). The key is to sum cancellation sets together with their powers of Q^{-1} before estimating the random support geometry.

6.1 Weighted exact-gap transforms

For $u > 0$, define

$$\rho_{\mathrm{sc}}(u) := \frac{1}{2}\exp\bigl(2\sqrt{2u}\bigr), \qquad K(u) := \frac{\rho_{\mathrm{sc}}(u)}{1-\rho_{\mathrm{sc}}(u)} \quad \text{when } \rho_{\mathrm{sc}}(u) < 1. \tag{41}$$

Lemma 6.1 (Weighted second-accumulator tail)**.** *If $\rho_{\mathrm{sc}}(u) < 1$, then, for every $1 \le h \le N$,*

$$\sum_{b=0}^{\lfloor h/2 \rfloor} D(h,b)u^b g_{2,N,1/2}(h,b) \le \rho_{\mathrm{sc}}(u)^h. \tag{42}$$

Proof. Put $k = h - b$. By lemma 3.6, the gap probability is the probability that at least k of the h sampled labels land in a fixed set of size $\lfloor N/2 \rfloor$. If this occurs, some specified k-subset of those labels lies in the marked set. For any fixed k labels,

$$\mathbb{P}[\text{all } k \text{ lie in the marked set}] = \frac{(\lfloor N/2 \rfloor)_k}{(N)_k} \le 2^{-k}.$$

A union bound gives

$$g_{2,N,1/2}(h,b) \le \binom{h}{k} 2^{-k} = \binom{h}{b} 2^{-(h-b)}.$$

Using $D(h,b) = \binom{h-b}{b} \le \binom{h}{b}$ and $\binom{h}{b} \le h^b/b!$,

$$\begin{aligned}\sum_b D(h,b)u^b g_{2,N,1/2}(h,b) &\le 2^{-h}\sum_{b\ge 0}\binom{h}{b}^2 (2u)^b \\ &\le 2^{-h}\sum_{b\ge 0}\frac{(2uh^2)^b}{(b!)^2}.\end{aligned}$$

For $z \ge 0$,

$$\sum_{b\ge 0}\frac{z^b}{(b!)^2} \le \left(\sum_{b\ge 0}\frac{z^{b/2}}{b!}\right)^2 = e^{2\sqrt{z}}.$$

Taking $z = 2uh^2$ proves (42). □

Proposition 6.2 (Scaled exponential-moment bound)**.** *Let Q_0 be a real number with $1 \le Q_0 \le Q$, put $u = Q_0^{-1}$, and assume $\rho_{\mathrm{sc}}(u) < 1$. For $m = rw$ and $x = m/N$,*

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc},w}(1/2)] \le \binom{n}{w} Q_0^{w-1}\bigl(K(u)x\bigr)^m \exp\left(\frac{Nu}{K(u)}\right). \tag{43}$$

Proof. Write $k = a + b$ and $s = (k - (w-1))_+$. Since $Q_0 \le Q$,

$$Q^{-s} \le Q_0^{-s} \le Q_0^{w-1-a-b};$$

the second inequality also holds when $a + b \le w - 1$, because its right-hand side is then at least one. Use the scalar branch in (15), factor out Q_0^{w-1}, and apply lemma 6.1 with $u = Q_0^{-1}$. This gives

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc},w}(1/2)] \le \binom{n}{w} Q_0^{w-1}\sum_{a=0}^{\lfloor m/2 \rfloor} D(m,a)u^a \sum_h g_{1,N}(m,a,h)\rho_{\mathrm{sc}}(u)^h.$$

By lemma 3.7, the inner sum is at most $[K(u)x]^{m-a}$. Finally, $D(m,a) = \binom{m-a}{a} \le m^a/a!$, and therefore

$$\sum_a D(m,a)u^a[K(u)x]^{m-a} \le [K(u)x]^m \sum_{a\ge 0} \frac{1}{a!}\left(\frac{mu}{K(u)x}\right)^a$$
$$= [K(u)x]^m \exp\left(\frac{Nu}{K(u)}\right),$$

because $x = m/N$. □

6.2 All sparse message weights

Define

$$W_N := \left\lfloor \frac{N^{(r-2)/(r-1)}}{\log N} \right\rfloor. \tag{44}$$

Lemma 6.3 (Sparse scaled weights)**.** *Fix $r \ge 4$ and $\gamma > 0$, and suppose $Q \ge \gamma N$. Then*

$$\sum_{w=1}^{W_N} \mathbb{P}[\mathcal{E}_{\mathrm{sc},w}(1/2)] = O_{r,\gamma}(N^{1-r}). \tag{45}$$

Proof. For all sufficiently large N, take $Q_0 = \gamma N$ in proposition 6.2. Then

$$\rho_{\mathrm{sc}}(Q_0^{-1}) < 1, \qquad K(Q_0^{-1}) \le 2, \qquad \exp\left(\frac{N}{Q_0 K(Q_0^{-1})}\right) \le e^{1/\gamma}.$$

Using $x = rw/N$ and $\binom{n}{w} \le (e/x)^w$, we obtain

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc},w}(1/2)] \le \frac{C_\gamma}{\gamma N}\left[B_{r,\gamma}\frac{w^{r-1}}{N^{r-2}}\right]^w, \qquad B_{r,\gamma} := e\gamma 2^r r^{r-1}, \tag{46}$$

where C_γ depends only on γ. Let U_w be the right-hand side. For $w < W_N$,

$$\frac{U_{w+1}}{U_w} \le \frac{B_{r,\gamma}}{N^{r-2}}(w+1)^{r-1}\left(1+\frac{1}{w}\right)^{(r-1)w}$$
$$\le \frac{B_{r,\gamma}e^{r-1}(W_N+1)^{r-1}}{N^{r-2}} = o(1).$$

Thus the ratio is at most $1/2$ for large N, and

$$\sum_{w=1}^{W_N} U_w \le 2U_1 = O_{r,\gamma}(N^{1-r}).$$

□

For the high-weight range, we need a transform restricted to genuine cancellation surplus.

Lemma 6.4 (Surplus generating transform)**.** *Fix w, put $m = rw$ and $x = m/N$, and retain only master-bound terms with $a + b \ge w$. Let $1 \le Q_0 \le Q$, let $z \ge 1$, put $u = z/Q_0$, and assume $\rho_{\mathrm{sc}}(u) < 1$. Their total contribution is at most*

$$\binom{n}{w} Q_0^{w-1} z^{-w} \big(K(u)x\big)^m \exp\left(\frac{Nu}{K(u)}\right). \tag{47}$$

Proof. For $k = a + b \geq w$,

$$Q^{-(k-w+1)} \leq Q_0^{w-1-k} z^{k-w} = Q_0^{w-1} z^{-w} (z/Q_0)^k.$$

Insert this estimate into the scalar branch of (15) and repeat the proof of proposition 6.2, now with $u = z/Q_0$. □

6.3 The small-surplus entropy exponent

For a master-bound term, normalize

$$x := \frac{rw}{N}, \qquad \alpha := \frac{a}{N}, \qquad t := \frac{h}{N}, \qquad \beta := \frac{b}{N}.$$

Define

$$\begin{aligned}\Phi_r(x, \alpha, t, \beta) := \frac{1}{r}\mathsf{h}(x) + (x+t)\mathsf{h}\left(\frac{x}{r(x+t)}\right) \\ + t\mathsf{h}\left(\frac{x-\alpha}{t}\right) + (1-t)\mathsf{h}\left(\frac{\alpha}{1-t}\right) - \mathsf{h}(x) \\ + \frac{1}{2}\mathsf{h}\big(2(t-\beta)\big) + \frac{1}{2}\mathsf{h}(2\beta) - \mathsf{h}(t).\end{aligned} \tag{48}$$

Lemma 6.5 (Arrangement-branch finite type bound)**.** *There is an absolute constant C such that every nonzero arrangement-branch summand in (15), including the factor $\binom{n}{w}$, is at most*

$$(N+1)^C \exp\{N\Phi_r(x, \alpha, t, \beta)\}. \tag{49}$$

Proof. The message-support factor contributes at most $\exp\{(N/r)\mathsf{h}(x)\}$. Since $m + h \geq rw \geq 4w$, the binomial coefficients in $J(m, h, w)$ are increasing through w, and

$$J(m, h, w) \leq (w+1)\binom{m+h}{w}.$$

This contributes a polynomial factor and the exponent

$$(x+t)\mathsf{h}\left(\frac{x}{r(x+t)}\right).$$

For the first exact-gap kernel,

$$\binom{h-1}{m-a-1} \leq \binom{h}{m-a},$$

while (22) lower-bounds $\binom{N}{m}$. Thus

$$g_{1,N}(m, a, h) \leq (N+1)\exp\left\{N\left[t\mathsf{h}\left(\frac{x-\alpha}{t}\right) + (1-t)\mathsf{h}\left(\frac{\alpha}{1-t}\right) - \mathsf{h}(x)\right]\right\}.$$

This also handles the -1 shifts uniformly.

For the second transition, put $L = \lfloor N/2 \rfloor$ and $j = h - b$. As in (5.4), the hypergeometric point masses are nonincreasing for $k \geq h/2$; hence

$$g_{2,N,1/2}(h, b) \leq (h+1)\frac{\binom{L}{j}\binom{N-L}{b}}{\binom{N}{h}}.$$

By (23),

$$\binom{L}{j} \le \exp\left\{\frac{N}{2}\mathsf{h}(2j/N)\right\}.$$

Let $M = \lfloor N/2 \rfloor$. If N is odd, then $N - L = M + 1$, and $b \le h/2 \le M$. Therefore

$$\binom{M+1}{b} = \frac{M+1}{M+1-b}\binom{M}{b} \le (N+1)\binom{M}{b} \le (N+1)\exp\left\{\frac{N}{2}\mathsf{h}(2b/N)\right\}.$$

The same bound without the first factor holds when N is even. Finally, (22) lower-bounds $\binom{N}{h}$. Thus the second transition contributes, up to a polynomial factor,

$$\frac{1}{2}\mathsf{h}(2(t-\beta)) + \frac{1}{2}\mathsf{h}(2\beta) - \mathsf{h}(t).$$

Multiplication proves (49). In particular, no stochastic comparison with a nonexistent half-integer marked set is used. □

Lemma 6.6 (A binary fourth-moment information bound)**.** *Let B, C be binary random variables, put*

$$\theta := \mathbb{P}[B = 1], \qquad \Delta := \mathbb{P}[B = 1, C = 1] - \mathbb{P}[B = 1]\mathbb{P}[C = 1],$$

and assume $0 < \theta < 1$. Then

$$I(B;C) \ge \frac{2\Delta^2}{\theta(1-\theta)} + \frac{4\Delta^4}{3\theta^2(1-\theta)^2}. \tag{50}$$

Here mutual information is measured in nats.

Proof. For $|y| \le 1$,

$$\mathsf{h}\left(\frac{1-y}{2}\right) = \log 2 - \sum_{k\ge 1} \frac{y^{2k}}{2k(2k-1)}, \tag{51}$$

with the endpoint values understood by continuity. Let

$$Y := 1 - 2\mathbb{P}[C = 1 \mid B],$$

viewed as a two-valued random variable under the law of B. Subtracting conditional entropy from marginal entropy in (51) gives

$$I(B;C) = \sum_{k\ge 1} \frac{\mathbb{E}[Y^{2k}] - (\mathbb{E}Y)^{2k}}{2k(2k-1)}.$$

If $p_i = \mathbb{P}[C = 1 \mid B = i]$, then

$$\Delta = \theta(1-\theta)(p_1 - p_0), \qquad \mathrm{Var}(Y) = 4\theta(1-\theta)(p_1-p_0)^2 = \frac{4\Delta^2}{\theta(1-\theta)}.$$

The $k = 1$ term is one half of this variance. Moreover,

$$\mathbb{E}[Y^4] \ge (\mathbb{E}[Y^2])^2 = \big((\mathbb{E}Y)^2 + \mathrm{Var}(Y)\big)^2,$$

so

$$\mathbb{E}[Y^4] - (\mathbb{E}Y)^4 \ge \mathrm{Var}(Y)^2.$$

Keeping only the $k = 1, 2$ terms proves (50). □

Table 3: Rational logarithm bounds used in the analytic scaled-entropy proof. The column M specifies the truncation in (52).

y	M	Consequence
2	4	$\log 2 > 34657/50000$
2	2	$\log 2 < 1733/2500$
3	4	$\log 3 < 54931/50000$
3	6	$\log 3 > 5493/5000$
5/4	2	$\log(5/4) > 11157/50000$
83/80	0	$\log(83/80) < 1841/50000$
83/73	0	$\log(83/73) < 12839/100000$
35/32	0	$\log(35/32) < 4481/50000$
53/40	1	$\log(53/40) < 14071/50000$
4672/2905	1, 0	$47/100 < \log(4672/2905) < 239/500$
1553/1200	0	$\log(1553/1200) < 2579/10000$
1553/1403	0	$\log(1553/1403) < 127/1250$
5/3	1	$\log(5/3) < 511/1000$
1403/1200	0	$\log(1403/1200) < 157/1000$
3/2	1	$\log(3/2) < 811/2000$
451/400	0	$\log(451/400) < 1201/10000$
501/451	0	$\log(501/451) < 1053/10000$
451/401	0	$\log(451/401) < 59/500$

The analytic proof below uses only the following elementary logarithm bounds. They are collected in one place to keep the later constant checks transparent.

Lemma 6.7 (Rational logarithm bounds)**.** *For $y > 1$, put $\vartheta = (y-1)/(y+1)$. For every integer $M \geq 0$,*

$$L_M(y) := 2\sum_{k=0}^{M} \frac{\vartheta^{2k+1}}{2k+1},$$
$$U_M(y) := L_M(y) + \frac{2\vartheta^{2M+3}}{(2M+3)(1-\vartheta^2)}. \tag{52}$$

Then

$$L_M(y) \leq \log y \leq U_M(y).$$

In particular, the rational inequalities in Table 3 hold.

Proof. The identity

$$\log y = 2\sum_{k\geq 0} \frac{\vartheta^{2k+1}}{2k+1}$$

is the power series for $2\operatorname{artanh}\vartheta$. Its terms are positive, giving the lower bound. In the remainder, replace every denominator by $2M+3$ and sum the resulting geometric series, giving the upper bound. Every entry in Table 3 follows by substituting the displayed rational y and M into (52); all comparisons are between integers after clearing denominators. □

Lemma 6.8 (Analytic scaled entropy margin)**.** *For every $r \geq 4$ and every feasible quadruple satisfying*

$$0 < x \leq 1, \quad 0 \leq \alpha \leq x/2, \quad x - \alpha \leq t \leq 1 - \alpha,$$
$$0 \leq \beta \leq t/2, \quad t - \beta \leq 1/2, \quad \alpha + \beta \leq \frac{51x}{50r}, \tag{53}$$

one has

$$\Phi_r(x, \alpha, t, \beta) \leq -\frac{x}{100}. \tag{54}$$

Proof. At fixed (x, α, t, β), the only positive terms in (48) that depend on r are

$$\frac{1}{r}\mathsf{h}(x), \qquad (x+t)\mathsf{h}\left(\frac{x}{r(x+t)}\right).$$

Both are nonincreasing for $r \geq 4$, and the cancellation-budget domain becomes smaller as r grows. It is therefore enough to prove the result for $r = 4$. Put

$$c := \frac{51}{200}, \qquad \bar{c} := 1 - c = \frac{149}{200}. \tag{55}$$

Introduce binary random variables B_0, B_1, B_2 through the two joint probability tables

	$B_1 = 1$	$B_1 = 0$
$B_0 = 1$	$x - \alpha$	α
$B_0 = 0$	$t - x + \alpha$	$1 - t - \alpha$

	$B_2 = 1$	$B_2 = 0$
$B_1 = 1$	$t - \beta$	β
$B_1 = 0$	$1/2 - t + \beta$	$1/2 - \beta$.

All entries are nonnegative by feasibility. The marginals of B_0, B_1, B_2 are $x, t, 1/2$, respectively. Direct expansion of mutual information gives

$$\Phi_4(x, \alpha, t, \beta) = \mathsf{A}(x, t) - I(B_0; B_1) - I(B_1; B_2), \tag{56}$$

where

$$\mathsf{A}(x, t) := \frac{1}{4}\mathsf{h}(x) + (x+t)\mathsf{h}\left(\frac{x}{4(x+t)}\right). \tag{57}$$

We bound the two information losses differently in two overlapping density regimes.

Range I: $0 < x \leq 3/8$**.** Put

$$u := \frac{t}{x}, \qquad e := \frac{\alpha}{x}, \qquad f := \frac{\beta}{t}.$$

Here $t \geq x - \alpha \geq x/2 > 0$. The cancellation budget gives

$$e + uf = \frac{\alpha + \beta}{x} \leq c. \tag{58}$$

Moreover,

$$\bar{c} \leq u \leq c + \frac{1}{2x}, \tag{59}$$

because $t \geq x - \alpha \geq \bar{c}x$ and $t \leq 1/2 + \beta \leq 1/2 + cx$.

Let

$$q_1 := 2\big(x(1-t) - \alpha\big), \qquad q_2 := t - 2\beta.$$

Decompose $I(B_0;B_1)$ according to B_0. The contribution from $B_0=1$ satisfies

$$D\big((1-e,e)\|(t,1-t)\big)\ge (1-e)\log(1/t)-\mathsf{h}(e),$$

while Pinsker's inequality applied to the $B_0=0$ contribution gives

$$I(B_0;B_1)\ge x\big[(1-e)\log(1/t)-\mathsf{h}(e)\big]+\frac{q_1^2}{2(1-x)}. \tag{6.8.1}$$

The analogous decomposition according to B_1, using that B_2 is balanced, gives

$$I(B_1;B_2)\ge t\big[\log 2-\mathsf{h}(f)\big]+\frac{q_2^2}{2(1-t)}. \tag{6.8.2}$$

For $0<z\le 1$, binary entropy conjugacy and (58) imply

$$\begin{aligned}&\mathsf{h}(e)+e\log\frac{1}{xu}+u\mathsf{h}(f)\\&\quad\le\log\Big(1+\frac{z}{xu}\Big)+u\log(1+z)-c\log z.\end{aligned}$$

Indeed, apply $\mathsf{h}(y)+ys\le\log(1+e^s)$ first with $s=\log(z/(xu))$ and then with $s=\log z$; since $\log z\le 0$, replacing $e+uf$ by its upper bound c preserves the inequality. Taking $z=x/4$ in (6.8.1) $-$ $-$(6.8.2) gives, after division by x, the information lower bound

$$\begin{aligned}\frac{I(B_0;B_1)+I(B_1;B_2)}{x}&\ge u\log 2+c\log(x/4)-u\log(1+x/4)\\&\quad-\log\big(xu+x/4\big)+\mathcal{Q}(x,u),\end{aligned} \tag{60}$$

where Cauchy–Schwarz and $q_1+q_2\ge x[2\bar c+u(1-2x)]$ give

$$\mathcal{Q}(x,u):=\frac{x[2\bar c+u(1-2x)]^2}{2[2-x(1+u)]}. \tag{61}$$

The denominator is positive throughout the range under consideration. Combining (56) and (60), we obtain

$$\frac{\Phi_4(x,\alpha,t,\beta)}{x}+\frac{1}{100}\le\mathcal{R}(x,u), \tag{62}$$

where

$$\begin{aligned}\mathcal{R}(x,u):=\frac{\mathsf{h}(x)}{4x}+(1+u)\mathsf{h}\Big(\frac{1}{4(1+u)}\Big)+\frac{1}{100}-u\log 2\\+\bar c\log x+c\log 4+u\log(1+x/4)+\log(u+1/4)-\mathcal{Q}(x,u).\end{aligned} \tag{63}$$

We now maximize this explicit two-variable function analytically. Differentiation gives

$$\partial_x\mathcal{R}=\frac{\log(1-x)}{4x^2}+\frac{\bar c}{x}+\frac{u}{4+x}-\partial_x\mathcal{Q}, \tag{64}$$

$$\partial_u\mathcal{R}=\log\frac{4(u+1)}{4u+3}-\log 2+\log(1+x/4)+\frac{4}{4u+1}-\partial_u\mathcal{Q}, \tag{65}$$

$$\partial_u^2\mathcal{R}=-\frac{1}{(u+1)(4u+3)}-\frac{16}{(4u+1)^2}-\partial_u^2\mathcal{Q}. \tag{66}$$

A direct calculation gives

$$\partial_u^2 \mathcal{Q} = \frac{x(200x^2 - 351x + 200)^2}{10000[2 - x(1+u)]^3} \geq 0. \tag{67}$$

Thus $u \mapsto \mathcal{R}(x,u)$ is strictly concave.

We first show that its maximum occurs with $u \leq 3/2$. At $u = 3/2$,

$$\partial_u \mathcal{Q}(x, 3/2) = \frac{x(299 - 300x)(700x^2 - 1001x + 400)}{5000(4 - 5x)^2} \geq x. \tag{68}$$

For the last inequality, after subtracting x and clearing the positive denominator it is enough to check

$$P_1(x) := -210000x^3 + 384600x^2 - 219299x + 39600 > 0.$$

Now $P_1''' = -1260000 < 0$,

$$P_1''(3/8) = 296700 > 0, \quad P_1'(3/8) = -77771/4 < 0, \quad P_1(3/8) = 11937/32 > 0.$$

It follows successively that $P_1'' > 0$, $P_1' < 0$, and $P_1 > 0$ on $[0, 3/8]$. Substituting (68) into (65) yields

$$\partial_u \mathcal{R}(x, 3/2) \leq \log \frac{5(4+x)}{36} + \frac{4}{7} - x < 0.$$

The right side decreases in x, and its value at zero is negative because $\log(9/5) > 4/7$, the latter following from $\log y \geq 2(y-1)/(y+1)$. Concavity therefore implies that $\mathcal{R}(x,u)$ decreases for $u \geq 3/2$.

Next we show that $\mathcal{R}$ is increasing in x for $\bar{c} \leq u \leq 3/2$. The identity

$$\partial_x^3 \mathcal{Q} = \frac{3(100u^2 - 151u + 149)^2}{5000[2 - x(1+u)]^4} > 0 \tag{69}$$

shows that $x \mapsto \partial_x \mathcal{Q}(x,u)$ is convex. Its endpoint values are

$$\begin{aligned} \partial_x \mathcal{Q}(0,u) &= \frac{(100u + 149)^2}{40000}, \\ \partial_x \mathcal{Q}(3/8,u) &= \frac{(25u + 149)(225u^2 - 775u + 1192)}{1250(13 - 3u)^2}. \end{aligned}$$

Both are increasing in $u \in [\bar{c}, 3/2]$. For the second one,

$$\frac{\mathrm{d}}{\mathrm{d}u} \partial_x \mathcal{Q}(3/8, u) = \frac{-16875u^3 + 219375u^2 + 110875u - 48127}{1250(13 - 3u)^3} > 0.$$

Indeed, the negative of its numerator has derivative $125(405u^2 - 3510u - 887) < 0$ on this interval and is already negative at $u = \bar{c}$. Convexity and monotonicity of the endpoint values now give

$$\partial_x \mathcal{Q}(x,u) \leq \left(1 - \frac{8x}{3}\right) \frac{89401}{40000} + \frac{8x}{3} \frac{799339}{722500} =: B(x). \tag{70}$$

Using

$$\log(1-x) \geq -x - \frac{x^2}{2(1-x)}$$

in (64), then using $u \geq \bar{c}$ and (70), gives

$$\partial_x \mathcal{R}(x,u) \geq \frac{P_2(x)}{11560000x(1-x)(x+4)}, \tag{71}$$

where

$$\begin{aligned} P_2(x) := & -34793240x^4 - 78542831x^3 + 200904227x^2 \\ & - 117681956x + 22888800. \end{aligned}$$

Here

$$P_2'''(x) = -835037760x - 471256986 < 0,$$

and

$$P_2''(3/8) = 665493967/4 > 0, \quad P_2'(3/8) = -478607465/64 < 0, \quad P_2(3/8) = 69768525/32 > 0.$$

The same monotonicity argument proves $P_2 > 0$ on $[0, 3/8]$. Hence $\partial_x \mathcal{R} > 0$ on the required rectangle.

It remains to bound the one-variable section $\mathcal{R}(3/8, u)$. From (66) and (67), for $u \leq 3/2$,

$$\partial_u^2 \mathcal{R}(3/8, u) \leq -\frac{16}{49}. \tag{72}$$

At $u_0 = 43/40$, direct simplification gives

$$\begin{aligned} \mathcal{R}(3/8, u_0) = & -\frac{53}{60}\log 2 + \frac{99}{200}\log 3 - \frac{5}{12}\log(5/4) + \frac{1}{4}\log(83/80) \\ & + \frac{73}{40}\log(83/73) + \frac{43}{40}\log(35/32) + \log(53/40) \\ & + \frac{1}{100} - \frac{5938947}{12512000}, \\ \partial_u \mathcal{R}(3/8, u_0) = & \frac{3073181597}{6482154400} - \log(4672/2905). \end{aligned}$$

The bounds in Table 3 yield

$$\mathcal{R}(3/8, u_0) < -\frac{56463563}{11730000000} < -\frac{1}{250}, \qquad |\partial_u \mathcal{R}(3/8, u_0)| < \frac{1}{200}. \tag{73}$$

Strong concavity therefore gives, for every $u \in [\bar{c}, 3/2]$,

$$\mathcal{R}(3/8, u) \leq \mathcal{R}(3/8, u_0) + \partial_u \mathcal{R}(3/8, u_0)(u - u_0) - \frac{8}{49}(u - u_0)^2 < -\frac{1}{250} + \frac{49}{32 \cdot 200^2} < 0.$$

Together with the preceding monotonicity in u and x, this proves $\mathcal{R}(x, u) < 0$ throughout Range I. The desired margin follows from (62).

Range II: $3/8 \leq x \leq 100/149$**.** The upper endpoint follows already from feasibility:

$$x - \alpha \leq t \leq \frac{1}{2} + \beta \quad \Longrightarrow \quad x \leq \frac{1}{2} + \alpha + \beta \leq \frac{1}{2} + cx.$$

For the two binary pairs in (56), put

$$q_1 := 2\big(x(1-t) - \alpha\big), \qquad q_2 := t - 2\beta.$$

They are positive. Indeed,

$$x(1-t)-\alpha \ge x(1/2-\beta)-\alpha \ge (1/2-c)x, \qquad t-2\beta \ge x-2(\alpha+\beta) \ge (1-2c)x.$$

By lemma 6.6, with $X = x(1-x)$ and $T = t(1-t)$,

$$I(B_0;B_1)+I(B_1;B_2) \ge \frac{1}{2}\left(\frac{q_1^2}{X}+\frac{q_2^2}{T}\right) + \frac{1}{12}\left(\frac{q_1^4}{X^2}+\frac{q_2^4}{T^2}\right).$$

Cauchy–Schwarz and

$$q_1+q_2 \ge 2\bar{c}x+(1-2x)t$$

give

$$I(B_0;B_1)+I(B_1;B_2) \ge \frac{\mathsf{Z}(x,t)}{2}+\frac{\mathsf{Z}(x,t)^2}{24}, \tag{74}$$

where

$$\mathsf{Z}(x,t) := \frac{[2\bar{c}x+(1-2x)t]^2}{x(1-x)+t(1-t)}. \tag{75}$$

Define

$$\mathsf{Z}_*(x) := \frac{x(40000x^3-80000x^2+42401x+4900)}{2500(1+4x-4x^2)}. \tag{76}$$

The exact identity

$$\begin{aligned}&[2\bar{c}x+(1-2x)t]^2-\mathsf{Z}_*(x)\big(x(1-x)+t(1-t)\big)\\ &\quad=\frac{(49x+50)^2}{2500(1+4x-4x^2)}\left(t+\frac{x(400x^2-600x+51)}{2(49x+50)}\right)^2\end{aligned} \tag{77}$$

proves $\mathsf{Z}(x,t) \ge \mathsf{Z}_*(x)$.

The function $\mathsf{A}(x,t)$ in (57) is increasing in t. Since $t \le 1/2+\beta \le 1/2+cx$, put

$$\mathsf{A}_+(x) := \frac{1}{4}\mathsf{h}(x)+\left(\frac{1}{2}+(1+c)x\right)\mathsf{h}\left(\frac{x}{4(\frac{1}{2}+(1+c)x)}\right). \tag{78}$$

Combining (56), (74), and (77), it is enough to prove

$$\mathsf{G}(x) := \mathsf{A}_+(x)+\frac{x}{100}-\frac{\mathsf{Z}_*(x)}{2}-\frac{\mathsf{Z}_*(x)^2}{24} < 0. \tag{79}$$

At $x = 3/8$,

$$\begin{aligned}\mathsf{A}_+(3/8) = \frac{23}{32}\log 2-\frac{3}{32}\log 3-\frac{5}{32}\log(5/4)\\ +\frac{3}{32}\log(1553/1200)+\frac{1403}{1600}\log(1553/1403),\end{aligned}$$

and $\mathsf{Z}_*(3/8) = 139917/155000$. Hence Table 3 gives

$$\mathsf{G}(3/8) < -\frac{1518473613}{192200000000} < -\frac{7}{1000}. \tag{80}$$

It remains to prove that G decreases. The function A_+ is concave: the first term is concave, and the second is an affine restriction of the perspective of binary entropy. Its derivative at three rational points is

$$\begin{aligned}
\mathsf{A}'_+(3/8) &= \frac{1}{4}\log(5/3) + \frac{1}{4}\log(1403/1200) + \frac{3}{4}\log 2 + \frac{251}{200}\log(1553/1403),\\
\mathsf{A}'_+(2/5) &= \frac{1}{4}\log(3/2) + \frac{1}{4}\log(451/400) + \frac{3}{4}\log 2 + \frac{251}{200}\log(501/451),\\
\mathsf{A}'_+(1/2) &= \frac{1}{4}\log(401/400) + \frac{3}{4}\log 2 + \frac{251}{200}\log(451/401).
\end{aligned}$$

Using $\log(401/400) < 1/400$ and Table 3,

$$\mathsf{A}'_+(3/8) < \frac{41}{50}, \qquad \mathsf{A}'_+(2/5) < \frac{157}{200}, \qquad \mathsf{A}'_+(1/2) < \frac{67}{100}. \tag{81}$$

Differentiating (76) gives

$$\mathsf{Z}'_*(x) = -\frac{160000x^5 - 400000x^4 + 240000x^3 + 25398x^2 - 42401x - 2450}{1250(1 + 4x - 4x^2)^2}. \tag{82}$$

We claim

$$\begin{aligned}
&\mathsf{Z}'_*(x) \geq 89/50, \quad 3/8 \leq x \leq 2/5,\\
&\mathsf{Z}'_*(x) \geq 73/50, \quad 2/5 \leq x \leq 1/2,\\
&\mathsf{Z}'_*(x) \geq 6/5, \qquad 1/2 \leq x \leq 100/149.
\end{aligned} \tag{83}$$

For the first interval, after multiplying $\mathsf{Z}'_* - 89/50$ by the positive denominator in (82) and putting $y = 2/5 - x$, the numerator is

$$160000y^5 + 44400y^4 - 158240y^3 + 1666y^2 + 18175y + \frac{1019}{25}.$$

Here $0 \leq y \leq 1/40$, and $-158240y^2 + 1666y + 18175 > 0$, so the expression is positive. For the second interval, putting $y = 1/2 - x \in [0, 1/10]$ gives

$$160000y^5 - 29200y^4 - 160000y^3 + 43802y^2 + 12997y + 1.$$

The two negative terms are at least $-162920y^3$, which is at least $-16292y^2$; the expression is therefore positive. For the third interval, put $y = x - 1/2$, so $0 \leq y \leq 51/298 < 1/5$. The numerator for $\mathsf{Z}'_* - 6/5$ is

$$y^3(160000 - 24000y - 160000y^2) + 38602y^2 - 12997y + 1301.$$

The cubic factor is positive, while the quadratic is positive because its discriminant is

$$12997^2 - 4(38602)(1301) = -31962799 < 0.$$

This proves (83). In particular, Z_* is increasing, and

$$\mathsf{Z}_*(3/8) > \frac{9}{10}, \qquad \mathsf{Z}_*(2/5) > \frac{47}{50}, \qquad \mathsf{Z}_*(1/2) > \frac{111}{100}. \tag{84}$$

Finally,

$$\mathsf{G}'(x) = \mathsf{A}'_+(x) + \frac{1}{100} - \left(\frac{1}{2} + \frac{\mathsf{Z}_*(x)}{12}\right)\mathsf{Z}'_*(x).$$

By concavity of A_+, together with (81)– (84),

$$\begin{aligned}
\mathsf{G}'(x) &\le \frac{41}{50} + \frac{1}{100} - \left(\frac{1}{2} + \frac{9}{120}\right)\frac{89}{50} < 0, && 3/8 \le x \le 2/5,\\
\mathsf{G}'(x) &\le \frac{157}{200} + \frac{1}{100} - \left(\frac{1}{2} + \frac{47}{600}\right)\frac{73}{50} < 0, && 2/5 \le x \le 1/2,\\
\mathsf{G}'(x) &\le \frac{67}{100} + \frac{1}{100} - \left(\frac{1}{2} + \frac{111}{1200}\right)\frac{6}{5} < 0, && 1/2 \le x \le 100/149.
\end{aligned}$$

Thus $\mathsf{G}(x) \le \mathsf{G}(3/8) < 0$ throughout Range II. Together with Range I, this proves (54) for $r = 4$, and the initial monotonicity reduction proves it for every $r \ge 4$. □

6.4 High message weights

For a master-bound term, define its surplus

$$\sigma := \big(a + b - (w-1)\big)_+.$$

Call the term *small-surplus* when $\sigma \le w/50$ and *large-surplus* otherwise.

Lemma 6.9 (Small-surplus high weights)**.** *For fixed $r \ge 4$,*

$$\sum_{w > W_N} \big(\textit{small-surplus contribution at weight } w\big) = \exp\big(-\Omega_r(W_N)\big).$$

Proof. If $\sigma \le w/50$, then

$$a + b \le \frac{51}{50} w,$$

including the case of zero surplus. Hence $\alpha + \beta \le 51x/(50r)$. Use the arrangement branch of (15). By lemmas 6.5 and 6.8, every nonzero summand is at most

$$(N+1)^C e^{-xN/100} = (N+1)^C e^{-rw/100}.$$

There are only polynomially many choices of (a, h, b) and at most N message weights. Since $W_N/\log N \to \infty$, the sum is $\exp(-\Omega_r(W_N))$. □

Lemma 6.10 (Large surplus at small density)**.** *Fix $r \ge 4$ and $\gamma > 0$, assume $Q \ge \gamma N$, and let $w > W_N$ satisfy $x = rw/N \le 1/50$. The total contribution of all positive-surplus terms at weight w is at most*

$$\frac{1}{\gamma N}(0.08)^w$$

for all sufficiently large N.

Proof. Take $Q_0 = \gamma N$ and $z = Q_0 w/N$. Since $w > W_N \to \infty$, $z \ge 1$ for large N. In lemma 6.4,

$$u = \frac{z}{Q_0} = \frac{w}{N} = \frac{x}{r} \le \frac{1}{200}.$$

Consequently,

$$\rho_{\mathrm{sc}}(u) \le \frac{1}{2} e^{1/5}, \qquad K(u) \le \frac{8}{5}.$$

Using $\binom{n}{w} \le (e/x)^w$ and $Nu = w$, (47) becomes

$$\frac{1}{Q_0}\left[er\, e^{1/K(u)} K(u)^r x^{r-2}\right]^w.$$

Because $K(u) > 1$, the bracket is at most

$$e^2 r \left(\frac{8}{5}\right)^r 50^{-(r-2)}.$$

At $r = 4$ this is less than 0.08. Its ratio at $r+1$ versus r is at most

$$\frac{r+1}{r}\frac{8/5}{50} \le 0.04,$$

so $r = 4$ is extremal. □

Lemma 6.11 (Large surplus at linear density)**.** *Fix $r \ge 4$ and $\gamma > 0$, and assume $Q \ge \gamma N$. The total contribution of terms satisfying*

$$w > W_N, \qquad \frac{rw}{N} > \frac{1}{50}, \qquad \sigma > \frac{w}{50}$$

is $\exp(-\Omega_r(N \log N))$.

Proof. Here

$$w > \frac{N}{50r}, \qquad \sigma > \frac{N}{2500r}.$$

Use the scalar branch in (15) and the crude estimates

$$\binom{n}{w} \le 2^n, \qquad D(m,a) \le 2^m, \qquad D(h,b) \le 2^h, \qquad g_{1,N}, g_{2,N,1/2} \le 1.$$

There are only polynomially many summation indices, and $n, m, h \le N$. Thus the total is at most

$$(N+1)^4 2^{3N} (\gamma N)^{-N/(2500r)} = \exp(-\Omega_r(N \log N)).$$

□

Theorem 6.12 (Scaled upper bound)**.** *Under the assumptions of theorem 1.1,*

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc}}(\delta)] = O_{r,\gamma}(N^{1-r}).$$

Proof. It is enough to bound $\mathcal{E}_{\mathrm{sc}}(1/2)$. Weights $1 \le w \le W_N$ contribute $O_{r,\gamma}(N^{1-r})$ by lemma 6.3. For $w > W_N$, split every master term by its surplus. Small-surplus terms contribute $\exp(-\Omega_r(W_N))$ by lemma 6.9. Every large-surplus term has $a + b \ge w$. Those with $rw/N \le 1/50$ are summable by lemma 6.10, while those with $rw/N > 1/50$ contribute $\exp(-\Omega_r(N \log N))$ by lemma 6.11. These cases cover every high-weight master term, and the total high-weight contribution is $o(N^{1-r})$. □

7 Matching lower bounds and the change of failure mechanism

The upper bounds are attained by two different minimal witnesses. Without scaling, a weight-two message creates deterministic cancellations. With scaling over a linear-size multiplicative group, the dominant event instead comes from a weight-one message whose repeated copies concentrate at a boundary. A third construction shows how paired cancellations return when the scalar alphabet is too small.

7.1 Unscaled weight-two mixed pairs

For a message coordinate u, let $B_u \subseteq [N]$ denote its r repeated positions before the first interleaver. For an unordered pair $e = \{u, v\}$, let $\mathcal{A}_e$ be the event that $\Pi_1(B_u) \cup \Pi_1(B_v)$ is the union of r disjoint adjacent pairs and every pair contains one point from each image set.

Lemma 7.1 (Mixed-pair lower bound)**.** *For every fixed $r \geq 4$ and $0 < \delta \leq 1/2$,*

$$\mathbb{P}[\mathcal{E}_0(\delta)] = \Omega_{r,\delta}(N^{2-r}). \tag{85}$$

This lower bound holds over every finite field.

Proof. A path on N vertices has $\binom{N-r}{r}$ matchings of cardinality r. For each matching, its endpoints can be assigned to $\Pi_1(B_u)$ and $\Pi_1(B_v)$ in 2^r ways. Since the two image supports are a uniform ordered pair of disjoint r-subsets,

$$\mathbb{P}(\mathcal{A}_e) = \frac{2^r \binom{N-r}{r}}{\binom{N}{r}\binom{N-r}{r}} = \frac{2^r}{\binom{N}{r}}. \tag{86}$$

On $\mathcal{A}_e$, take the message $x = e_u - e_v$. Before every adjacent pair, the running sum is zero. The first endpoint changes it to either 1 or -1, and the second returns it to zero. Hence the first accumulator output is nonzero on exactly one coordinate from each pair and has weight exactly r. This remains valid in characteristic two, where $-1 = 1$.

Let $H \subseteq [N]$ be the terminal interval of length $L = \lfloor \delta N \rfloor$. Conditional on the first-stage realization, the second interleaver maps the weight-r support into H with probability

$$\frac{\binom{L}{r}}{\binom{N}{r}} \geq c_{r,\delta} > 0$$

for all sufficiently large N. On this event, the input to the final accumulator is supported inside H. The accumulator output is zero before the first active coordinate and hence supported inside H; it is nonzero because the accumulator is invertible. Define $\mathcal{F}_e$ as the intersection of these two events. Then

$$\mathbb{P}(\mathcal{F}_e) = \Theta_{r,\delta}(N^{-r}).$$

There are $\binom{n}{2} = \Theta_r(N^2)$ message pairs. For distinct pairs e, f,

$$\mathbb{P}(\mathcal{F}_e \cap \mathcal{F}_f) = O_r(N^{-2r}). \tag{87}$$

Indeed, condition on $\mathcal{A}_e$. If e and f share a message coordinate, the new block in f must choose one neighboring position for each of the r already placed copies of the shared block; there are at most 2^r candidate supports among $\Theta_r(N^r)$ possible supports. If e and f are disjoint, condition additionally on the image support of one block of f; the support of the other block again has at most 2^r neighboring choices. Thus $\mathbb{P}(\mathcal{A}_f \mid \mathcal{A}_e) = O_r(N^{-r})$. Ignoring the second-interleaver conditions only enlarges the intersection probability.

By the first two terms of inclusion–exclusion,

$$\begin{aligned}\mathbb{P}\left[\bigcup_e \mathcal{F}_e\right] &\geq \sum_e \mathbb{P}(\mathcal{F}_e) - \sum_{e<f} \mathbb{P}(\mathcal{F}_e \cap \mathcal{F}_f) \\ &= \Theta_{r,\delta}(N^{2-r}) - O_r(N^{4-2r}) = \Omega_{r,\delta}(N^{2-r}),\end{aligned}$$

because $r \geq 4$. □

Combining theorem 5.9 and lemma 7.1 proves theorem 1.2.

7.2 Scaled weight-one boundary concentration

Lemma 7.2 (Weight-one boundary lower bound)**.** *For every fixed $r \geq 2$ and $0 < \delta \leq 1/2$,*

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc}}(\delta)] = \Omega_{r,\delta}(N^{1-r}), \tag{88}$$

uniformly in the field size.

Proof. Let

$$I := \{N - r + 1, \ldots, N\}.$$

For each message coordinate u, let $\mathcal{B}_u$ be the event $\Pi_1(B_u) = I$. Then

$$\mathbb{P}(\mathcal{B}_u) = \binom{N}{r}^{-1},$$

and the events $\mathcal{B}_u$ are mutually exclusive. On $\mathcal{B}_u$, the first accumulator output produced by the weight-one message e_u is nonzero and supported inside I, regardless of the first diagonal scalars. Let its weight be h; then $1 \leq h \leq r$.

Conditional on this realization, the second interleaver maps that support into the terminal interval H of length $L = \lfloor \delta N \rfloor$ with probability

$$\frac{\binom{L}{h}}{\binom{N}{h}} \geq c_{r,\delta} > 0$$

uniformly over $1 \leq h \leq r$ for sufficiently large N. The second scaling preserves support. The final accumulator output is supported inside H, and it is nonzero because both the scaling and accumulator are invertible. Since the $\mathcal{B}_u$ are disjoint,

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc}}(\delta)] \geq \sum_{u=1}^{n} c_{r,\delta}\mathbb{P}(\mathcal{B}_u) = \frac{N}{r}\, c_{r,\delta}\binom{N}{r}^{-1} = \Omega_{r,\delta}(N^{1-r}).$$

□

Together with theorem 6.12, this proves theorem 1.1.

7.3 Paired scalar cancellations and the field threshold

Proof of theorem 1.4. Put

$$s = \lfloor r/2 \rfloor, \qquad t = \lceil r/2 \rceil, \qquad Q = q - 1.$$

Fix a message coordinate u. If $r = 2s$ is even, let $\mathcal{P}_u$ be the event that the r positions in $\Pi_1(B_u)$ form a size-s matching of adjacent pairs. If $r = 2s + 1$ is odd, require that one copy land at coordinate N and that the remaining $2s$ copies form a size-s matching in the path on $[N - 1]$. In the two cases,

$$\mathbb{P}(\mathcal{P}_u) = \frac{\binom{N-s}{s}}{\binom{N}{2s}} = \Theta_r(N^{-s}), \qquad r = 2s,$$

$$\mathbb{P}(\mathcal{P}_u) = \frac{\binom{N-s-1}{s}}{\binom{N}{2s+1}} = \Theta_r(N^{-(s+1)}), \qquad r = 2s + 1.$$

Thus $\mathbb{P}(\mathcal{P}_u) = \Theta_r(N^{-t})$ in both cases.

For every adjacent pair, require the two corresponding first-stage scalars to sum to zero. Conditional on the first scalar, exactly one of the Q possible values of the second scalar works. The s pair equations use disjoint diagonal entries and therefore hold with probability Q^{-s}. On their intersection with $\mathcal{P}_u$, the first accumulator output has weight exactly t: one nonzero position per pair and, in the odd case, the singleton at coordinate N.

The second interleaver maps this weight-t support into the terminal interval H with probability $\binom{L}{t}/\binom{N}{t} \geq c_{r,\delta} > 0$. Let $\mathcal{F}_u$ denote the complete event. Then

$$\mathbb{P}(\mathcal{F}_u) = \Theta_{r,\delta}(N^{-t}Q^{-s}). \tag{89}$$

When r is odd, the events $\mathcal{P}_u$ are mutually exclusive because each requires a copy of block u at coordinate N. Summing (89) over $n = N/r$ coordinates proves the claim.

When r is even, use Bonferroni. For $u \neq v$, conditioning on $\mathcal{P}_u$ leaves the image of B_v uniform among the remaining positions, and the probability that it also forms adjacent pairs is $O_r(N^{-s})$. The scalar equations use disjoint diagonal entries. Ignoring the two second-interleaver concentration conditions gives

$$\mathbb{P}(\mathcal{F}_u \cap \mathcal{F}_v) = O_r(N^{-2s}Q^{-2s}).$$

It follows that

$$\begin{aligned}\mathbb{P}\left[\bigcup_u \mathcal{F}_u\right] &\geq \Theta_{r,\delta}(N^{1-s}Q^{-s}) - O_r(N^{2-2s}Q^{-2s}) \\ &= \Omega_{r,\delta}(N^{1-s}Q^{-s}),\end{aligned}$$

because the ratio of the second term to the first is $O_r(N^{1-s}Q^{-s}) = o_r(1)$ for $s \geq 2$ and $Q \geq 1$. Here $t = s$, completing the proof. □

The ratio between the paired-cancellation lower bound and N^{1-r} is

$$\frac{N^{1-t}Q^{-s}}{N^{1-r}} = \left(\frac{N}{Q}\right)^s,$$

because $r = s + t$. This proves corollary 1.5. Combining the two sharp theorems on their common parameter range proves corollary 1.3.

8 What random scaling changes

The two sharp laws identify different dominant witnesses:

ensemble	minimal mechanism	tail order
G_0	deterministic mixed weight-two pairs	N^{2-r}
$G,\ Q \geq \gamma N$	weight-one boundary concentration	N^{1-r}.

The result is therefore not only that scaling improves an upper bound. It removes the witness that actually controls the unscaled ensemble. Once that witness is destroyed, weight one becomes the leading obstruction. The field lower bound gives the intermediate interpretation: over smaller multiplicative groups, adjacent copies of one symbol can again cancel, now with probability Q^{-1} per pair. Linear growth of Q is precisely the order at which this algebraic mechanism is pushed down to the geometric N^{1-r} scale.

The common proof architecture makes this comparison explicit. Both ensembles have the same adaptive trace count and exact support kernels. The sole additional branch for the scaled ensemble is

$$Q^{-(a+b-(w-1))_+},$$

which charges zero equations beyond the one shared projective-rank budget. The sharp change in exponent is thus traced to a concrete algebraic source, not to a different interleaver geometry.

9 Generality and limitations

9.1 A reusable layered template

The proof applies beyond the literal two-accumulator encoder whenever four features are present:

1. witnesses lie in a low-dimensional projective parameter space;
2. each layer exposes an adaptively selected family of homogeneous linear zero tests;
3. random permutations separate active order from a uniform unordered support; and
4. surplus equations involve fresh random coefficients.

The shared-rank argument extends directly to an arbitrary fixed number of adaptive layers. Exact random-gap kernels compose from one layer to the next, while fresh random coefficients can be charged through analogues of lemma 4.1. The exposure order used in lemma 4.2 is part of the template: a canonical algebraic label is fixed before the random support gaps on which its geometric cost depends.

This abstraction also explains why counting traces is materially different from counting witnesses. A direct union bound sees Q^{w-1} projective messages. A separate layer-by-layer trace count may spend $w-1$ dimensions at every layer. The shared-rank argument instead records only the cumulative rank inside one hyperplane $x^{\perp}$.

9.2 Why the two distance ranges differ

The scaled theorem reaches the endpoint $\delta = 1/2$. Its weighted transform sums second-stage traces with their factors u^b and retains the strictly decaying base

$$\rho_{\mathrm{sc}}(u) = \frac{1}{2} e^{2\sqrt{2u}} < 1$$

for the values $u = O(1/N)$ used in the proof. In contrast, the unscaled fixed-trace bound has base $2\sqrt{\delta(1-\delta)}$, which tends to one as $\delta \uparrow 1/2$. Our unscaled theorem is therefore stated for $\delta \le 1/10$. Extending its sharp N^{2-r} law to larger relative thresholds requires a stronger high-weight argument; no failure of the claimed exponent is known from the present analysis.

9.3 The repetition factor and field interpolation

The scaled entropy reduction uses $r = 4$ as its extremal case and proves all $r \ge 4$ by monotonicity. The corresponding functional is not controlled by the same inequality at $r = 3$, so the theorem makes no claim there. Whether the N^{1-r} law itself extends to $r = 3$, or a different witness dominates, remains open.

Below linear field growth, theorem 1.4 identifies a larger paired-cancellation lower bound. A natural full interpolation conjecture is that, over a broad range of Q,

$$\mathbb{P}[\mathcal{E}_{\mathrm{sc}}(\delta)] \asymp N^{1-r} + N^{1-\lceil r/2 \rceil} Q^{-\lfloor r/2 \rfloor}.$$

We do not claim this statement. The present paper determines the sharp constant-factor law in the linear-field regime and the asymptotically necessary field-size order for that law.

10 Conclusion

We determined sharp minimum-distance failure probabilities for both unscaled and randomly scaled RAA codes. In characteristic greater than the repetition factor, the unscaled ensemble has tail $\Theta(N^{2-r})$ for $\delta \le 1/10$, driven by deterministic mixed weight-two pairs. With independent nonzero coordinate scalings over any field satisfying $q - 1 \ge \gamma N$, the tail becomes $\Theta(N^{1-r})$ for every $\delta \le 1/2$, driven by a weight-one boundary event. A paired-cancellation lower bound shows that linear field growth has the correct asymptotic order for this sharper law.

The proof replaces witness enumeration by adaptive cancellation traces. The traces across the accumulator stages share one projective-rank budget, while random interleavers produce exact support-transition kernels. Canonical trace labels are exposed before the support gaps whose geometry they charge, and exact supports are fixed whenever they determine later linear forms. Summing algebraic cancellation penalties before estimating geometry is the step that makes a linear field sufficient.

The principal open directions are a full interpolation theorem in the field size, the $r = 3$ regime, larger distance thresholds for the unscaled ensemble, and sharp analyses for more accumulator layers.

References

[1] Aliazam Abbasfar, Dariush Divsalar, and Kung Yao. Accumulate–repeat–accumulate codes. *IEEE Transactions on Communications*, 55(4):692–702, April 2007. doi:10.1109/TCOMM.2007.894114.

[2] Pariya Akhiani and Yupeng Zhang. Distance of RAA codes over large finite fields (with applications in zkSNARKs and PCGs). Cryptology ePrint Archive, Paper 2026/524, 2026.

[3] Louay Bazzi, Mohammad Mahdian, and Daniel A. Spielman. The minimum distance of turbo-like codes. *IEEE Transactions on Information Theory*, 55(1):6–15, January 2009. doi:10.1109/TIT.2008.2008114.

[4] Dariush Divsalar, Hui Jin, and Robert J. McEliece. Coding theorems for "turbo-like" codes. In *Proceedings of the 36th Annual Allerton Conference on Communication, Control, and Computing*, pages 201–210, 1998.

[5] Venkatesan Guruswami and Widad Machmouchi. Explicit interleavers for a repeat accumulate accumulate (RAA) code construction. In *2008 IEEE International Symposium on Information Theory*, pages 1968–1972, 2008. doi:10.1109/ISIT.2008.4595333.

[6] Jörg Kliewer, Kamil Sh. Zigangirov, Christian Koller, and Daniel J. Costello, Jr. Coding theorems for repeat multiple accumulate codes. *CoRR*, abs/0810.3422, 2008.