Asymptotically Optimal Distance-Tail Bounds for Large-Field RAA Codes

Majid Khabbazian

University of Alberta, Canada

Abstract. Repeat–accumulate–accumulate (RAA) codes combine a very simple linear-time encoding procedure with strong distance behavior, making them attractive both in classical coding theory and in recent cryptographic applications such as code-based polynomial commitments, zkSNARKs, and pseudorandom correlation generators. Existing concrete analyses of RAA codes are strongest over the binary field, while large-field cryptographic applications require distance guarantees over fields whose size grows with the block length. In this regime, the usual binary-field weight-enumerator and union-bound arguments lose the large-field cancellation gains needed to obtain sharp tails.

We give a gap–covering proof of an optimal-tail distance bound for the large-field RAA ensemble $G = RP_1AP_2A$. For every fixed repetition factor $r \geq 9$ and every field size satisfying $q - 1 \geq (eN)^2$, we prove

$$\mathbb{P}[d_{\min}(G) \leq \delta N] \leq \widetilde{O}_r(N^{1-r})$$

for every fixed $0 < \delta < 1/2$, where N denotes the code length. We also prove the matching large-field lower bound $\Omega_r(N^{1-r})$, showing that the upper bound is optimal up to polylogarithmic factors.

In addition, we prove a binary-field companion lower bound. In particular, for every fixed $0 < \delta < 1/2$ and even r,

$$\mathbb{P}_{\mathbb{F}_2}[d_{\min}(G) \leq \delta N] \geq c_{r,\delta} N^{1-r/2}.$$

For even r, together with the improved binary upper bound in the literature, this identifies the binary tail up to polylogarithmic factors, while our large-field result gives the tight tail $\widetilde{\Theta}_r(N^{1-r})$. Thus the binary and large-field RAA ensembles have genuinely different low-distance tail exponents: the large-field improvement is an actual polynomial separation, not merely a separation between available proof techniques.

We also record two complementary obstructions clarifying the role of the large-field assumptions. If $q - 1 = N^\gamma$ with $0 \leq \gamma < 1$, then weight-one paired cancellations give a lower bound with polynomial exponent strictly larger than $1-r$, so a sublinear field cannot support an N^{1-r}-order tail. If the random diagonal scalings are removed, then opposite-valued weight-two messages give low-distance events of order $\Omega_{r,\delta}(N^{2-r})$, showing that the scalings are essential for the optimal large-field exponent.

1 Introduction

Error-correcting codes with fast encoding and large distance are a basic tool in both coding theory and modern cryptography. In code-based proof systems, a linear code is often used to encode a large witness or polynomial-evaluation table; the encoding time contributes directly to prover time, while the relative distance controls soundness and proof size. Similar requirements arise in pseudorandom correlation generators (PCGs), where security is tied to the hardness of finding low-weight codewords in a structured linear code. These applications motivate the search for codes that are simultaneously simple to encode, concretely efficient, and provably far from having low-weight nonzero codewords.

Repeat–accumulate (RA) and repeat–accumulate–accumulate (RAA) codes are among the simplest candidates in this direction. An RA code repeats each message symbol, randomly permutes the repeated coordinates, and applies an accumulator, i.e., a prefix-sum map. This gives a highly structured linear-time encoder, but RA codes are not asymptotically good: their minimum distance does not grow linearly with the block length in the usual random-interleaver ensemble [6,2]. RAA codes add a second permutation and a second accumulator. This additional accumulator changes the distance behavior substantially: random RAA ensembles are known to achieve linear minimum distance with high probability, and related repeat–multiple–accumulate ensembles approach strong rate–distance tradeoffs as the number of accumulators increases [2,9]. The resulting codes are attractive because the encoder remains essentially just repetition, permutation, and prefix sums. This simplicity also motivated work on explicit or partially explicit RAA interleavers [8].

The recent cryptographic interest in these codes comes from a different parameter regime. Code-based SNARKs and polynomial commitment schemes require codes with efficient encoders and concrete distance guarantees. Brakedown showed that linear-time, field-agnostic SNARKs can be built from suitable linear codes, but the concrete distance of the underlying code is a major driver of proof size [7]. Blaze later used interleaved RAA codes to obtain a very efficient multilinear polynomial commitment scheme over binary extension fields, relying on a refined binary-field distance analysis of RAA codes [5]. In parallel, expand–accumulate (EA) codes were introduced for PCGs, where sparse expansion followed by an accumulator gives efficient generation of correlated randomness [4]. EA codes were later generalized to arbitrary finite fields and used in field-agnostic SNARKs [3].

These developments leave a central large-field question. Many cryptographic systems naturally operate over large prime fields, either because the computation itself is over such a field or because algebraic proof systems use random challenges and Schwartz–Zippel-type soundness bounds. Binary RAA analyses do not directly transfer to this setting. A naive large-field union bound must range over roughly $(q-1)^w$ nonzero assignments for a message of weight w, which can overwhelm the combinatorial decay coming from the accumulator. Earlier finite-field analyses of related repeat–multiple–accumulate codes typically treated regimes where the field size is fixed or small compared with the

block length, while cryptographic applications often require q to be polynomially or even super-polynomially large in the security parameter. Recent work of Akhiani and Zhang gave the first large-field RAA analysis showing that RAA codes achieve constant relative distance over large finite fields and that large fields improve the provable distance behavior compared with the binary case [1]. The goal of the present work is to sharpen this large-field picture and identify the correct polynomial order of the low-distance tail.

We study the large-field RAA ensemble

$$G = RP_1AP_2A, \qquad P_i = \Pi_i V_i,$$

where R repeats each message coordinate r times, A is the accumulator matrix, each Π_i is a uniform random permutation, and each V_i is an independent random diagonal matrix with entries in $\mathbb{F}_q^*$. The diagonal scalings are the large-field analogue of random signs: they do not change the support geometry, but they make accidental prefix cancellations unlikely. Our main theorem proves that, for every fixed $r \geq 9$ and every $q - 1 \geq (eN)^2$,

$$\mathbb{P}[d_{\min}(G) \leq N/2] \leq \widetilde{O}_r(N^{1-r}).$$

Consequently, the same upper bound holds for every smaller fixed threshold δN with $0 < \delta < 1/2$. We also prove a matching lower bound $\Omega_r(N^{1-r})$ at the half-distance threshold. Thus the large-field tail is determined, up to polylogarithmic factors, by a simple weight-one obstruction: all r repeated copies of a message coordinate may land near the end of the first accumulator and then remain confined to the last half of the second accumulator. No stronger polynomial upper bound is possible for this ensemble at this threshold.

At a high level, the proof separates the geometry of accumulator gaps from the algebra of large-field cancellation. The geometric part counts the possible locations of active positions and low-weight accumulator outputs. The algebraic part uses the random nonzero diagonal scalings to show that every surplus zero-prefix cancellation beyond the message degrees of freedom costs a factor of order $(q-1)^{-1}$. This surplus penalty is the main reason the large-field tail is smaller than the binary one.

The rest of the upper-bound proof sums this reduction over message weights. Weight-one messages give the dominant obstruction and the matching lower bound. Sparse weights are controlled directly by the covering count and the surplus cancellation penalty. High weights are ruled out by combining a deterministic zero-budget lower bound with a deficit summation showing that the remaining surplus-cancellation configurations are negligible.

Our results also make the large-field versus binary separation precise at the level of tail exponents. Prior work had identified the improvement of RAA distance over large fields as an open conjectural phenomenon: large-field random scalars should suppress the deterministic cancellations available over $\mathbb{F}_2$. We show that this intuition is reflected in the actual low-distance tails. The large-field ensemble has tight tail $\widetilde{\Theta}_r(N^{1-r})$ at the half-distance threshold, while the

binary ensemble has unavoidable low-distance events of much larger probability. In particular, for every fixed $0 < \delta < 1/2$ and even r, we prove

$$\mathbb{P}_{\mathbb{F}_2}[d_{\min}(G) \le \delta N] \ge c_{r,\delta} N^{1-r/2}.$$

Together with the improved binary upper bound in the literature, this identifies the binary tail up to polylogarithmic factors. Thus the binary and large-field RAA ensembles have genuinely different low-distance tail exponents, not merely different known provable guarantees.

We also include two short complementary appendices that clarify which ingredients of the large-field ensemble are responsible for the exponent $1-r$. First, the field size cannot be made sublinear without changing the polynomial tail: if $q - 1 = N^\gamma$ with $0 \le \gamma < 1$, then weight-one paired cancellations already give a low-distance lower bound with exponent strictly larger than $1 - r$. Thus some essentially linear field-size growth is necessary for an N^{1-r}-order tail, even though our upper bound assumes the stronger condition $q - 1 \ge (eN)^2$. Second, the random diagonal scalings are not merely a proof device. If they are removed, opposite-valued weight-two messages create deterministic first-accumulator cancellations and yield low-distance events of order $\Omega_{r,\delta}(N^{2-r})$, polynomially larger than the $\widetilde{O}_r(N^{1-r})$ large-field tail proved here.

The rest of the paper is structured as follows. Section 2 defines the large-field RAA ensemble, records the accumulator gap identities, and collects elementary counting estimates used throughout the proof. Section 3 proves the large-field distance tail bounds: it develops the deterministic and scalar lemmas, reduces bad codewords to marked covering sums, estimates the weight-one, sparse, and high-weight regimes, and proves the matching lower bound. The conclusion summarizes the main consequences and remaining directions. Appendix A gives the binary companion lower bound. Appendix B records the sublinear-field paired-cancellation obstruction, and Appendix C shows that removing the random diagonal scalings creates a larger weight-two obstruction.

2 Setup

Fix an integer $r \ge 4$, a prime-power field size $q \ge 3$, and an even integer N divisible by r. Write

$$N = rn.$$

Throughout, log denotes the natural logarithm and $\log_2$ denotes the base-2 logarithm. We use one-based indexing:

$$[N] = \{1, \dots, N\}.$$

We use the convention that

$$\binom{u}{v} = 0 \qquad \text{whenever } v < 0 \text{ or } v > u.$$

Let

$$G = R\Pi_1 V_1 A \Pi_2 V_2 A \in \mathbb{F}_q^{n\times N}.$$

Here R is the repetition matrix, A is the accumulator matrix, Π_1, Π_2 are independent uniform permutation matrices on $[N]$, and V_1, V_2 are independent diagonal matrices whose diagonal entries are i.i.d. uniform in $\mathbb{F}_q^*$. All vectors are row vectors. The accumulator satisfies

$$(yA)_j = \sum_{i=1}^{j} y_i, \qquad 1 \le j \le N.$$

Put

$$P_i := \Pi_i V_i, \qquad i = 1, 2,$$

so that

$$G = RP_1 AP_2 A.$$

For a realization of G, define

$$d_{\min}(G) := \min_{x\in\mathbb{F}_q^n\setminus\{0\}} \mathrm{wt}(xG), \qquad \delta(G) := d_{\min}(G)/N.$$

For a set $S \subseteq [N]$, write $\mathbf{1}_S$ for its indicator row vector. For permutation matrix Π, define

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

2.1 First-stage active positions

For a vector $v \in \mathbb{F}_q^N$, we call a coordinate $j \in [N]$ *active* if $v_j \neq 0$. Thus the active positions of v are exactly $\mathrm{supp}(v)$.

Let $x \in \mathbb{F}_q^n \setminus \{0\}$, and write

$$w := \mathrm{wt}(x), \qquad T := \mathrm{supp}(x) \subseteq [n].$$

For $u \in [n]$, let

$$B_u := \mathrm{supp}(e_u R) \subseteq [N]$$

be the set of the r repeated copies of coordinate u. For this support T, put

$$B_T := \bigcup_{u\in T} B_u, \qquad |B_T| = rw.$$

The vector entering the first accumulator is

$$xR\Pi_1 V_1.$$

Since V_1 is diagonal with nonzero diagonal entries, it does not change support. Hence the active positions entering the first accumulator are

$$\mathrm{supp}(xR\Pi_1 V_1) = \mathrm{supp}(xR\Pi_1) = \Pi_1(B_T).$$

We order these positions as

$$\Pi_1(B_T) = \{p_1 < p_2 < \cdots < p_m\}, \qquad m := rw.$$

For each $\ell \in [m]$, let $\lambda_\ell \in T$ be the unique message coordinate whose repeated copy lands at p_ℓ, equivalently the unique coordinate satisfying

$$p_\ell \in \Pi_1(B_{\lambda_\ell}).$$

Set

$$\alpha_\ell := (V_1)_{p_\ell, p_\ell} \in \mathbb{F}_q^*.$$

Thus the nonzero increment entering the first accumulator at position p_ℓ is

$$\alpha_\ell x_{\lambda_\ell}.$$

Conditional on Π_1, the active scalars $\alpha_1, \ldots, \alpha_m$ are independent uniform elements of $\mathbb{F}_q^*$.

Define the first-stage prefix sums

$$\sigma_\ell(x, \alpha) := \sum_{s=1}^{\ell} \alpha_s x_{\lambda_s}, \qquad 1 \le \ell \le m.$$

With $p_{m+1} := N + 1$, the first accumulator output

$$z := xRP_1A$$

satisfies $z_j = 0$ for $j < p_1$ and

$$z_j = \sigma_\ell(x, \alpha) \qquad \text{for } p_\ell \le j < p_{\ell+1}.$$

Define the first-stage gap lengths

$$\Delta_\ell := p_{\ell+1} - p_\ell, \qquad 1 \le \ell \le m.$$

Then

$$\mathrm{wt}(xRP_1A) = \sum_{\ell=1}^{m} \Delta_\ell \, \mathbf{1}[\sigma_\ell(x, \alpha) \neq 0]. \tag{1}$$

Finally, define the first-stage zero-prefix index set

$$Z_1(x) := \{\ell \in [m] : \sigma_\ell(x, \alpha) = 0\}.$$

2.2 Second-stage active positions

Let $y \in \mathbb{F}_q^N \setminus \{0\}$, and write

$$h := \mathrm{wt}(y), \qquad S := \mathrm{supp}(y) \subseteq [N].$$

The vector entering the second accumulator is

$$y\Pi_2 V_2.$$

Since V_2 is diagonal with nonzero diagonal entries, it does not change support. Hence the active positions entering the second accumulator are

$$\operatorname{supp}(y\Pi_2 V_2) = \operatorname{supp}(y\Pi_2) = \Pi_2(S).$$

We order these positions as

$$\Pi_2(S) = \{q_1 < q_2 < \cdots < q_h\}.$$

For each $j \in [h]$, let $\mu_j \in S$ be the unique original coordinate of y whose image under Π_2 is q_j, equivalently the unique coordinate satisfying

$$q_j \in \Pi_2(\{\mu_j\}).$$

Set

$$\beta_j := (V_2)_{q_j,q_j} \in \mathbb{F}_q^*.$$

Thus the nonzero increment entering the second accumulator at position q_j is

$$\beta_j y_{\mu_j}.$$

Conditional on Π_2, the active second-stage scalars $\beta_1, \ldots, \beta_h$ are independent uniform elements of $\mathbb{F}_q^*$.

Define the second-stage prefix sums

$$\Theta_j(y, \beta) := \sum_{\ell=1}^{j} \beta_\ell y_{\mu_\ell}, \qquad 1 \le j \le h.$$

With $q_{h+1} := N + 1$, the second accumulator output

$$yP_2A$$

satisfies $(yP_2A)_i = 0$ for $i < q_1$ and

$$(yP_2A)_i = \Theta_j(y, \beta) \qquad \text{for } q_j \le i < q_{j+1}.$$

Define the second-stage gap lengths

$$\Gamma_j := q_{j+1} - q_j, \qquad 1 \le j \le h.$$

Then

$$\operatorname{wt}(yP_2A) = \sum_{j=1}^{h} \Gamma_j \, \mathbf{1}[\Theta_j(y, \beta) \neq 0]. \tag{2}$$

Define the second-stage zero-prefix index set

$$Z_2(y) := \{j \in [h] : \Theta_j(y, \beta) = 0\}.$$

When $y = xRP_1A$, we also record the corresponding zero-prefix positions in $[N]$ by writing

$$\widehat{Z}_2(x) := \{q_j : \Theta_j(xRP_1A, \beta) = 0\}.$$

Thus

$$|\widehat{Z}_2(x)| = |Z_2(xRP_1A)|.$$

2.3 Elementary counting inequalities

We repeatedly use the following elementary estimates. All binomial coefficients are interpreted with the convention stated above.

Lemma 1 (Basic binomial estimates). *Let all parameters below be nonnegative integers. The following inequalities hold.*

(i) *For* $0 \le b \le a$ *and* $0 \le d \le c$,

$$\binom{a}{b}\binom{c}{d} \le \binom{a+c}{b+d}. \tag{3}$$

(ii) *For* $0 \le K \le M \le N$,

$$\frac{\binom{M}{K}}{\binom{N}{K}} \le \left(\frac{M}{N}\right)^K. \tag{4}$$

(iii) *For* $1 \le K \le M$,

$$\binom{M}{K} \le \left(\frac{eM}{K}\right)^K. \tag{5}$$

(iv) *The sequence* $\binom{M}{j}$ *is increasing for* $0 \le j \le \lfloor M/2 \rfloor$.

(v) *If* $0 \le t \le K$ *and* $K - t \le H$, *then*

$$\binom{H}{K-t} \le \binom{H+t}{K}. \tag{6}$$

Proof. For (3), view the two factors as choosing b elements from a set of size a and d elements from a disjoint set of size c; this is one subfamily of all choices of $b+d$ elements from the union.

For (4),

$$\frac{\binom{M}{K}}{\binom{N}{K}} = \prod_{i=0}^{K-1} \frac{M-i}{N-i} \le \left(\frac{M}{N}\right)^K,$$

since $M \le N$. The estimate (5) is the standard bound obtained from $\binom{M}{K} \le M^K/K!$ and $K! \ge (K/e)^K$. The monotonicity claim follows from

$$\frac{\binom{M}{j+1}}{\binom{M}{j}} = \frac{M-j}{j+1} \ge 1 \qquad \text{for } j < M/2.$$

Finally, (6) follows by adding t dummy elements: each choice of $K-t$ elements from a set of size H, together with all t dummy elements, gives a choice of K elements from a set of size $H+t$.

3 Optimal Distance-Tail Bounds

We now prove the large-field distance-tail bounds. The upper bound is proved by reducing bad codewords to marked covering sums and then estimating those sums in three weight regimes: weight one, sparse weights, and high weights. The same section also proves the matching lower bound, which comes from a weight-one terminal-window obstruction. Together these results determine the large-field tail up to polylogarithmic factors.

3.1 Basic deterministic and scalar lemmas

Lemma 2 (Zero prefixes are nonadjacent). *Let $v \in \mathbb{F}_q^N$ be an accumulator input with active positions*

$$i_1 < i_2 < \cdots < i_m, \qquad \operatorname{supp}(v) = \{i_1, \dots, i_m\}.$$

For $1 \le \ell \le m$, define the active-order prefix sums $\sigma_\ell := \sum_{s=1}^{\ell} v_{i_s}$. Let $Z := \{\ell \in [m] : \sigma_\ell = 0\}$. Then $|Z| \le \left\lfloor \frac{m}{2} \right\rfloor$.

Proof. Since $i_1, \dots, i_m$ are active positions, each increment v_{i_ℓ} is nonzero. Hence

$$\sigma_1 = v_{i_1} \neq 0.$$

Moreover, two consecutive active-order prefixes cannot both vanish: if $\sigma_{\ell-1} = \sigma_\ell = 0$, then

$$v_{i_\ell} = \sigma_\ell - \sigma_{\ell-1} = 0,$$

contradicting the activity of i_ℓ. Thus Z is a nonadjacent subset of $\{2, \dots, m\}$, whose maximum possible size is $\lfloor m/2 \rfloor$.

Lemma 3 (Deterministic two-stage zero-prefix bound). *Let $x \in \mathbb{F}_q^n \setminus \{0\}$ have message weight w. Then*

$$\operatorname{wt}(xG) \ge rw - |Z_1(x)| - |\widehat{Z}_2(x)|. \tag{DHW}$$

Proof. Before the first accumulator, $xR\Pi_1 V_1$ has exactly rw active positions. By (1), and since every gap length is at least one,

$$\operatorname{wt}(xRP_1A) \ge rw - |Z_1(x)|.$$

Writing $y = xRP_1A$, the second-stage identity (2) gives

$$\operatorname{wt}(yP_2A) \ge \operatorname{wt}(y) - |\widehat{Z}_2(x)|.$$

Combining the two inequalities proves (DHW).

Lemma 4 (Joint scalar anti-cancellation). *Fix a message support $T \subseteq [n]$ of size w. Fix the first-stage active positions together with their labels $\lambda_1, \dots, \lambda_{rw} \in T$, and fix a set $I_1 \subseteq [rw]$. Fix also a possible first-stage support $S \subseteq [N]$ of size h, the second-stage active positions together with their labels $\mu_1, \dots, \mu_h \in S$, and a set $I_2 \subseteq [h]$. Put*

$$a := |I_1|, \qquad b := |I_2|.$$

Then the probability, over V_1, V_2, that there exists a message x with $\mathrm{supp}(x) = T$ such that

$$I_1 \subseteq Z_1(x), \qquad \mathrm{supp}(xRP_1A) = S, \qquad I_2 \subseteq Z_2(xRP_1A)$$

is at most

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

where $t_+ := \max\{t, 0\}$.

Proof. The claim is trivial when $a+b \le w-1$, so assume $a+b > w-1$. Quotient the nonzero message values on T by global nonzero scaling. There are at most $(q-1)^{w-1}$ projective assignments.

Fix one projective assignment. We first expose the active first-stage scalars in active order. Whenever an index $\ell \in I_1$ is encountered, the condition $\ell \in Z_1(x)$ has the form

$$c + \alpha_\ell x_{\lambda_\ell} = 0,$$

where c is determined by previously exposed scalars. Since $\mathrm{supp}(x) = T$, we have $x_{\lambda_\ell} \neq 0$. Hence at most one value of the fresh scalar $\alpha_\ell \in \mathbb{F}_q^*$ satisfies this constraint. Therefore the first-stage prescribed zero-prefix constraints contribute a factor at most $(q-1)^{-a}$.

Now condition on any realization of the first-stage scalars for which the first-stage constraints hold and for which

$$y := xRP_1A \qquad \text{satisfies} \qquad \mathrm{supp}(y) = S.$$

If no such realization exists, there is nothing to prove. Otherwise, for every $j \in I_2$, the condition $j \in Z_2(y)$ has the form

$$c + \beta_j y_{\mu_j} = 0,$$

where c is determined by previously exposed second-stage scalars. Since $\mu_j \in S = \mathrm{supp}(y)$, we have $y_{\mu_j} \neq 0$. Hence each prescribed second-stage zero-prefix constraint is satisfied by at most one value of the fresh scalar $\beta_j \in \mathbb{F}_q^*$. Thus the second-stage constraints contribute a factor at most $(q-1)^{-b}$.

Consequently, for the fixed projective message assignment, the probability of satisfying all prescribed zero-prefix constraints and the additional support condition $\mathrm{supp}(xRP_1A) = S$ is at most

$$(q-1)^{-(a+b)}.$$

Union bounding over the at most $(q-1)^{w-1}$ projective assignments gives

$$(q-1)^{w-1}(q-1)^{-(a+b)} = (q-1)^{-(a+b-(w-1))}.$$

Since $a + b > w - 1$, this is the claimed bound.

3.2 Marked covering reduction

For $2 \le w \le n$, $0 \le a \le \lfloor rw/2 \rfloor$, and $1 \le h \le N$, define

$$\mathcal{C}_1^\bullet(w,h,a) := \binom{h}{a}\binom{N-h+1}{a}\frac{\binom{h}{rw-a}}{\binom{N}{rw}}. \tag{C1$^\bullet$}$$

For $1 \le h \le N$ and $0 \le b \le \lfloor h/2 \rfloor$, define

$$\mathcal{C}_2^\bullet(h,b) := \binom{N/2}{b}\binom{N/2+1}{b}\frac{\binom{N/2}{h-b}}{\binom{N}{h}}. \tag{C2$^\bullet$}$$

Lemma 5 (Marked joint covering reduction). *Fix $2 \le w \le n$, exact zero-prefix counts a, b, and an intermediate weight h. For a fixed message support of size w, the joint geometric cost of having first-stage output weight h, exactly a first-stage zero-prefixes, final weight at most $N/2$, and exactly b second-stage zero-prefixes is at most*

$$\mathcal{C}_1^\bullet(w,h,a)\mathcal{C}_2^\bullet(h,b).$$

After including the scalar penalty and union bounding over message supports, the corresponding contribution is at most

$$\binom{n}{w}\mathcal{C}_1^\bullet(w,h,a)\mathcal{C}_2^\bullet(h,b)(q-1)^{-(a+b-(w-1))_+}.$$

Proof. We first record the covering count used for both accumulators. Consider an accumulator input with m active positions and output support of size H. Suppose that exactly A active-order prefixes are zero, and mark those A zero-prefix positions. The marked zero-prefix gaps split the nonzero output into at most $A+1$ nonzero pieces. Such a marked output pattern is encoded by choosing the A separations among the H nonzero coordinates and the A marked zero gaps among the $N-H+1$ zero gaps, giving at most

$$\binom{H}{A}\binom{N-H+1}{A}$$

possibilities. Once this marked output pattern is fixed, the remaining $m-A$ active positions, namely those with nonzero prefix value, must lie inside the H nonzero coordinates. This gives at most

$$\binom{H}{m-A}$$

choices. This count is only an upper bound: it forgets ordering restrictions and the nonadjacency of zero prefixes, which can only reduce the number of patterns.

For the first stage, take $m = rw$, $H = h$, and $A = a$. For a fixed message support T, the first-stage active set $\Pi_1(B_T)$ is a uniform rw-subset of $[N]$. Therefore the first-stage geometric cost is at most

$$\binom{h}{a}\binom{N-h+1}{a}\frac{\binom{h}{rw-a}}{\binom{N}{rw}} = \mathcal{C}_1^\bullet(w,h,a).$$

The bound is uniform in the labels $\lambda_1, \dots, \lambda_{rw}$, which will be fixed when the scalar lemma is applied.

Now condition on a first-stage support S of size h. In the second stage, the final output is required to have weight at most $N/2$. Enlarging its support to a set of size exactly $N/2$ can only increase the count. Applying the same covering count with $m = h$, $H = N/2$, and $A = b$, and using that $\Pi_2(S)$ is a uniform h-subset of $[N]$, gives the second-stage geometric cost

$$\binom{N/2}{b}\binom{N/2+1}{b}\frac{\binom{N/2}{h-b}}{\binom{N}{h}} = \mathcal{C}_2^{\bullet}(h,b).$$

Again the estimate is uniform in the second-stage labels $\mu_1, \dots, \mu_h$.

Finally, for each fixed labeled geometric pattern, Lemma 4 bounds the probability of realizing the prescribed first- and second-stage zero-prefix constraints by

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

There are $\binom{n}{w}$ choices for the message support. Multiplying the two geometric costs, the scalar penalty, and the number of supports gives the stated contribution.

3.3 Weight-one messages

Lemma 6 (Weight-one end-to-end bound). *Assume N is even and divisible by r. Fix $r \geq 4$, and write $n = N/r$. Assume $q-1 \geq N$. Then, for all sufficiently large N,*

$$\mathbb{P}_{P_1,P_2}\left[\exists x \in \mathbb{F}_q^n,\ \mathrm{wt}(x) = 1 : \mathrm{wt}(xG) \leq \frac{N}{2}\right] \leq O_r(N^{1-r}).$$

Proof. Fix a message coordinate $u \in [n]$. Every nonzero message supported on u is a global scalar multiple of e_u, so one representative suffices.

Let $p_1 < \cdots < p_r$ be the first-stage active positions of the r repeated copies of u, and let $a = |Z_1|$. Since the first active prefix cannot be zero, we have $0 \leq a \leq r-1$. If the first stage has output weight h and exactly a zero prefixes, the marked first-stage cover and the scalar penalty give

$$\mathbb{P}[\mathrm{wt}(e_u R P_1 A) = h,\ |Z_1| = a] \leq \binom{h}{a}\binom{N-h+1}{a}\frac{\binom{h}{r-a}}{\binom{N}{r}}(q-1)^{-a}.$$

We bound this uniformly in a. Since r is fixed,

$$\frac{1}{\binom{N}{r}} \leq C_r N^{-r}$$

for all sufficiently large N. Also,

$$\binom{h}{a} \leq h^a, \qquad \binom{h}{r-a} \leq h^{r-a},$$

and, using $q-1 \geq N$,

$$\binom{N-h+1}{a}(q-1)^{-a} \leq \left(\frac{N-h+1}{q-1}\right)^{a} \leq C_r,$$

because $0 \leq a \leq r-1$. Therefore

$$\begin{aligned}\binom{h}{a}\binom{N-h+1}{a}\frac{\binom{h}{r-a}}{\binom{N}{r}}(q-1)^{-a} &\leq C_r\, h^a h^{r-a} N^{-r} \\ &= C_r \left(\frac{h}{N}\right)^{r}.\end{aligned}$$

Summing over the $O_r(1)$ possible values of a, we obtain

$$\mathbb{P}[\mathrm{wt}(e_u R P_1 A) = h] \leq C_r \left(\frac{h}{N}\right)^{r}. \tag{7}$$

Now condition on a first-stage support of size h, and write the corresponding first-stage output as y. If the second stage has exactly b zero-prefix active positions and output weight at most $N/2$, the marked second-stage cover and the scalar penalty give

$$\mathbb{P}[\mathrm{wt}(yP_2A) \leq N/2,\ |Z_2(y)| = b \mid \mathrm{supp}(y)] \leq \binom{N/2}{b}\binom{N/2+1}{b}\frac{\binom{N/2}{h-b}}{\binom{N}{h}}(q-1)^{-b}.$$

For all sufficiently large N,

$$\binom{N/2+1}{b}(q-1)^{-b} \leq (2/3)^b.$$

Hence the conditional second-stage bad probability is at most

$$\frac{1}{\binom{N}{h}}\sum_{b}\binom{N/2}{b}\binom{N/2}{h-b}\left(\frac{2}{3}\right)^{b}.$$

The numerator is the coefficient of X^h in

$$(1+(2/3)X)^{N/2}(1+X)^{N/2}.$$

Since

$$(1+(2/3)X)(1+X) \leq (1+(5/6)X)^2$$

coefficientwise, this coefficient is at most

$$\left(\frac{5}{6}\right)^{h}\binom{N}{h}.$$

Thus the conditional second-stage bad probability is at most $(5/6)^h$.

Combining this with (7), for fixed u,

$$\mathbb{P}[\mathrm{wt}(e_u G) \le N/2] \le C_r \sum_{h \ge 1} \left(\frac{h}{N}\right)^r \left(\frac{5}{6}\right)^h \le C'_r N^{-r}.$$

Finally, union bounding over the $n = N/r$ message coordinates gives

$$\mathbb{P}_{P_1,P_2}\left[\exists x \in \mathbb{F}_q^n,\ \mathrm{wt}(x) = 1 : \mathrm{wt}(xG) \le \frac{N}{2}\right] \le O_r(N^{1-r}).$$

3.4 Sparse weights

Lemma 7 (Sparse marked joint covering bound with surplus zeros). *Assume N is even and divisible by r. Fix $r \ge 4$, and write $n = N/r$. Assume $q - 1 \ge (eN)^2$. Then, for all sufficiently large N,*

$$\begin{aligned}\sum_{w=2}^{\lfloor N/\log_2 N \rfloor} \binom{n}{w} \sum_{a=0}^{\lfloor rw/2 \rfloor} \sum_{h=1}^{N} \sum_{b=0}^{\lfloor h/2 \rfloor} (eN)^{-2s(w,a,b)} \mathcal{C}_1^\bullet(w,h,a)\mathcal{C}_2^\bullet(h,b) \\ \le \widetilde{O}_r(N^{1-r}),\end{aligned} \tag{SM}$$

where $s(w,a,b) = (a + b - (w-1))_+$. Consequently, the same bound holds with $(q-1)^{-s(w,a,b)}$ in place of $(eN)^{-2s(w,a,b)}$.

Proof. All constants may depend on r. Put $k = a + b$ and $s = (k - (w-1))_+$. By Vandermonde,

$$\binom{h}{a}\binom{N-h+1}{a} \le \binom{N+1}{2a}, \qquad \binom{N/2}{b}\binom{N/2+1}{b} \le \binom{N+1}{2b}.$$

Thus, after summing over pairs $a + b = k$,

$$\sum_{a+b=k} \binom{h}{a}\binom{N-h+1}{a}\binom{N/2}{b}\binom{N/2+1}{b} \le \binom{2N+2}{2k}.$$

For $k \le w - 1$, the sum over k is at most

$$w\binom{2N+2}{2(w-1)} \le C^w \left(\frac{CN}{w}\right)^{2(w-1)}.$$

For $k = w - 1 + s$, $s \ge 1$,

$$\begin{aligned}(eN)^{-2s}\binom{2N+2}{2k} &\le (eN)^{-2s}\left(\frac{e(N+1)}{k}\right)^{2k} \\ &\le C\left(\frac{CN}{w}\right)^{2(w-1)} (w+s)^{-2s}.\end{aligned}$$

Since $\sum_{s\ge 0}(w+s)^{-2s} = O(1)$ uniformly for $w \ge 2$,

$$\sum_{a,b}(eN)^{-2s(w,a,b)}\binom{h}{a}\binom{N-h+1}{a}\binom{N/2}{b}\binom{N/2+1}{b} \le C^w\left(\frac{CN}{w}\right)^{2(w-1)}. \tag{8}$$

For a nonzero first-stage marked ratio, $rw - a \le h$. Since $a \le rw/2$, we have $h \ge rw/2$ and hence $a \le h$. Thus, by (6),

$$\binom{h}{rw-a} \le \binom{h+a}{rw} \le \binom{2h}{rw}.$$

If $2h \le N$, then (4) gives

$$\frac{\binom{h}{rw-a}}{\binom{N}{rw}} \le \left(\frac{2h}{N}\right)^{rw}.$$

If $2h > N$, the right-hand side is at least 1, while the ratio on the left is at most 1 by binomial monotonicity, since $rw \le N/2$ in the sparse range for large N. Hence, uniformly,

$$\frac{\binom{h}{rw-a}}{\binom{N}{rw}} \le \left(\frac{2h}{N}\right)^{rw}.$$

For the second-stage marked ratio,

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

If $k > N/4$, then, because $w \le N/\log_2 N$, we have $s \ge N/5$ for all sufficiently large N; the factor $(w+s)^{-2s}$ in the preceding compression gives $\exp(-\Omega(N\log N))$, which dominates all remaining factors. Hence it remains to consider $k \le N/4$, where $b \le k \le N/4$ and

$$\left(\frac{1}{2}+\frac{b}{N}\right)^{h} \le \left(\frac{3}{4}\right)^{h}.$$

Combining the estimates gives, for fixed w,

$$\begin{aligned}\binom{n}{w}&\sum_{a,h,b}(eN)^{-2s(w,a,b)}\mathcal{C}_1^{\bullet}(w,h,a)\mathcal{C}_2^{\bullet}(h,b)\\ &\le C_r^w\binom{n}{w}\left(\frac{N}{w}\right)^{2(w-1)}\sum_{h\ge 1}\left(\frac{2h}{N}\right)^{rw}\left(\frac{3}{4}\right)^{h}+\exp(-\Omega(N\log N)).\end{aligned}$$

Since

$$\sum_{h\ge 1}h^{rw}\left(\frac{3}{4}\right)^{h} \le (C_r w)^{rw},$$

we get

$$\sum_{h\geq 1}\left(\frac{2h}{N}\right)^{rw}\left(\frac{3}{4}\right)^{h}\leq\left(\frac{C_r w}{N}\right)^{rw}.$$

Using $\binom{n}{w}\leq(eN/(rw))^w$, the fixed-$w$ contribution is at most

$$N^{-2}\left(C_r\frac{w}{N}\right)^{(r-3)w}+\exp(-\Omega(N\log N)).$$

The maximum over $2\leq w\leq N/\log_2 N$ occurs at $w=2$ up to harmless constants and is $O_r(N^{4-2r})$; summing over w gives $\widetilde{O}_r(N^{1-r})$ for $r\geq 4$. Finally, $(q-1)^{-s}\leq(eN)^{-2s}$.

3.5 High weights

Throughout this section all logarithms are natural unless explicitly marked otherwise. Set

$$\kappa_r:=\frac{1}{2}+\frac{1}{2(r-1)}=\frac{r}{2(r-1)},\qquad c_r:=-\log\kappa_r>0.$$

For $0<p<1$, define

$$C(p):=(1-p)^{-(1-p)/p}.$$

For every fixed $p>0$ and every $\varepsilon>0$, for all sufficiently large M,

$$\binom{M}{K}\leq\left((1+\varepsilon)C(p)\frac{M}{K}\right)^{K}\qquad\text{whenever } pM\leq K\leq M/2.\tag{EBC}$$

For the rest of this section fix $r\geq 9$, put

$$p_r:=\frac{1}{2r^{4/3}},$$

and choose $\varepsilon_r>0$ so small that

$$D_r:=(1+\varepsilon_r)C(p_r)<e.\tag{9}$$

This is possible because $C(p)\to e$ from below as $p\downarrow 0$, and p_r is fixed once r is fixed. Define

$$D(w):=\begin{cases}e, & w<N/r^{4/3},\\ D_r, & w\geq N/r^{4/3}.\end{cases}$$

Lemma 8 (Elementary constants for $r\geq 9$). *For every integer $r\geq 9$ the following three assertions hold.*

(i) *There exists $\eta_r^{(1)}>0$ such that, for all $u\geq r$,*

$$-\frac{4}{3}(r-3)\log r+3-\log r+r\log u-c_r(u-1)\leq-\eta_r^{(1)}.\tag{C1}$$

(ii)
$$e^3 r^3 \kappa_r^{2r-1} < 1. \tag{C2}$$

(iii) *With*
$$\Phi_r(\alpha) := \alpha \log \frac{e^3}{r\alpha^3} + r\alpha \log(r\alpha) + \left(\frac{1}{2} + \alpha\right) \log\left(\frac{1}{2} + \alpha\right), \qquad \alpha_* := \frac{1}{2(r-1)},$$
one has
$$\Phi_r(\alpha_*) < 0. \tag{C3}$$

Proof. For (i), the function $u \mapsto r \log u - c_r(u-1)$ is concave and has its maximum on $[r, \infty)$ at $u = r/c_r$, since $0 < c_r < 1$. Therefore the left side of (C1) is at most
$$M_r := -\frac{4}{3}(r-3)\log r + 3 - \log r + r \log \frac{r}{c_r} - r + c_r.$$
Let $c_0 := \log(16/9)$. Since $r \geq 9$, $c_r \geq c_0$ and $c_r < \log 2$. Hence
$$M_r \leq \left(3 - \frac{r}{3}\right) \log r + r \log \frac{1}{c_0} - r + 3 + \log 2.$$
The right side is decreasing for $r \geq 9$ and is negative at $r = 9$. This proves (i).

For (ii), use
$$\log \kappa_r = -\log 2 - \log(1 - 1/r) \leq -\log 2 + \frac{1}{r-1}.$$
Then
$$\log(e^3 r^3 \kappa_r^{2r-1}) \leq 3 + 3\log r - (2r-1)\log 2 + \frac{2r-1}{r-1}.$$
The final expression is decreasing for $r \geq 9$ and is negative at $r = 9$.

For (iii), note that $\alpha_* = \kappa_r / r$, so
$$\Phi_r(\alpha_*) = \kappa_r \left[2 \log \kappa_r + \frac{3 + 2\log r - 3 \log \kappa_r}{r}\right].$$
Since $1/2 \leq \kappa_r \leq 9/16$ for $r \geq 9$,
$$2 \log \kappa_r + \frac{3 + 2\log r - 3\log \kappa_r}{r} \leq 2 \log \frac{9}{16} + \frac{3 + 2\log r + 3 \log 2}{r}.$$
The second term on the right is decreasing for $r \geq 9$, and at $r = 9$ the whole right side is negative. Hence $\Phi_r(\alpha_*) < 0$.

Lemma 9 (Endpoint compression). *Let*
$$E_N(h, a, b) := \binom{h}{a} \binom{N - h + 1}{a} \binom{N/2}{b} \binom{N/2 + 1}{b}, \qquad k := a + b.$$
Assume $w > N/\log_2 N$ and $w \leq n = N/r$. For all sufficiently large N, uniformly in h, a, b, the following bounds hold.

(i) *If $k \le w-1$, then*

$$E_N(h,a,b) \le C_r \left(\frac{D(w)N}{w}\right)^{2w}. \tag{EC0}$$

(ii) *If $k = w-1+s$ with $s \ge 1$ and $k \le N/4$, then*

$$(eN)^{-2s} E_N(h,a,b) \le C_r \left(\frac{D(w)N}{w}\right)^{2w} k^{-2s}. \tag{ECs}$$

Proof. By Vandermonde,

$$\binom{h}{a}\binom{N-h+1}{a} \le \binom{N+1}{2a}, \qquad \binom{N/2}{b}\binom{N/2+1}{b} \le \binom{N+1}{2b}.$$

A second application of Vandermonde gives the pointwise bound

$$E_N(h,a,b) \le \binom{2N+2}{2k}. \tag{10}$$

If $k \le w-1$, then $2k \le 2w-2$. Since $w \le N/r \le N/9$, the coefficients $\binom{2N+2}{j}$ are increasing for $0 \le j \le 2w$. In the range $w < N/r^{4/3}$,

$$\binom{2N+2}{2k} \le \binom{2N+2}{2w} \le \left(\frac{e(N+1)}{w}\right)^{2w} \le C\left(\frac{eN}{w}\right)^{2w}.$$

In the range $w \ge N/r^{4/3}$, apply (EBC) with $M = 2N+2$ and $K = 2w$. Since $p_r(2N+2) \le 2w \le (2N+2)/2$ for large N,

$$\binom{2N+2}{2k} \le \binom{2N+2}{2w} \le C_r \left(\frac{D_r N}{w}\right)^{2w}.$$

This proves (EC0).

Now let $k = w-1+s \ge w$ and $k \le N/4$. In the range $w < N/r^{4/3}$,

$$\begin{aligned}(eN)^{-2s}\binom{2N+2}{2k} &\le (eN)^{-2s}\left(\frac{e(N+1)}{k}\right)^{2k} \\ &= \left(\frac{e(N+1)}{k}\right)^{2(w-1)}\left(\frac{N+1}{Nk}\right)^{2s} \\ &\le C\left(\frac{eN}{w}\right)^{2w} k^{-2s}.\end{aligned}$$

In the range $w \ge N/r^{4/3}$, apply (EBC) with $M = 2N+2$ and $K = 2k$. The hypotheses hold because $k/(N+1) \ge p_r$ and $k \le N/4$. Thus

$$\begin{aligned}(eN)^{-2s}\binom{2N+2}{2k} &\le (eN)^{-2s}\left(D_r\frac{N+1}{k}\right)^{2k} \\ &= \left(D_r\frac{N+1}{k}\right)^{2(w-1)}\left(\frac{D_r}{e}\right)^{2s}\left(\frac{N+1}{N}\right)^{2s} k^{-2s} \\ &\le C_r\left(\frac{D_r N}{w}\right)^{2w} k^{-2s},\end{aligned}$$

where $D_r < e$ and $s \le N$ were used. This proves (ECs).

Lemma 10 (Feasible no-surplus high-weight kernel). *Assume N is even and divisible by r. Fix $r \ge 9$ and write $n = N/r$. For all sufficiently large N,*

$$\sum_{w>N/\log_2 N}^{\lfloor N/(2(r-1))\rfloor} \binom{n}{w} \sum_{a+b\le w-1} \sum_{1\le h\le N:\ rw-a\le h\le N/2+b} \left(\frac{D(w)N}{w}\right)^{2w} \left(\frac{h+a}{N}\right)^{rw} \left(\frac{1}{2}+\frac{b}{N}\right)^{h} \le \exp\left(-\Omega_r\left(\frac{N}{\log N}\right)\right). \tag{NSK}$$

Proof. We split according to $D(w)$.

Range I: $N/\log_2 N < w < N/r^{4/3}$. Put $u = (h+a)/w$. Then $u \ge r$, and since $a \le a+b \le w-1$, one has $h \ge (u-1)w$. Also $b \le w-1$ and $w \le N/(2(r-1))$, so $1/2 + b/N \le \kappa_r$. Using $\binom{n}{w} \le (eN/(rw))^w$, one Range I summand is at most

$$\exp\left\{w\left[-(r-3)\log\frac{N}{w} + 3 - \log r + r\log u - c_r(u-1)\right]\right\}.$$

Since $w < N/r^{4/3}$, $\log(N/w) > (4/3)\log r$. Lemma 8(i) gives a constant $\eta_r > 0$ such that every Range I summand is at most $\exp(-\eta_r w)$. There are at most N^3 choices of (w, a, b, h), and $w > N/\log_2 N$, so Range I contributes $\exp(-\Omega_r(N/\log N))$.

Range II: $w \ge N/r^{4/3}$. Here $w \le N/(2(r-1))$. Since $a+b \le w-1$, the feasibility constraints imply

$$h + a \le \frac{N}{2} + a + b \le \frac{N}{2} + w - 1 \le \kappa_r N, \qquad h \ge rw - a \ge (r-1)w.$$

Moreover $b \le w-1$, so $1/2 + b/N \le \kappa_r$. Hence each summand is at most

$$\left[\frac{eD_r^2}{r}\left(\frac{N}{w}\right)^3 \kappa_r^{2r-1}\right]^w.$$

Since $w \ge N/r^{4/3}$, $(N/w)^3 \le r^4$, and the bracket is at most $eD_r^2 r^3 \kappa_r^{2r-1} < 1$ by (9) and Lemma 8(ii). Thus each Range II summand is $\exp(-\Omega_r(w)) = \exp(-\Omega_r(N))$, and the polynomial number of summands is harmless.

Lemma 11 (Surplus perturbations with leftover gain). *Assume $w > N/\log_2 N$. Let*

$$s_1, s_2 \ge 0, \qquad s = s_1 + s_2, \qquad k = w - 1 + s.$$

Let a_0, b_0, h be nonnegative integers, and set

$$a := a_0 + s_1, \qquad b := b_0 + s_2.$$

Assume

$$h + a \geq rw, \qquad h \leq \frac{N}{2} + b, \qquad a \leq \frac{rw}{2}, \qquad 1 \leq h \leq N, \qquad b \leq \frac{h}{2}.$$

Define $U_0 := \max\{h + a_0, rw\}$. *Then, for all sufficiently large* N,

$$k^{-2s} \left(\frac{h+a}{U_0}\right)^{rw} \left(\frac{\frac{1}{2} + \frac{b}{N}}{\frac{1}{2} + \frac{b_0}{N}}\right)^{h} \leq \left(\frac{k}{2}\right)^{-2s}. \tag{SP-left}$$

Proof. The case $s = 0$ is immediate. Assume $s \geq 1$. First, $U_0 \geq rw$ and $h + a \leq U_0 + s_1$. Also $s_1 \leq a \leq rw/2 \leq U_0/2$. For $0 \leq t \leq 1/2$, $\log_2(1 + t) \leq 2t$, so

$$rw \log_2 \frac{h+a}{U_0} \leq rw \log_2 \left(1 + \frac{s_1}{U_0}\right) \leq 2s_1.$$

For the second-stage perturbation, put $M = N/2 + b_0$. Then

$$\frac{\frac{1}{2} + b/N}{\frac{1}{2} + b_0/N} = 1 + \frac{s_2}{M}.$$

The feasibility hypothesis $h \leq N/2 + b = M + s_2$ and the condition $b \leq h/2$ imply $s_2 \leq h/2 \leq (M + s_2)/2$, hence $s_2 \leq M$. With $x = s_2/M \in [0, 1]$ and $(1 + x) \log_2(1 + x) \leq 2x$, we get

$$h \log_2 \left(1 + \frac{s_2}{M}\right) \leq M(1 + x) \log_2(1 + x) \leq 2s_2.$$

Thus the total perturbation in base-2 logarithm is at most $2s$, and

$$\begin{aligned} rw \log_2 \frac{h+a}{U_0} + h \log_2 \left(\frac{\frac{1}{2} + b/N}{\frac{1}{2} + b_0/N}\right) - 2s \log_2 k \\ \leq 2s - 2s \log_2 k = -2s \log_2(k/2). \end{aligned}$$

Exponentiating proves the claim.

Lemma 12 (Deficit summability). *Assume* $w > N/\log_2 N$. *For all nonnegative integers* d_1, d_2 *and all sufficiently large* N,

$$\sum_{s_1 \geq d_1,\ s_2 \geq d_2} \left(\frac{w - 1 + s_1 + s_2}{2}\right)^{-2(s_1+s_2)} \leq C \left(\frac{w}{2}\right)^{-2(d_1+d_2)}. \tag{DS}$$

Proof. Put $d = d_1 + d_2$ and write $s_i = d_i + t_i$. For fixed $m = t_1 + t_2$ there are $m + 1$ choices of (t_1, t_2). If $d + m \geq 1$, then $w - 1 + d + m \geq w$, and hence

$$\left(\frac{w - 1 + d + m}{2}\right)^{-2(d+m)} \leq \left(\frac{w}{2}\right)^{-2d} \left(\frac{w}{2}\right)^{-2m}.$$

The exceptional case $d = m = 0$ contributes 1. Therefore the whole sum is at most

$$\left(\frac{w}{2}\right)^{-2d}\left(1+\sum_{m\geq 1}(m+1)\left(\frac{w}{2}\right)^{-2m}\right),$$

and the parenthesized factor is bounded by an absolute constant for large N.

Lemma 13 (Deficit kernel summation). *Fix $w > N/\log_2 N$ and nonnegative integers a_0, b_0 satisfying $a_0 + b_0 = w - 1$. Put*

$$L := rw - a_0, \qquad U := \frac{N}{2} + b_0, \qquad v_0 := \frac{1}{2} + \frac{b_0}{N}, \qquad \theta := \left(\frac{w}{2}\right)^{-2},$$

and $U_0(h) := \max\{h+a_0, rw\}$. Then the following bounds hold for all sufficiently large N.

(i) *If $L \leq U$, then*

$$\sum_{h=1}^{N}\left(\frac{U_0(h)}{N}\right)^{rw} v_0^h \theta^{(L-h)_+ + (h-U)_+} \leq C\sum_{h=L}^{U}\left(\frac{h+a_0}{N}\right)^{rw} v_0^h. \quad \text{(DK-below)}$$

(ii) *If $L > U$, then*

$$\sum_{h=1}^{N}\left(\frac{U_0(h)}{N}\right)^{rw} v_0^h \theta^{(L-h)_+ + (h-U)_+} \leq CN\left(\frac{rw}{N}\right)^{rw} v_0^U \theta^{L-U}. \quad \text{(DK-above)}$$

Proof. Assume first that $L \leq U$. If $h < L$, then $U_0(h) = rw = U_0(L)$, and

$$v_0^h \theta^{L-h} = v_0^L\left(\frac{\theta}{v_0}\right)^{L-h} \leq v_0^L\left(\frac{8}{w^2}\right)^{L-h},$$

because $v_0 \geq 1/2$. Hence the lower tail is bounded by a constant times the boundary term at $h = L$. If $h > U$, then $U + a_0 \geq rw$, so $U_0(h) = h + a_0$ and $U_0(U) = U + a_0$. Writing $t = h - U$,

$$\left(\frac{U_0(h)}{U_0(U)}\right)^{rw} v_0^{h-U}\theta^{h-U} \leq \left(1+\frac{t}{U+a_0}\right)^{rw}\theta^t \leq \left(\frac{4e}{w^2}\right)^t,$$

since $U + a_0 \geq rw$. Thus the upper tail is bounded by a constant times the boundary term at $h = U$. The interval contribution is exactly the right side of (DK-below).

Now assume $L > U$. If $h < U$, comparison with $h = U$ gives the geometric factor $(8/w^2)^{U-h}$. If $h > L$, comparison with $h = L$ gives the geometric factor $(4e/w^2)^{h-L}$. For $U \leq h \leq L$, one has $U_0(h) = rw$ and $(L-h)_+ + (h-U)_+ = L - U$. Also $v_0^h \leq v_0^U$, since $0 < v_0 < 1$. There are at most N values of h in this middle interval. This proves (DK-above).

Lemma 14 (Above-cutoff master sum). *For fixed $r \geq 9$ and all sufficiently large N,*

$$\sum_{w>(N/2-1)/(r-1)}^{N/r} \binom{n}{w} \left(\frac{D_r N}{w}\right)^{2w} \left(\frac{rw}{N}\right)^{rw} \left(\frac{1}{2}+\frac{w}{N}\right)^{N/2+w} \left(\frac{w}{2}\right)^{-2((r-1)w+1-N/2)} \leq \exp(-\Omega_r(N)). \tag{ACM}$$

Proof. Put $\alpha = w/N$ and $\alpha_* = 1/(2(r-1))$. The positive part of the summand, that is, the summand without the final surplus factor, is at most

$$\exp\{N\Psi_r(\alpha) + O_r(\log N)\},$$

where

$$\Psi_r(\alpha) := \alpha \log \frac{eD_r^2}{r\alpha^3} + r\alpha \log(r\alpha) + \left(\frac{1}{2}+\alpha\right) \log\left(\frac{1}{2}+\alpha\right).$$

Since $D_r < e$, we have $\Psi_r(\alpha) \leq \Phi_r(\alpha)$, where Φ_r is the function from Lemma 8(iii). That lemma gives $\Phi_r(\alpha_*) < 0$. By continuity, there are constants $\gamma_r, \eta_r > 0$ such that

$$\Psi_r(\alpha) \leq -\gamma_r \qquad \text{whenever } |\alpha - \alpha_*| \leq \eta_r.$$

The part of the sum with $\alpha \leq \alpha_* + \eta_r$ is therefore $\exp(-\Omega_r(N))$; the surplus factor is at most 1.

For $\alpha \geq \alpha_* + \eta_r$, the positive exponent $\Psi_r(\alpha)$ is bounded above by a constant B_r, because $\alpha \in [\alpha_*, 1/r]$. On the other hand, $(r-1)w + 1 - N/2 \geq (r-1)\eta_r N$ for all sufficiently large N, and hence the surplus factor contributes at most

$$\exp\{-2(r-1)\eta_r N \log(w/2) + O_r(N)\} \leq \exp(-\Omega_r(N \log N)).$$

This dominates $\exp(B_r N)$. Summing over at most N possible values of w proves the lemma.

Lemma 15 (High-weight marked joint covering bound with surplus zeros). *Assume N is even and divisible by r. Fix $r \geq 9$, and write $n = N/r$. Assume $q - 1 \geq (eN)^2$. Then, for all sufficiently large N,*

$$\sum_{w=\lfloor N/\log_2 N\rfloor+1}^{n} \binom{n}{w} \sum_{a=0}^{\lfloor rw/2\rfloor} \sum_{h=1}^{N} \sum_{b=0}^{\lfloor h/2\rfloor} (eN)^{-2s(w,a,b)} \mathcal{C}_1^{\bullet}(w,h,a)\mathcal{C}_2^{\bullet}(h,b) \leq \exp\left(-\Omega_r\left(\frac{N}{\log N}\right)\right),$$

where $s(w,a,b) = (a+b-(w-1))_+$. Consequently, the same bound holds with $(q-1)^{-s(w,a,b)}$ in place of $(eN)^{-2s(w,a,b)}$.

Proof. Write $k = a + b$ and $s = (k - (w-1))_+$. A nonzero summand necessarily satisfies

$$h + a \geq rw, \qquad h \leq N/2 + b. \tag{11}$$

Indeed these are the nonvanishing conditions for $\binom{h}{rw-a}$ and $\binom{N/2}{h-b}$.

No-surplus terms. Assume $k \le w-1$. Then (11) implies

$$rw \le h + a \le N/2 + a + b \le N/2 + w - 1,$$

so no such term exists unless $w \le (N/2-1)/(r-1)$. By Lemma 9(i), the four endpoint factors are bounded by $C_r(D(w)N/w)^{2w}$. Also $h + a \le N/2 + k \le N/2 + w - 1 < N$ and $N/2 + b < N$. Hence the standard binomial-ratio bound $\binom{X}{M}/\binom{N}{M} \le (X/N)^M$, valid for $X \le N$, gives

$$\frac{\binom{h}{rw-a}}{\binom{N}{rw}} \le \frac{\binom{h+a}{rw}}{\binom{N}{rw}} \le \left(\frac{h+a}{N}\right)^{rw},$$

and

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

Therefore the total no-surplus contribution is bounded by a constant multiple of (NSK), and is $\exp(-\Omega_r(N/\log N))$.

Large surplus totals. If $k > N/4$, then, since $w \le N/r \le N/9$,

$$s = k - (w-1) \ge N/8$$

for all sufficiently large N. The scalar factor contributes at most $(eN)^{-N/4}$, while the remaining combinatorial and ratio factors are at most $\exp(O_r(N))$ and there are only polynomially many indices. Thus the total contribution of $k > N/4$ is $\exp(-\Omega(N \log N))$.

Surplus terms with $k \le N/4$. It remains to consider $s \ge 1$ and $k \le N/4$. For each surplus pair choose a surplus decomposition

$$a = a_0 + s_1, \qquad b = b_0 + s_2, \qquad a_0 + b_0 = w - 1, \qquad s_1 + s_2 = s.$$

For example, take $a_0 = \min\{a, w-1\}$ and $b_0 = w - 1 - a_0$. In the upper bound we may sum over all such decompositions.

By Lemma 9(ii), the endpoint factors together with $(eN)^{-2s}$ contribute at most

$$C_r \left(\frac{D(w)N}{w}\right)^{2w} k^{-2s}.$$

Since $k \le N/4$, the feasibility inequalities give $h + a \le N/2 + k \le 3N/4$ and $N/2 + b \le 3N/4$. Therefore the same marked-ratio bounds used in the no-surplus case are valid. They, followed by Lemma 11, give the pointwise comparison

$$\begin{aligned}(eN)^{-2s}&\mathcal{C}_1^{\bullet}(w,h,a)\mathcal{C}_2^{\bullet}(h,b) \\ &\le C_r \left(\frac{D(w)N}{w}\right)^{2w} \left(\frac{U_0}{N}\right)^{rw} \left(\frac{1}{2} + \frac{b_0}{N}\right)^{h} \left(\frac{k}{2}\right)^{-2s}, \end{aligned} \tag{S-point}$$

where $U_0 = \max\{h + a_0, rw\}$.

For fixed w, a_0, b_0, h, set

$$L := rw - a_0, \qquad U := \frac{N}{2} + b_0.$$

The feasibility inequalities (11) imply

$$s_1 \geq (L - h)_+, \qquad s_2 \geq (h - U)_+.$$

Summing (S-point) over all surplus decompositions and using Lemma 12, the surplus contribution is at most a constant multiple of

$$\sum_{w > N/\log_2 N} \binom{n}{w} \sum_{a_0 + b_0 = w-1} \sum_{h=1}^{N} \left(\frac{D(w)N}{w}\right)^{2w} \left(\frac{\max\{h + a_0, rw\}}{N}\right)^{rw} \cdot \left(\frac{1}{2} + \frac{b_0}{N}\right)^{h} \left(\frac{w}{2}\right)^{-2((rw - a_0 - h)_+ + (h - N/2 - b_0)_+)}. \tag{D-kernel}$$

If $w \leq (N/2 - 1)/(r - 1)$, then $L \leq U$. Lemma 13(i) bounds the h-sum in (D-kernel) by a constant times the feasible exact-budget kernel

$$\sum_{h=L}^{U} \left(\frac{h + a_0}{N}\right)^{rw} \left(\frac{1}{2} + \frac{b_0}{N}\right)^{h}.$$

After summing over w, a_0, b_0, this is a sub-sum of the feasible no-surplus kernel in Lemma 10; hence the below-cutoff surplus contribution is

$$\exp\left(-\Omega_r\left(\frac{N}{\log N}\right)\right).$$

It remains to consider $w > (N/2 - 1)/(r - 1)$. For $r \geq 9$ this range lies inside $w \geq N/r^{4/3}$ for all sufficiently large N, so $D(w) = D_r$. Now $L > U$, and

$$L - U = (r - 1)w + 1 - N/2.$$

Lemma 13(ii) bounds the h-sum in (D-kernel) by

$$CN \left(\frac{rw}{N}\right)^{rw} \left(\frac{1}{2} + \frac{b_0}{N}\right)^{N/2 + b_0} \left(\frac{w}{2}\right)^{-2((r-1)w + 1 - N/2)}.$$

The function $v \mapsto v^{Nv}$ is increasing on $[1/2, 1]$, because its logarithmic derivative is $N(\log v + 1) > 0$. Since $b_0 \leq w - 1$, the preceding display is at most

$$CN \left(\frac{rw}{N}\right)^{rw} \left(\frac{1}{2} + \frac{w}{N}\right)^{N/2 + w} \left(\frac{w}{2}\right)^{-2((r-1)w + 1 - N/2)}.$$

There are at most w choices of (a_0, b_0), so the above-cutoff surplus contribution is bounded by a polynomial factor times the master sum in Lemma 14. Hence it is $\exp(-\Omega_r(N))$.

Combining the no-surplus part, the $k > N/4$ part, the below-cutoff surplus part, and the above-cutoff surplus part proves the high-weight estimate. Finally, $(q - 1)^{-s} \leq (eN)^{-2s}$ under $q - 1 \geq (eN)^2$.

3.6 Main theorem

Theorem 1 (Distance tail bound for $r \geq 9$). *Assume N is even and divisible by r. Fix $r \geq 9$ and a constant $0 < \delta \leq 1/2$. Assume $q - 1 \geq (eN)^2$. Let $G = RP_1AP_2A$ be the large-field RAA generator matrix above. Then, for all sufficiently large N,*

$$\mathbb{P}\left[d_{\min}(G) \leq \delta N\right] \leq \widetilde{O}_r(N^{1-r}).$$

Proof. It suffices to prove the claim for $\delta = 1/2$, since

$$\{d_{\min}(G) \leq \delta N\} \subseteq \{d_{\min}(G) \leq N/2\} \qquad (0 < \delta \leq 1/2).$$

We therefore bound the latter event.

Split nonzero messages by weight. The case $w = 1$ is Lemma 6. For $2 \leq w \leq N/\log_2 N$, apply the marked covering reduction, the joint scalar anti-cancellation lemma, and Lemma 7. For $w > N/\log_2 N$, apply the same reduction and Lemma 15. The three contributions are respectively $O_r(N^{1-r})$, $\widetilde{O}_r(N^{1-r})$, and $\exp(-\Omega_r(N/\log N))$.

3.7 Matching lower bound

Lemma 16 (Weight-one lower bound). *Fix $r \geq 4$ and a constant $0 < \delta \leq 1/2$. Assume N is even and divisible by r, and write $n = N/r$. Then there exists a constant $c_{r,\delta} > 0$ such that, for all sufficiently large N,*

$$\mathbb{P}\left[d_{\min}(G) \leq \delta N\right] \geq c_{r,\delta} N^{1-r}.$$

Proof. Let $I = \{N - r + 1, \ldots, N\}$. For each message coordinate $u \in [n]$, let

$$\mathcal{A}_u := \{\Pi_1(B_u) = I\}.$$

Then

$$\mathbb{P}[\mathcal{A}_u] = \binom{N}{r}^{-1}.$$

The events $\mathcal{A}_u$ are pairwise disjoint, because the sets B_u are pairwise disjoint and two disjoint r-sets cannot both be mapped onto I.

On $\mathcal{A}_u$, the first-stage output $y = e_u RP_1A$ is nonzero and supported inside I, hence $1 \leq h := \mathrm{wt}(y) \leq r$. Conditional on the first-stage realization, let $S = \mathrm{supp}(y)$.

Put

$$L := \lfloor \delta N \rfloor, \qquad H := \{N - L + 1, \ldots, N\}.$$

The event

$$\Pi_2(S) \subseteq H$$

has probability

$$\frac{\binom{L}{h}}{\binom{N}{h}} \geq c'_{r,\delta} > 0$$

uniformly for $1 \le h \le r$ and all sufficiently large N. If it occurs, the second accumulator output is supported in the final L coordinates, so

$$\mathrm{wt}(e_u G) \le L \le \delta N.$$

Hence, for $\mathcal{E}_u$ equal to the intersection of these two events,

$$\mathbb{P}[\mathcal{E}_u] \ge c'_{r,\delta} \binom{N}{r}^{-1}.$$

The events $\mathcal{E}_u$ are pairwise disjoint, so

$$\mathbb{P}[d_{\min}(G) \le \delta N] \ge \sum_{u=1}^{n} \mathbb{P}[\mathcal{E}_u] \ge \frac{N}{r} c'_{r,\delta} \binom{N}{r}^{-1} \ge c_{r,\delta} N^{1-r}.$$

Corollary 1 (Tightness up to polylogarithmic factors). *Assume N is even and divisible by r. Fix $r \ge 9$, a constant $0 < \delta \le 1/2$, and assume*

$$q - 1 \ge (eN)^2.$$

Then, for all sufficiently large N,

$$c_{r,\delta} N^{1-r} \le \mathbb{P}\left[d_{\min}(G) \le \delta N\right] \le \widetilde{O}_r(N^{1-r})$$

for some constant $c_{r,\delta} > 0$.

4 Conclusion

We proved an asymptotically optimal distance-tail bound for the large-field RAA ensemble. For every fixed $r \ge 9$, every fixed $0 < \delta \le 1/2$, and every field size satisfying $q - 1 \ge (eN)^2$, the probability that the minimum distance falls below δN is at most $\widetilde{O}_r(N^{1-r})$. A matching weight-one lower bound shows that this polynomial order is unavoidable, up to polylogarithmic factors.

The proof identifies the mechanism behind the large-field improvement. The random diagonal scalings turn accumulator cancellations into scalar equations, so cancellations beyond the message degrees of freedom incur field-size penalties. Combined with the gap-covering reduction and the high-weight deficit summation, this suppresses all bad-codeword contributions except the weight-one terminal-window obstruction.

The binary companion lower bound shows that this improvement reflects a genuine difference between the ensembles. Over $\mathbb{F}_2$, adjacent repeated copies can cancel deterministically, leading to low-distance events of order $N^{1-\lceil r/2 \rceil}$, and of order $N^{1-r/2}$ for even r. Thus the large-field and binary RAA ensembles have different low-distance tail exponents, not merely different currently known proof guarantees.

The complementary obstructions in the appendices clarify the assumptions needed for the exponent $1-r$. First, the field size cannot be made sublinear without changing the polynomial tail: if $q - 1 = N^\gamma$ with $0 \le \gamma < 1$, then weight-one

paired cancellations give a lower bound with exponent strictly larger than $1-r$. Thus some essentially linear field-size growth is necessary for an N^{1-r}-order tail, although our upper bound assumes the stronger condition $q-1 \geq (eN)^2$. Second, the random diagonal scalings are essential to the large-field exponent. If the scalings are removed, opposite-valued weight-two messages create deterministic first-accumulator cancellations and yield low-distance events of order $\Omega_{r,\delta}(N^{2-r})$, polynomially larger than the $\widetilde{O}_r(N^{1-r})$ tail for the randomly scaled ensemble.

Several natural questions remain. One is to sharpen the field-size requirement between the linear lower obstruction and the quadratic-size assumption used in the upper bound. Another is to sharpen the constants and evaluate the pre-compression covering sums numerically for cryptographic parameter sets. It would also be interesting to extend the optimal-tail analysis to smaller repetition factors, especially $r = 4$, which is important in applications, and to develop matching upper bounds for related variants such as the no-scaling ensemble in large characteristic. Finally, the gap-covering and scalar anti-cancellation framework may be useful for other accumulator-based code families.

A Binary Companion Lower Bound

In this appendix we record a lower bound for the binary RAA ensemble

$$G = R\Pi_1 A\Pi_2 A \in \mathbb{F}_2^{n\times N}.$$

Theorem 2 (Binary lower bound). *Fix an integer $r \geq 4$ and a constant $0 < \delta < 1/2$. For the binary RAA ensemble*

$$G = R\Pi_1 A\Pi_2 A \in \mathbb{F}_2^{n\times N}, \qquad N = rn,$$

where Π_1, Π_2 are independent uniform permutation matrices on $[N]$, we have, for all sufficiently large even N divisible by r,

$$\mathbb{P}[\delta(G) \leq \delta] \geq \left(\frac{r!}{r\lfloor r/2\rfloor!}\,\delta^{\lceil r/2\rceil} + o_{r,\delta}(1)\right) N^{1-\lceil r/2\rceil}.$$

In particular,

$$\mathbb{P}[\delta(G) \leq \delta] \geq c_{r,\delta} N^{1-\lceil r/2\rceil}$$

for some $c_{r,\delta} > 0$.

Proof. Set

$$s := \left\lfloor \frac{r}{2} \right\rfloor, \qquad t := \left\lceil \frac{r}{2} \right\rceil.$$

For each $u \in [n]$, let $B_u = \mathrm{supp}(e_u R)$, so $|B_u| = r$ and the sets B_u are pairwise disjoint.

Call an r-subset $T \subseteq [N]$ admissible as follows. If $r = 2s$, then T consists of s disjoint adjacent pairs. If $r = 2s + 1$, then T consists of s disjoint adjacent pairs together with the final coordinate N. The number of admissible subsets is

$$M_r(N) = \binom{N - t}{s}.$$

Indeed, the starting positions of the adjacent pairs are transformed by $b_i = a_i - (i - 1)$ into an arbitrary s-subset of $\{1, \ldots, N - t\}$.

Let

$$E_u := \{\Pi_1(B_u) \text{ is admissible}\}, \qquad X := \sum_{u=1}^{n} \mathbf{1}_{E_u}.$$

Then

$$\mathbb{P}(E_u) = \frac{M_r(N)}{\binom{N}{r}} = \left(\frac{r!}{s!} + o_r(1)\right) N^{-t},$$

and hence

$$\mathbb{E}X = \left(\frac{r!}{rs!} + o_r(1)\right) N^{1-t}.$$

If r is odd, all admissible sets contain N, so the events E_u are mutually exclusive and $\mathbb{P}(X > 0) = \mathbb{E}X$. If r is even, Bonferroni gives the same asymptotic lower bound: for $u \neq v$,

$$\mathbb{P}(E_u \cap E_v) \leq \frac{M_r(N)}{\binom{N}{r}} \frac{M_r(N)}{\binom{N-r}{r}} = O_r(N^{-2t}),$$

so

$$\mathbb{E}\binom{X}{2} = O_r(N^{2-2t}) = o_r(N^{1-t})$$

because $t \geq 2$. Therefore

$$\mathbb{P}(X > 0) \geq \left(\frac{r!}{rs!} + o_r(1)\right) N^{1-t}. \tag{12}$$

On E_u, the first accumulator output $z_u = e_u R \Pi_1 A$ has weight t. Each adjacent pair contributes exactly one nonzero prefix coordinate over $\mathbb{F}_2$, and in the odd case the forced final coordinate N contributes one additional nonzero coordinate. Conditional on $X > 0$, choose one witness u. Let $L = \lfloor \delta N \rfloor$ and $H = \{N - L + 1, \ldots, N\}$. Since Π_2 is independent,

$$\mathbb{P}[\mathrm{supp}(z_u \Pi_2) \subseteq H \mid \Pi_1, X > 0] = \frac{\binom{L}{t}}{\binom{N}{t}} = \delta^t + o_{r,\delta}(1).$$

If this event occurs, then $z_u \Pi_2 A = e_u G$ is supported in the final L coordinates and is nonzero, so $\delta(G) \leq \delta$. Combining with (12) proves the theorem.

Remark 1. When r is even, $\lceil r/2 \rceil = r/2$, so the lower bound is of order $N^{1-r/2}$.

B A Small-Field Weight-One Obstruction

In this appendix we record a complementary lower bound showing that the large-field tail N^{1-r} cannot hold uniformly over sublinear field sizes. The obstruction is already visible for weight-one messages: pairs of adjacent repeated copies can cancel after the first random diagonal scaling.

Theorem 3 (Small-field paired-cancellation lower bound). *Fix an integer $r \geq 4$ and a constant $0 < \delta < 1/2$. Put*

$$s := \left\lfloor \frac{r}{2} \right\rfloor, \qquad t := \left\lceil \frac{r}{2} \right\rceil.$$

For the large-field RAA ensemble

$$G = R\Pi_1 V_1 A \Pi_2 V_2 A \in \mathbb{F}_q^{n \times N}, \qquad N = rn,$$

where V_1, V_2 have independent uniform diagonal entries in $\mathbb{F}_q^*$, we have, for all sufficiently large even N divisible by r,

$$\mathbb{P}[d_{\min}(G) \leq \delta N] \geq \left(\frac{r!}{rs!}\, \delta^t + o_{r,\delta}(1) \right) N^{1-t} (q-1)^{-s}.$$

In particular, for even r,

$$\mathbb{P}[d_{\min}(G) \leq \delta N] \geq c_{r,\delta} N^{1-r/2} (q-1)^{-r/2} = c_{r,\delta} N \big(N(q-1)\big)^{-r/2}.$$

Proof. For $u \in [n]$, let $B_u = \operatorname{supp}(e_u R)$, so $|B_u| = r$ and the sets B_u are pairwise disjoint.

Call an r-subset $T \subseteq [N]$ paired-admissible as follows. If $r = 2s$, then T consists of s disjoint adjacent pairs. If $r = 2s+1$, then T consists of s disjoint adjacent pairs together with the final coordinate N. As in Appendix A, the number of such subsets is

$$M_r(N) = \binom{N-t}{s}.$$

Indeed, if the adjacent pairs start at positions $a_1 < \cdots < a_s$, with $a_{i+1} \geq a_i + 2$, then $b_i = a_i - (i-1)$ gives an arbitrary s-subset of $\{1, \ldots, N-t\}$.

Let F_u be the event that $\Pi_1(B_u)$ is paired-admissible and that, within each adjacent pair, the two first-stage scaled copies cancel. Conditional on a paired-admissible first-stage active set, each pair-cancellation condition has the form

$$\alpha + \alpha' = 0$$

with α, α' independent uniform elements of $\mathbb{F}_q^*$. Hence each pair cancels with probability $(q-1)^{-1}$, independently across the s pairs. Therefore

$$\mathbb{P}(F_u) = \frac{M_r(N)}{\binom{N}{r}} (q-1)^{-s} = \left(\frac{r!}{s!} + o_r(1) \right) N^{-t} (q-1)^{-s}.$$

Let

$$X := \sum_{u=1}^{n} \mathbf{1}_{F_u}.$$

If r is odd, every paired-admissible set contains N, so the events F_u are mutually exclusive. Hence

$$\mathbb{P}(X > 0) = \mathbb{E}X = \left(\frac{r!}{rs!} + o_r(1) \right) N^{1-t}(q-1)^{-s}.$$

If r is even, Bonferroni gives the same asymptotic lower bound. Indeed, for $u \neq v$,

$$\mathbb{P}(F_u \cap F_v) \leq \frac{M_r(N)}{\binom{N}{r}} \frac{M_r(N)}{\binom{N-r}{r}} (q-1)^{-2s} = O_r\left(N^{-2t}(q-1)^{-2s}\right).$$

Thus

$$\mathbb{E}\binom{X}{2} = O_r\left(N^{2-2t}(q-1)^{-2s}\right) = o_r\left(N^{1-t}(q-1)^{-s}\right),$$

because $t \geq 2$ and $q - 1 \geq 1$. Therefore, in both parity cases,

$$\mathbb{P}(X > 0) \geq \left(\frac{r!}{rs!} + o_r(1) \right) N^{1-t}(q-1)^{-s}. \tag{13}$$

On F_u, the first accumulator output

$$z_u := e_u R \Pi_1 V_1 A$$

has weight exactly t. Each adjacent canceling pair contributes exactly one nonzero output coordinate, namely the first coordinate of the pair. In the odd case, the forced final coordinate N contributes one additional nonzero output coordinate.

Condition on $X > 0$, and choose a deterministic witness u, for example the least u such that F_u occurs. Let

$$L := \lfloor \delta N \rfloor, \qquad H := \{N - L + 1, \ldots, N\}.$$

The second permutation Π_2 is independent of Π_1, V_1. Since $\mathrm{wt}(z_u) = t$, we have

$$\mathbb{P}[\Pi_2(\mathrm{supp}(z_u)) \subseteq H \mid \Pi_1, V_1, X > 0] = \frac{\binom{L}{t}}{\binom{N}{t}} = \delta^t + o_{r,\delta}(1).$$

If this event occurs, then $z_u \Pi_2 V_2$ is supported inside the final L coordinates. Therefore the second accumulator output

$$z_u \Pi_2 V_2 A = e_u G$$

is also supported inside the final L coordinates. It is nonzero because $z_u \Pi_2 V_2 \neq 0$ and the accumulator matrix A is invertible. Hence

$$\mathrm{wt}(e_u G) \leq L \leq \delta N,$$

and so $d_{\min}(G) \leq \delta N$.

Combining this conditional second-stage probability with (13) proves the theorem.

Corollary 2 (Sublinear fields cannot have the N^{1-r} tail). *Fix $r \geq 4$ and $0 < \delta < 1/2$. If $q - 1 = N^{\gamma}$ with $0 \leq \gamma < 1$, then*

$$\mathbb{P}[d_{\min}(G) \leq \delta N] \geq N^{\,1-\lceil r/2\rceil-\gamma\lfloor r/2\rfloor-o(1)}.$$

In particular, for even r,

$$\mathbb{P}[d_{\min}(G) \leq \delta N] \geq N^{\,1-(1+\gamma)r/2-o(1)}.$$

This exponent is strictly larger than $1 - r$ whenever $\gamma < 1$. Thus a field of at least essentially linear size is necessary for an N^{1-r}-order low-distance tail.

C A No-Scaling Weight-Two Obstruction

In this appendix we record a simple obstruction showing that the random diagonal scalings in the large-field ensemble are not merely a technical convenience. If the diagonal scalings are removed, then weight-two messages already create low-distance events of order N^{2-r}.

Let

$$G_0 := R\Pi_1 A\Pi_2 A \in \mathbb{F}_q^{n\times N}, \qquad N = rn,$$

where Π_1, Π_2 are independent uniform permutation matrices on $[N]$. No random diagonal matrices are used in this ensemble.

Theorem 4 (No-scaling weight-two obstruction). *Fix an integer $r \geq 4$ and a constant $0 < \delta < 1/2$. For the no-scaling RAA ensemble $G_0 = R\Pi_1 A\Pi_2 A$, over any finite field, we have, for all sufficiently large even N divisible by r,*

$$\mathbb{P}[d_{\min}(G_0) \leq \delta N] \geq \left(\frac{2^{r-1}r!}{r^2}\,\delta^r + o_{r,\delta}(1)\right) N^{2-r}.$$

In particular,

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

for some constant $c_{r,\delta} > 0$.

Proof. For $u \in [n]$, let $B_u = \mathrm{supp}(e_u R)$, so $|B_u| = r$ and the sets B_u are pairwise disjoint.

For an unordered pair $e = \{u, v\} \subseteq [n]$, $u \neq v$, define F_e to be the event that the $2r$ positions

$$\Pi_1(B_u) \cup \Pi_1(B_v)$$

form r disjoint adjacent pairs, and that each adjacent pair contains exactly one point of $\Pi_1(B_u)$ and exactly one point of $\Pi_1(B_v)$.

We first compute the probability of F_e. The images $\Pi_1(B_u)$ and $\Pi_1(B_v)$ are two ordered disjoint r-subsets of $[N]$. There are

$$\binom{N}{r}\binom{N-r}{r}$$

possible choices. The number of ways to choose r disjoint adjacent pairs in $[N]$ is

$$\binom{N-r}{r},$$

by the standard transformation from starting positions $a_1 < \cdots < a_r$, $a_{i+1} \geq a_i + 2$, to $b_i = a_i - (i-1)$. For each adjacent pair there are two choices for which point belongs to $\Pi_1(B_u)$, and the other point then belongs to $\Pi_1(B_v)$. Therefore

$$\mathbb{P}(F_e) = \frac{2^r \binom{N-r}{r}}{\binom{N}{r}\binom{N-r}{r}} = \frac{2^r}{\binom{N}{r}} = (2^r r! + o_r(1))\, N^{-r}.$$

Let

$$X := \sum_{\{u,v\} \subseteq [n]} \mathbf{1}_{F_{\{u,v\}}}.$$

Then

$$\mathbb{E}X = \binom{n}{2} \frac{2^r}{\binom{N}{r}} = \left(\frac{2^{r-1} r!}{r^2} + o_r(1) \right) N^{2-r}.$$

We next show that $\mathbb{P}(X > 0)$ has the same order. If $e \neq f$ are two distinct unordered message pairs, then

$$\mathbb{P}(F_e \cap F_f) = O_r(N^{-2r}).$$

Indeed, condition on F_e. If f is disjoint from e, then the event F_f is still a mixed-adjacent-pair event for two fresh r-sets in a remaining universe of size $N - O_r(1)$, and has probability $O_r(N^{-r})$. If f shares one message coordinate with e, then the positions of that shared coordinate are already fixed; for F_f to occur, the other r-set must occupy one adjacent neighbor of each of those r fixed positions, giving at most $O_r(1)$ possible r-sets among $\Theta_r(N^r)$ choices. Again the conditional probability is $O_r(N^{-r})$. Thus the displayed intersection bound holds in all cases.

Consequently,

$$\mathbb{E}\binom{X}{2} = O_r(N^{4-2r}) = o_r(N^{2-r}),$$

because $r \geq 4$. Bonferroni gives

$$\mathbb{P}(X > 0) \geq \mathbb{E}X - \mathbb{E}\binom{X}{2} = \left(\frac{2^{r-1} r!}{r^2} + o_r(1) \right) N^{2-r}. \tag{14}$$

Now suppose $F_{\{u,v\}}$ occurs. Consider the weight-two message

$$x = e_u - e_v.$$

This has weight 2 over every field; in characteristic 2, it is simply $e_u + e_v$. In each adjacent mixed pair, the two increments entering the first accumulator are 1 and -1, in some order. Hence the running sum is nonzero at the first position

of the pair and returns to its previous value at the second position. Since each pair has total increment zero, the first accumulator output

$$z := xR\Pi_1 A$$

has weight exactly r.

Condition on $X > 0$, and choose a deterministic witness $\{u, v\}$, for example the lexicographically first pair for which $F_{\{u,v\}}$ occurs. Let

$$L := \lfloor \delta N \rfloor, \qquad H := \{N - L + 1, \ldots, N\}.$$

The second permutation Π_2 is independent of Π_1. Since $\mathrm{wt}(z) = r$,

$$\mathbb{P}[\Pi_2(\mathrm{supp}(z)) \subseteq H \mid \Pi_1, X > 0] = \frac{\binom{L}{r}}{\binom{N}{r}} = \delta^r + o_{r,\delta}(1).$$

If this event occurs, then $z\Pi_2$ is supported inside the final L coordinates. Therefore the second accumulator output

$$z\Pi_2 A = xG_0$$

is also supported inside the final L coordinates. It is nonzero because the accumulator matrix A is invertible and $z\Pi_2 \neq 0$. Hence

$$\mathrm{wt}(xG_0) \leq L \leq \delta N,$$

so $d_{\min}(G_0) \leq \delta N$.

Combining this conditional second-stage probability with (14) proves the theorem.

References

1. 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. URL: https://eprint.iacr.org/2026/524.
2. Louay Bazzi, Mohammad Mahdian, and Daniel A. Spielman. The minimum distance of turbo-like codes. *IEEE Transactions on Information Theory*, 55(1):6–15, 2009. doi:10.1109/TIT.2008.2008114.
3. Alexander R. Block, Zhiyong Fang, Jonathan Katz, Justin Thaler, Hendrik Waldner, and Yupeng Zhang. Field-agnostic SNARKs from expand-accumulate codes. Cryptology ePrint Archive, Paper 2024/1871, 2024. URL: https://eprint.iacr.org/2024/1871, doi:10.1007/978-3-031-68403-6_9.
4. Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch, and Peter Scholl. Correlated pseudorandomness from expand-accumulate codes. Cryptology ePrint Archive, Paper 2022/1014, 2022. URL: https://eprint.iacr.org/2022/1014.
5. Martijn Brehm, Binyi Chen, Ben Fisch, Nicolas Resch, Ron D. Rothblum, and Hadas Zeilberger. Blaze: Fast SNARKs from interleaved RAA codes. Cryptology ePrint Archive, Paper 2024/1609, 2024. URL: https://eprint.iacr.org/2024/1609.
6. 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, Monticello, Illinois, 1998.
7. Alexander Golovnev, Jonathan Lee, Srinath Setty, Justin Thaler, and Riad S. Wahby. Brakedown: Linear-time and field-agnostic SNARKs for R1CS. Cryptology ePrint Archive, Paper 2021/1043, 2021. URL: https://eprint.iacr.org/2021/1043.
8. 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.
9. Jörg Kliewer, Kamil Sh. Zigangirov, Christian Koller, and Daniel J. Jr. Costello. Coding theorems for repeat multiple accumulate codes. arXiv preprint arXiv:0810.3422, 2008. URL: https://arxiv.org/abs/0810.3422.