Linear Distance for Fixed-Row-Weight Expand–Accumulate Codes over Arbitrary Fields

Majid Khabbazian

University of Alberta, Canada

Abstract. Expand–accumulate (EA) codes are sparse linear codes underlying constructions of correlated pseudorandomness and field-agnostic succinct arguments. In *Field-Agnostic SNARKs from Expand–Accumulate Codes* (CRYPTO 2024), Block et al. conjectured that a single fixed-row-weight EA component already achieves constant relative distance with inverse-polynomial failure probability.

We prove this conjecture in a stronger, field-uniform form. For every rate $R \in (0,1)$, there exists $\delta_R > 0$ such that, for every target exponent $C > 0$, one can choose $\gamma = \gamma(R, C) > 0$ for which the fixed-row ensemble with $t = \lceil \gamma \log N \rceil$ satisfies

$$\mathbb{P}\left[\min_{x \in \mathbb{F}_q^{\lfloor RN \rfloor} \setminus \{0\}} \mathrm{wt}(xEA) \le \delta_R N\right] \le N^{-C}.$$

The same constants work for every prime power q; in particular, the field may vary arbitrarily with the block length. Thus a single fixed-row-weight EA component is asymptotically good over all finite fields, and its polynomial reliability exponent can be made arbitrarily large by increasing the row-weight constant.

The proof separates sparse and high-weight messages. Sparse messages are handled through expansion and a compact analysis of accumulator cancellations, while larger messages are controlled by surplus linear constraints over large fields and a stochastic accumulator analysis over bounded fields. A boundary obstruction shows that, for $t = \Theta(\log N)$, inverse-polynomial failure is qualitatively optimal.

Keywords: expand–accumulate codes · sparse random codes · minimum distance · field-agnostic SNARKs · correlated pseudorandomness

1 Introduction

Expand–accumulate codes combine a sparse expansion layer with a simple prefix-sum accumulator. Their sparse structure makes encoding inexpensive, while the accumulator spreads local expansion events across long output intervals. This combination has proved useful in cryptographic constructions based on correlated pseudorandomness [4] and, more recently, in field-agnostic SNARKs [3].

The relevant ensemble is particularly simple. Let $n \approx RN$, let each row of an $n \times N$ matrix E contain exactly $t = \Theta(\log N)$ independently chosen

nonzero entries, and let A be the length-N prefix-sum accumulator. The resulting generator is

$$G_{\mathrm{EA}} = EA.$$

The fixed row weight is attractive both conceptually and algorithmically: every input symbol contributes exactly t sparse impulses. The main unresolved coding-theoretic question is whether this single component already has constant relative distance with inverse-polynomial failure probability.

Block et al. [3] prove the required behavior for messages of weight $O(N/\log N)$, but their high-weight analysis uses an additional Bernoulli component. They state the single fixed-row-weight assertion as Conjecture 4.6. The difficulty is not ordinary small-set expansion. Once the message support exceeds N/t, an unsaturated estimate of the form $\mathrm{wt}(xE) \gtrsim t\,\mathrm{wt}(x)$ is impossible because the right-hand side can exceed N. Moreover, over small fields, many algebraic cancellations can occur and a direct union bound over all field-valued messages is too expensive in the sparse range.

1.1 Our Result

We prove the fixed-row-weight conjecture and strengthen its field quantifier. The theorem is uniform over all finite fields, including fields whose size changes arbitrarily with N. It also permits any prescribed inverse-polynomial failure exponent while keeping the distance constant dependent only on the rate.

Theorem 1 (Informal main theorem). *For every fixed rate $R \in (0,1)$ there exists $\delta_R > 0$ such that the following holds. For every $C > 0$, one can choose $\gamma = \gamma(R, C) > 0$ so that, for $t = \lceil \gamma \log N \rceil$ and every prime power $q = q(N)$,*

$$\mathbb{P}_E\left[\min_{x \in \mathbb{F}_q^{\lfloor RN \rfloor} \setminus \{0\}} \mathrm{wt}(xEA) \le \delta_R N\right] \le N^{-C}$$

for all sufficiently large N.

Since the accumulator is invertible, positivity of the message minimum weight also implies that E has full row rank. Thus, with the same probability bound, EA generates an $[N, \lfloor RN \rfloor]_q$ code with relative distance at least δ_R.

At the coding-theoretic level, theorem 1 proves that the fixed-row-weight EA component alone achieves the conjectured distance; the additional Bernoulli component used in the previous high-weight analysis is therefore unnecessary for establishing distance. Whether that component can also be removed from a particular cryptographic construction depends on what other roles it plays in the construction.

1.2 Technical Overview

Let x be a nonzero message of Hamming weight w. Its nonzero entries activate w rows of E, which together place tw nonzero contributions among the N

coordinates before cancellations. The accumulator then replaces these coordinates by their running sums. Thus, for xEA to have small weight, many contributions must either cancel or leave nonzero running sums only over unusually short intervals.

The proof separates messages according to their Hamming weight w and, in the high-weight range, according to the field size. The three regimes use genuinely different mechanisms.

Sparse messages: forced noncancellations and random spacing. Suppose $w \leq \beta N/t$. With high probability, every set of w active rows touches almost tw distinct columns, and a constant fraction of these columns are touched by exactly one active row. At each such column, the corresponding coordinate of xE is necessarily nonzero, regardless of the values of the nonzero message entries. A simple deterministic argument then shows that $\Omega(tw)$ of the running sums immediately following touched columns are nonzero.

It remains to control all possible message values without taking a union bound over the q^w messages on a fixed support. Once the ordered collision pattern and coefficients are fixed, the running sums are linear forms in the w active message symbols. The set of positions at which these forms vanish can realize at most

$$\sum_{d=0}^{w-1} \binom{M}{d}$$

different patterns, where M is the number of touched columns. Crucially, this bound is independent of the field size.

The actual locations of the M touched columns are uniform and independent of these vanishing patterns. Each nonzero running sum persists until the next touched column, so $\mathrm{wt}(xEA)$ is the sum of the corresponding random gaps. If $k = \Omega(tw)$ running sums are nonzero, an exact random-gap calculation gives

$$\mathbb{P}[\mathrm{wt}(xEA) \leq \delta N] \leq \left(\frac{eM\delta}{k}\right)^k = \exp(-\Omega(tw))$$

for sufficiently small δ. This decay pays for the remaining union over supports and vanishing patterns. Since $t = \gamma \log N$, increasing γ yields any prescribed inverse-polynomial failure exponent.

High-weight messages over large fields: more equations than degrees of freedom. Now suppose $w > \beta N/t$. In this range one cannot expect the active rows to touch $\Theta(tw)$ distinct columns, since tw may exceed N. Instead, a saturated expansion property ensures that they touch at least

$$M \geq w + \eta N$$

columns.

If $\mathrm{wt}(xEA) \le \delta N$, then at most δN of the running sums following these columns can be nonzero. Hence at least

$$M - \delta N \ge w + (\eta - \delta)N$$

running sums must vanish. These are linearly independent prefix equations.

For a fixed message up to nonzero scaling, the random expansion values in distinct touched columns are independent, and every possible value has probability at most $1/(q-1)$. Thus b prescribed prefix equations hold with probability at most $(q-1)^{-b}$. There are only $(q-1)^{w-1}$ projective message choices on a fixed support, so after paying for all message values there remains the factor

$$(q-1)^{-(\eta-\delta)N}.$$

This linear surplus of equations dominates the exponentially many choices of supports and vanishing patterns once the field is larger than a suitable constant.

High-weight messages over bounded fields: a Markov-chain view of the accumulator. For the finitely many smaller fields, the preceding field-size gain is not strong enough. We instead fix a message and analyze the accumulator directly.

A weight-w message generates tw random nonzero contributions. We first replace the fixed-row sampling rule by sampling positions with replacement; this changes probabilities by only a subexponential factor. We then Poissonize the number of contributions. Under this model, the coordinates of xE become independent and identically distributed.

Let Y_j be the accumulated value after coordinate j. Because the increment distribution assigns the same probability to every nonzero field element, the process that records only whether Y_j is zero or nonzero is an exact two-state Markov chain. Its transition probabilities give an exponential bound on the probability that at most δN of the states are nonzero.

There are two ways this bound closes the union over messages. If $w = o(N)$, then the number of such messages is subexponential, while $tw/N \ge \beta$ keeps the Markov-chain penalty bounded away from zero. If $w/N \to \alpha > 0$, then $tw/N \to \infty$ and the accumulated value approaches the uniform distribution on $\mathbb{F}_q$. The resulting low-weight exponent dominates the full q-ary message entropy because

$$R\mathsf{h}(\alpha/R) + \alpha \log(q-1) \le R \log q$$

and δ is chosen so that

$$\mathsf{h}(\delta) + \delta \log(q-1) < (1-R)\log q.$$

This completes the high-weight argument over every bounded field.

Why the failure probability is only inverse polynomial. The polynomial tail is inherent when $t = \gamma \log N$. If all t nonzero entries of one row lie in the final δN coordinates, then the accumulated codeword generated by that row is also

supported in those final coordinates and therefore has weight at most δN. For a fixed row, this event has probability

$$\delta^{\,t+o(t)} = N^{-\gamma \log(1/\delta)+o(1)}.$$

Consequently, for fixed γ, the failure probability cannot in general be exponentially small. Increasing γ can make its polynomial exponent arbitrarily large, but cannot change its qualitative scale.

1.3 Related Work

EA codes were introduced by Boyle et al. [4] in the study of correlated pseudorandomness. Block et al. [3] developed their use in field-agnostic SNARKs, proved low-weight distance bounds for the fixed-row component, and formulated the single-component distance conjecture resolved here. Raghuraman, Rindal, and Tanguy [9] introduced expand–convolute codes and studied related concrete-distance phenomena. A different encode–repeat–accumulate construction was proposed by Baweja et al. [1].

Classical repeat–accumulate and repeat–multiple–accumulate codes are typically analyzed through input–output weight enumerators and random interleavers [6,7,2]. Our setting differs in that the expansion layer has fixed logarithmic row weight and the theorem must hold uniformly over every finite field. Sparse random generator matrices and sparse random matrix rank have also been studied in other coding regimes; see, for example, [8,5].

Organization. Section 2 defines the ensemble and states the precise theorem. Section 3 establishes the two support-expansion properties. Section 4 proves the sparse-message bound. Section 5 treats high-weight messages, separating the large-field argument in section 5.1 from the bounded-field analysis in section 5.2. Section 6 combines the regimes, while section 7 proves the terminal boundary obstruction.

2 Model and Main Theorem

Fix a rate $R \in (0,1)$. For each block length N, let

$$n = n_N := \lfloor RN \rfloor .$$

Let $q = q(N)$ be an arbitrary prime power, write $\mathbb{F}_q$ for the field of size q, and put

$$Q := q - 1, \qquad t = t_N := \lceil \gamma \log N \rceil ,$$

where logarithms are natural. All asymptotic statements are for sufficiently large N, so in particular $1 \le t < N/2$.

The rows of $E \in \mathbb{F}_q^{n \times N}$ are independent. In each row, the support is a uniform t-subset of $[N]$, and conditional on the support, the t nonzero entries are independent uniform elements of $\mathbb{F}_q^*$.

The accumulator $A \in \mathbb{F}_q^{N\times N}$ is the invertible matrix satisfying

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

for every row vector $z \in \mathbb{F}_q^N$.

For a generator matrix B, not assumed a priori to have full row rank, define its *message minimum weight* by

$$d_*(B) := \min_{x\in\mathbb{F}_q^n\setminus\{0\}} \mathrm{wt}(xB). \tag{2}$$

Thus $d_*(B) = 0$ exactly when B fails to have full row rank, equivalently, when there exists a nonzero row vector $x \in \mathbb{F}_q^n$ such that $xB = 0$. If $d_*(B) > 0$, then B has full row rank and $d_*(B)$ is the usual minimum distance of the code generated by B.

Theorem 2 (Uniform fixed-row EA distance theorem). *For every fixed $R \in (0, 1)$ and every $C > 0$, there exist constants*

$$\delta = \delta(R) > 0, \qquad \gamma = \gamma(R, C) > 0, \qquad N_0 = N_0(R, C),$$

such that the following holds for every $N \ge N_0$ and every prime power q, including a field size $q = q(N)$ chosen as an arbitrary function of N. If E is sampled as above with $n = \lfloor RN \rfloor$ and $t = \lceil \gamma \log N \rceil$, then

$$\mathbb{P}_E[d_*(EA) \le \delta N] \le N^{-C}. \tag{3}$$

Consequently, except with probability at most N^{-C}, the matrix E has row rank n and EA generates an $[N, n]_q$ code of relative distance at least δ.

The inverse-polynomial scale in (3) is qualitatively optimal for fixed γ; see section 7.

Remark 1 (Uniformity in the field). The same constants δ, γ, and N_0 work for all characteristics and all field sizes. In particular, the theorem is stronger than a statement in which q is fixed before N tends to infinity.

3 Expansion Properties

Ignoring coefficients, view the row supports of E as a bipartite graph with left vertex set $[n]$, right vertex set $[N]$, and left degree t. For $T \subseteq [n]$, let $\Gamma(T)$ denote its right-neighbor set.

We use two complementary expansion statements. Small row sets enjoy nearly lossless expansion, while all larger row sets have a linear additive surplus of neighbors.

3.1 Small-Set Expansion

Lemma 1 (Uniform small-set expansion). *Fix $\varepsilon \in (0, 1/2)$. There exists $\beta = \beta(\varepsilon) > 0$ with the following property. For every $D > 0$, if $t = \lceil \gamma \log N \rceil$ and γ is sufficiently large as a function of ε and D, then, except with probability at most N^{-D},*

$$|\Gamma(T)| \geq (1-\varepsilon)t|T| \tag{4}$$

for every $T \subseteq [n]$ satisfying $1 \leq |T| \leq \beta N/t$.

Proof. Fix $s = |T| \geq 2$ and put $u = \lfloor (1-\varepsilon)ts \rfloor$. If $|\Gamma(T)| \leq u$, then all row supports indexed by T lie in some u-subset of $[N]$. Hence

$$\mathbb{P}[\exists T : |T| = s,\ |\Gamma(T)| \leq u] \leq \binom{n}{s}\binom{N}{u}\left(\frac{\binom{u}{t}}{\binom{N}{t}}\right)^s \tag{5}$$

$$\leq \binom{n}{s}\binom{N}{u}\left(\frac{u}{N}\right)^{ts}. \tag{6}$$

The event $u < t$ is impossible, since a row support of size t cannot be contained in a set of size u. Hence assume $u \geq t$, and write

$$y := \frac{ts}{N} \leq \beta.$$

Using the standard estimate

$$\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b,$$

we bound the three factors in (6) separately. Since $n \leq RN$ and $s = yN/t$,

$$\frac{1}{s}\log\binom{n}{s} \leq \log\frac{en}{s} \leq \log\frac{eRt}{y}.$$

For the second binomial factor, the same estimate gives

$$\log\binom{N}{u} \leq u\log\frac{eN}{u}.$$

Because $u \leq (1-\varepsilon)ts = (1-\varepsilon)yN$ and $v \mapsto v\log(eN/v)$ is increasing for $v < N$, it follows that

$$\frac{1}{s}\log\binom{N}{u} \leq (1-\varepsilon)t\log\frac{e}{(1-\varepsilon)y}.$$

Finally,

$$\frac{1}{s}\log\left(\frac{u}{N}\right)^{ts} = t\log\frac{u}{N} \leq t\log\big((1-\varepsilon)y\big).$$

Adding these bounds and simplifying, the logarithm of the right-hand side of (6), divided by s, is at most

$$F_t(y) := \log\frac{eRt}{y} + (1-\varepsilon)t - \varepsilon t\log\frac{1}{(1-\varepsilon)y}. \tag{7}$$

For $\varepsilon t > 1$,

$$F_t'(y) = \frac{\varepsilon t - 1}{y} > 0,$$

so $F_t(y) \le F_t(\beta)$ whenever $y \le \beta$. Choose $\beta \in (0,1)$ sufficiently small that

$$\varepsilon \log \frac{1}{(1-\varepsilon)\beta} > 1 - \varepsilon + 2.$$

Then

$$F_t(y) \le \log \frac{eRt}{\beta} - 2t \le -t$$

for all sufficiently large N, since $\log(eRt/\beta) = o(t)$. Thus the probability in (6) is at most e^{-ts} for every $s \ge 2$ in the stated range.

The case $s = 1$ cannot fail because each row has exactly t distinct neighbors. Therefore

$$\sum_{s=2}^{\lfloor \beta N/t \rfloor} e^{-ts} \le \frac{e^{-2t}}{1 - e^{-t}} = O(e^{-2t}),$$

which is at most N^{-D} after increasing γ.

□

Corollary 1 (Unique neighbors). *On the event in lemma 1, every T with $|T| = w \le \beta N/t$ has at least*

$$u_T \ge (1 - 2\varepsilon)tw \tag{8}$$

right vertices of degree exactly one in the subgraph induced by T.

Proof. Let $M = |\Gamma(T)|$ and let u_T be the number of degree-one neighbors. Counting the tw incident edges gives

$$tw \ge u_T + 2(M - u_T) = 2M - u_T.$$

Therefore $u_T \ge 2M - tw \ge (1 - 2\varepsilon)tw$. □

3.2 Saturated Expansion

Lemma 2 (Saturated expansion). *Fix $R \in (0,1)$ and $\beta > 0$. There exist constants $\eta = \eta(R,\beta) > 0$ and $\kappa = \kappa(R,\beta) > 0$ such that, whenever $n \le RN$, $t \to \infty$, and $t = o(N)$,*

$$\mathbb{P}\left[\exists T \subseteq [n] : |T| > \frac{\beta N}{t}, \quad |\Gamma(T)| < |T| + \eta N\right] \le e^{-\kappa N} \tag{9}$$

for all sufficiently large N.

Proof. Choose $\eta > 0$ sufficiently small that

$$R + 2\eta < 1, \qquad \beta \log \frac{1}{2\eta} > (1 + R) \log 2 + 1. \tag{10}$$

Such a choice is possible because the left-hand side of the second inequality tends to infinity as $\eta \downarrow 0$.

Fix $s > \beta N/t$ and let $u = \lceil s + \eta N \rceil$. As in (6),

$$\mathbb{P}[\exists T : |T| = s,\ |\Gamma(T)| \le u] \le U_s := \binom{n}{s}\binom{N}{u}\left(\frac{u}{N}\right)^{ts}. \tag{11}$$

First suppose $s \le \eta N/2$. For all sufficiently large N,

$$\frac{u}{N} \le \frac{s}{N} + \eta + \frac{1}{N} \le 2\eta.$$

Using $\binom{n}{s} \le 2^n$, $\binom{N}{u} \le 2^N$, and $ts > \beta N$, we obtain

$$\begin{aligned} U_s &\le 2^{n+N}(2\eta)^{ts} \\ &\le \exp\left\{N\left((1+R)\log 2 - \beta \log \frac{1}{2\eta}\right)\right\} \\ &\le e^{-N} \end{aligned}$$

by (10).

Now suppose $s > \eta N/2$. Put $\rho := R + 2\eta < 1$. Since $s \le n \le RN$, for all sufficiently large N,

$$\frac{u}{N} \le \frac{s}{N} + \eta + \frac{1}{N} \le \rho.$$

Therefore

$$\begin{aligned} U_s &\le 2^{n+N}\rho^{ts} \\ &\le \exp\left\{N\left((1+R)\log 2 - \frac{\eta t}{2} \log \frac{1}{\rho}\right)\right\} \\ &\le e^{-N}, \end{aligned}$$

where the last inequality holds for all sufficiently large N because $t \to \infty$.

There are at most N possible values of s. Hence

$$\mathbb{P}\left[\exists T \subseteq [n] : |T| > \frac{\beta N}{t}, \quad |\Gamma(T)| < |T| + \eta N\right] \le N e^{-N} \le e^{-N/2}$$

for all sufficiently large N. Thus (9) holds, for example with $\kappa = 1/2$. □

4 Sparse Messages

The sparse regime combines support expansion with a field-independent compression of accumulator zero traces and an exact analysis of the gaps between occupied columns.

4.1 Marks, Positions, and Prefix Traces

Fix a message support $T \subseteq [n]$ of size w. Restrict the support graph to the rows in T, and suppose its union contains M occupied columns. List their actual positions as

$$1 \le p_1 < \cdots < p_M \le N, \qquad p_{M+1} := N + 1.$$

For each occupied column p_j, define its *mark* by

$$\mu_j := \big\{(i, E_{i,p_j}) : i \in T,\ E_{i,p_j} \neq 0\big\}.$$

Thus μ_j records which rows in T are incident to column p_j, together with their nonzero coefficients. For message variables $X = (X_i)_{i\in T}$, define

$$L_j(X) := \sum_{i\in T} E_{i,p_j} X_i, \qquad P_j(X) := \sum_{\ell=1}^{j} L_\ell(X). \tag{12}$$

Thus L_j is the expansion value at the jth occupied column and P_j is the accumulated value immediately after that column.

Order–support separation.

Lemma 3 (Order–support separation). *Conditional on M and on the multiset of the M nonempty column marks, the occupied set $\{p_1, \dots, p_M\}$ is a uniform M-subset of $[N]$. Conditional on that set, the marks appear in a uniform ordering. Consequently, the occupied set and the ordered mark sequence are independent.*

The same conclusion remains valid after conditioning on any event that depends only on the restricted incidence pattern and is invariant under permutations of the N column labels.

Proof. The restricted random matrix is invariant under every permutation of its N columns. Given M marked nonempty columns, placing them into $[N]$ is therefore a uniform injection. Every M-subset is the range of exactly $M!$ injections, and conditional on the range all assignments of the marks to the ordered locations are equally likely. Conditioning on a column-permutation-invariant event preserves equal probabilities within each orbit. □

Prefix-trace compression. For integers $H, w \ge 1$, define

$$\mathcal{A}(H, w) := \sum_{d=0}^{w-1} \binom{H}{d}. \tag{13}$$

For a nonzero witness $x \in \mathbb{F}_q^w$, let

$$Z_A(x) := \{j \in [M] : P_j(x) = 0\}$$

be its exact occupied-prefix zero trace.

Lemma 4 (Prefix-trace compression). *After the ordered marks and their coefficients are fixed, the number of distinct exact traces $Z_A(x)$ realized by nonzero $x \in \mathbb{F}_q^w$ is at most*

$$\mathcal{A}(M, w) = \sum_{d=0}^{w-1} \binom{M}{d}. \tag{14}$$

The bound allows zero forms, repeated forms, and arbitrary linear dependencies among the prefix forms.

Proof. Display the forms $P_1, \ldots, P_M$ in a fixed order. For a nonzero witness x, let U_x be the span of all displayed forms that vanish at x. Every such form lies in the hyperplane $x^\perp \subseteq (\mathbb{F}_q^w)^*$, so $\dim U_x \le w - 1$.

Choose canonically, for example lexicographically, a basis of U_x from the vanishing displayed forms. The chosen set has at most $w-1$ indices and determines U_x. A displayed form vanishes at x exactly when it belongs to U_x, so the canonical basis determines the entire zero trace. Distinct traces therefore receive distinct canonical labels, and the number of labels is at most (14). □

4.2 Random Gaps and Nonzero Prefixes

The accumulator gap kernel. Define the occupied gaps

$$\Delta_j := p_{j+1} - p_j, \qquad 1 \le j \le M.$$

Lemma 5 (EA gap identity). *For every witness x supported on T,*

$$\mathrm{wt}(xEA) = \sum_{j : P_j(x) \neq 0} \Delta_j. \tag{15}$$

Proof. Between coordinates p_j and $p_{j+1}-1$, the accumulated value is the constant $P_j(x)$. It is nonzero on that entire interval exactly when $P_j(x) \neq 0$. □

For $0 \le k \le M$ and $L = \lfloor \delta N \rfloor$, define

$$\mathcal{K}_{N,M,L}(k) := \begin{cases} 1, & k = 0, \\ \dfrac{1}{\binom{N}{M}} \displaystyle\sum_{h=k}^{L} \binom{h-1}{k-1}\binom{N-h}{M-k}, & 1 \le k \le M. \end{cases} \tag{16}$$

Lemma 6 (Exact random-gap kernel). *Let $C \subseteq [M]$ be fixed with $|C| = k$. If $\{p_1 < \cdots < p_M\}$ is a uniform M-subset of $[N]$, then*

$$\mathbb{P}\left[\sum_{j \in C} \Delta_j \le L\right] = \mathcal{K}_{N,M,L}(k). \tag{17}$$

Equivalently, if $X \sim \mathrm{Hyp}(N, L, M)$, then

$$\mathcal{K}_{N,M,L}(k) = \mathbb{P}[X \ge k]. \tag{18}$$

In particular, for $k \geq 1$,

$$\mathcal{K}_{N,M,L}(k) \leq \binom{M}{k}\left(\frac{L}{N}\right)^k \leq \left(\frac{eM\delta}{k}\right)^k. \tag{19}$$

Proof. Put $g_0 = p_1 - 1$ and $g_j = p_{j+1} - p_j - 1$ for $1 \leq j \leq M$. The vector $(g_0, \ldots, g_M)$ is a uniform weak composition of $N - M$. For a fixed k-set C, the selected gaps sum to $k + \sum_{j \in C} g_j$. Requiring this sum to equal h gives $\binom{h-1}{k-1}$ choices for the selected composition variables and $\binom{N-h}{M-k}$ choices for the remaining variables. Summing over $h \leq L$ proves (17). The hypergeometric form is the equivalent count obtained by asking that at least k of the M sampled positions lie in a fixed L-set.

For the final bound, if at least k sampled positions lie in the L-set, then some specified k-subset of the M labels lies there. A union bound gives $\binom{M}{k}(L/N)^k$, followed by $\binom{M}{k} \leq (eM/k)^k$. □

From nonzero expansion values to nonzero prefixes.

Lemma 7 (Nonzero expansion values force nonzero prefixes). *For any x, let*

$$r := \big|\{j \in [M] : L_j(x) \neq 0\}\big|.$$

Then at least $\lceil r/2 \rceil$ of the occupied-prefix values

$$P_1(x), \ldots, P_M(x)$$

are nonzero.

Proof. The claim is immediate if $r = 0$. Otherwise, let

$$1 \leq i_1 < \cdots < i_r \leq M$$

be the indices for which $L_{i_k}(x) \neq 0$, and define

$$Q_k := P_{i_k}(x), \qquad 1 \leq k \leq r.$$

Because i_1 is the first index with a nonzero expansion value,

$$Q_1 = P_{i_1}(x) = L_{i_1}(x) \neq 0.$$

Moreover, for every $2 \leq k \leq r$, the indices i_{k-1} and i_k are consecutive among the nonzero expansion values, so

$$Q_k - Q_{k-1} = \sum_{j=i_{k-1}+1}^{i_k} L_j(x) = L_{i_k}(x) \neq 0.$$

Thus two consecutive values in $Q_1, \ldots, Q_r$ cannot both be zero. Since $Q_1 \neq 0$, at most $\lfloor r/2 \rfloor$ of these values vanish, and hence at least $\lceil r/2 \rceil$ are nonzero. Each Q_k is one of the occupied-prefix values $P_1(x), \ldots, P_M(x)$, which proves the claim. □

4.3 Closing the Sparse Range

A fixed support.

Lemma 8 (Sparse fixed-support bound). *Fix $\varepsilon \in (0, 1/2)$ and let $c_0 = 1-2\varepsilon$. Let $T \subseteq [n]$ have size $w \le \beta N/t$, and condition on*

$$|\Gamma(T)| \ge (1-\varepsilon)tw. \tag{20}$$

If

$$\theta := \frac{2e\delta}{c_0} < 1, \tag{21}$$

then, for all sufficiently large N,

$$\mathbb{P}\big[\exists x \in \mathbb{F}_q^n : \operatorname{supp}(x) = T,\ \operatorname{wt}(xEA) \le \delta N \ \big|\ |\Gamma(T)| \ge (1-\varepsilon)tw\big] \le (2et)^w \theta^{c_0 tw/2}. \tag{22}$$

The bound is uniform in q and in the characteristic.

Proof. Put $M = |\Gamma(T)|$ and $L = \lfloor \delta N \rfloor$. By corollary 1, the induced graph has at least $c_0 tw$ unique neighbors. For every exact-support witness x, the expansion value $L_j(x)$ is nonzero at each corresponding column. Hence, by lemma 7, at least

$$k_0 := \lceil c_0 tw/2 \rceil$$

of the occupied-prefix values $P_1(x), \dots, P_M(x)$ are nonzero.

Fix the ordered mark sequence. By lemma 4, exact-support witnesses realize at most $\mathcal{A}(M, w)$ distinct prefix-zero traces, and every such trace has at least k_0 nonzero prefixes. Moreover, by lemma 3, conditional on the ordered marks and on (20), the occupied positions remain a uniform M-subset of $[N]$.

For a trace with $k \ge k_0$ nonzero prefixes, we have $M \le tw$ and

$$k \ge k_0 \ge \frac{c_0 tw}{2}, \qquad \text{so} \qquad \frac{M}{k} \le \frac{2}{c_0}.$$

Therefore, lemma 6 gives

$$\mathcal{K}_{N,M,L}(k) \le \left(\frac{eM\delta}{k}\right)^k \le \left(\frac{2e\delta}{c_0}\right)^k = \theta^k \le \theta^{k_0}.$$

It remains to bound the number of traces. Condition (20) gives

$$M \ge (1-\varepsilon)tw,$$

and hence $M \ge w$ for all sufficiently large N. If $w = 1$, then

$$\mathcal{A}(M, 1) = 1 \le 2eM.$$

Now suppose $w \ge 2$, and put $r = w - 1$ and $z = r/M \le 1$. Since $z^r \le z^d$ for $0 \le d \le r$,

$$z^r \mathcal{A}(M, w) \le \sum_{d=0}^{r} \binom{M}{d} z^d \le (1+z)^M \le e^{Mz} = e^r.$$

Consequently,

$$\mathcal{A}(M,w) \le \left(\frac{eM}{w-1}\right)^{w-1} \le \left(\frac{2eM}{w}\right)^{w} \le (2et)^w,$$

where the middle inequality uses $w-1 \ge w/2$ and $M \ge w$, while the final inequality uses $M \le tw$.

A union bound over the realizable traces now gives, conditional on the ordered marks,

$$\mathbb{P}[\exists x : \mathrm{supp}(x) = T,\ \mathrm{wt}(xEA) \le \delta N] \le \mathcal{A}(M,w)\theta^{k_0} \le (2et)^w \theta^{c_0 tw/2},$$

where the final step uses $\theta < 1$ and $k_0 \ge c_0 tw/2$. Averaging over the ordered marks proves (22). □

Union over sparse supports.

Proposition 1 (All sparse messages). *For every $C > 0$, after choosing γ sufficiently large,*

$$\mathbb{P}\left[\exists x \ne 0 : \mathrm{wt}(x) \le \frac{\beta N}{t}, \quad \mathrm{wt}(xEA) \le \delta N\right] \le N^{-C}. \tag{23}$$

The bound is uniform over all finite fields, including $q = q(N)$.

Proof. After increasing γ, lemma 1 shows that the probability of any support in the stated range violating (20) is at most N^{-C-1}. On the complementary event, take a union bound only over message supports and apply lemma 8:

$$\sum_{w=1}^{\lfloor \beta N/t \rfloor} \binom{n}{w} (2et)^w \theta^{c_0 tw/2} \le \sum_{w \ge 1} \left(2e^2 Nt\, \theta^{c_0 t/2}\right)^w.$$

Since $t = \lceil \gamma \log N \rceil$,

$$2e^2 Nt\, \theta^{c_0 t/2} \le N^{1-(\gamma c_0/2)\log(1/\theta)+o(1)}.$$

Choose γ so that $(\gamma c_0/2)\log(1/\theta) > C+3$. The geometric sum is then $O(N^{-C-1})$. □

5 High-Weight Messages

For message supports above the sparse threshold, saturated expansion provides a linear neighbor surplus. We exploit this surplus directly over large fields and use a stochastic accumulator analysis over the remaining bounded fields.

5.1 Large Fields

The large-field argument converts the additive neighbor surplus from lemma 2 into a surplus of independent prefix-zero equations.

Point-probability and rank bounds.

Lemma 9 (Maximum atom of a nonzero-uniform sum). *Let $U_1, \dots, U_d$ be independent uniform elements of $\mathbb{F}_q^*$, where $d \geq 1$. Then*

$$\max_{a \in \mathbb{F}_q} \mathbb{P}[U_1 + \cdots + U_d = a] \leq \frac{1}{Q}. \tag{24}$$

The same conclusion holds after multiplying each U_i by an arbitrary fixed nonzero scalar.

Proof. For $d = 1$ the assertion is immediate. If a random variable Y is independent of U_d, then, for every a,

$$\mathbb{P}[Y + U_d = a] = \mathbb{E}\left[\frac{\mathbf{1}_{\{a-Y \neq 0\}}}{Q}\right] \leq \frac{1}{Q}.$$

Induct on d. Nonzero scalar multiplication preserves uniformity on $\mathbb{F}_q^*$. □

Lemma 10 (Rank-r systems with bounded atoms). *Let $Z_1, \dots, Z_M$ be independent $\mathbb{F}_q$-valued random variables, each having every atom at most $1/Q$. If $H \in \mathbb{F}_q^{s \times M}$ has rank r, then, for every $b \in \mathbb{F}_q^s$,*

$$\mathbb{P}[HZ = b] \leq Q^{-r}. \tag{25}$$

Proof. Choose r pivot columns of H. Conditional on all nonpivot variables, the system determines at most one value of the r pivot variables. Independence and the atom bound give probability at most Q^{-r}. □

Prefix anti-cancellation.

Lemma 11 (Projective prefix anti-cancellation). *Fix a message support T of size w, an ordered occupied-column incidence pattern with M columns, and a prescribed set $V \subseteq [M]$ of occupied prefixes. Over the random nonzero coefficients,*

$$\mathbb{P}\big[\exists x \in (\mathbb{F}_q^*)^T : P_j(x) = 0 \text{ for all } j \in V\big] \leq \min\{1, Q^{w-1-|V|}\}. \tag{26}$$

Proof. Fix one projective message class on T. For every occupied column, $L_j(x)$ is a sum of at least one independent uniform nonzero coefficient multiplied by a fixed nonzero message value. The variables $L_1(x), \dots, L_M(x)$ are independent across columns and satisfy lemma 9.

The equations $P_j(x) = 0$ for $j \in V$ have coefficient rows

$$(\underbrace{1, \dots, 1}_{j}, 0, \dots, 0) \qquad (j \in V)$$

in the variables $L_1(x), \dots, L_M(x)$. Distinct such prefix rows are linearly independent, so lemma 10 gives probability at most $Q^{-|V|}$ for the fixed projective class. There are exactly Q^{w-1} projective classes in $(\mathbb{F}_q^*)^T$, and all equations are homogeneous in x. A union bound proves (26). □

Closing the large-field range.

Proposition 2 (High weights over large fields). *Fix R and β, let η be supplied by lemma 2, and choose $0 < \delta < \eta$. Put $\zeta = \eta - \delta > 0$. There exists an integer $Q_* = Q_*(R, \beta, \delta) \geq 2$ such that, for every field with $Q = q - 1 \geq Q_*$,*

$$\mathbb{P}\left[\exists x : \mathrm{wt}(x) > \frac{\beta N}{t}, \quad \mathrm{wt}(xEA) \leq \delta N\right] \leq e^{-\Omega(N)}. \tag{27}$$

The implicit constant is uniform over all $q \geq Q_* + 1$.

Proof. Except with probability $e^{-\Omega(N)}$, lemma 2 holds for every support in the stated range. Condition on such a support graph. Fix one support T, write $w = |T|$, and let $M = |\Gamma(T)| \geq w + \eta N$.

For a bad witness on T, let b be the number of occupied prefixes $P_j(x)$ that vanish. Every nonzero occupied prefix contributes at least one coordinate to the output by lemma 5. Hence

$$M - b \leq \mathrm{wt}(xEA) \leq \delta N,$$

so

$$b \geq M - \delta N \geq w + \zeta N. \tag{28}$$

For a prescribed exact zero trace V of size b, the event of realizing that trace is contained in the prescribed-zero event. By lemma 11, its probability is at most

$$Q^{w-1-b} \leq Q^{-\zeta N - 1}.$$

There are at most $2^M \leq 2^N$ possible prefix traces for T, and at most $2^n \leq 2^N$ message supports. Conditional on the saturated support graph, the total probability is therefore at most

$$4^N Q^{-\zeta N}. \tag{29}$$

Choose Q_* so that $\zeta \log Q_* > 3 \log 2$. Then (29) is exponentially small, uniformly for $Q \geq Q_*$. □

5.2 Bounded Fields

For the remaining bounded fields, the large-field equation-counting bound is too weak to pay for the union over all messages. We therefore fix a message and analyze the accumulator directly as a random process. The fixed-row model has two inconvenient dependencies: positions within each active row are sampled without replacement, and the total number of impulses is fixed. We first pass to a with-replacement model and then Poissonize the impulse count. These changes incur only subexponential costs, while Poisson splitting makes the expansion coordinates independent and identically distributed. The zero/nonzero status of their running sum is then an exact two-state Markov chain, whose occupation-time bound is strong enough to absorb the full q-ary message count.

Fix a message x of weight w and put

$$m := tw, \qquad \lambda := \frac{m}{N}.$$

Multiplication by the nonzero message entries preserves uniformity of active coefficients. Thus xE is the sum of m nonzero impulses, grouped into w blocks of t impulses whose positions are distinct within each block.

Comparison with the with-replacement model.

Lemma 12 (Fixed-row to with-replacement comparison). *Fix a nonzero message $x \in \mathbb{F}_q^n$ of weight w. Let P_{fr} be the fixed-row law of xEA, and let P_{wr} be its law when the t positions in each active row are sampled independently with replacement. Then, for every $\mathcal{B} \subseteq \mathbb{F}_q^N$,*

$$\mathsf{P}_{\mathrm{fr}}[xEA \in \mathcal{B}] \le d_{N,t}^{-w} \mathsf{P}_{\mathrm{wr}}[xEA \in \mathcal{B}], \qquad d_{N,t} := \prod_{j=0}^{t-1} \left(1 - \frac{j}{N}\right). \tag{30}$$

If $t \le N/2$, then

$$d_{N,t}^{-w} \le \exp\left(\frac{2wt^2}{N}\right). \tag{31}$$

Proof. Under P_{wr}, let $\mathcal{D}$ be the event that the t positions are distinct within each of the w active rows. Then

$$\mathsf{P}_{\mathrm{wr}}(\mathcal{D}) = d_{N,t}^{w},$$

and, conditional on $\mathcal{D}$, the with-replacement law is exactly the fixed-row law. Hence

$$\mathsf{P}_{\mathrm{fr}}[xEA \in \mathcal{B}] = \mathsf{P}_{\mathrm{wr}}[xEA \in \mathcal{B} \mid \mathcal{D}] \le d_{N,t}^{-w} \mathsf{P}_{\mathrm{wr}}[xEA \in \mathcal{B}].$$

Moreover, using

$$-\log(1-z) \le \frac{z}{1-z}, \qquad 0 \le z < 1,$$

we get

$$-\log d_{N,t} \le \sum_{j=0}^{t-1} \frac{j}{N-j} \le \frac{t^2}{N-t} \le \frac{2t^2}{N},$$

where the final inequality uses $t \le N/2$. Multiplying by w and exponentiating proves (31). □

Poissonization and the status chain. Let K be a Poisson random variable with mean m, written $K \sim \mathrm{Pois}(m)$. Generate K independent impulses, each with a uniform position in $[N]$ and a uniform nonzero coefficient. Conditional on $K = m$, this is the with-replacement model.

For each $j \in [N]$, let $Z_j \in \mathbb{F}_q$ be the sum of the coefficients of all impulses that land at position j, with $Z_j = 0$ if no impulse lands there. Thus $Z = (Z_1, \ldots, Z_N)$ is the Poissonized analogue of the vector xE, before applying the accumulator.

Recall that

$$\lambda = \frac{m}{N} = \frac{tw}{N}$$

is the mean number of impulses landing at each coordinate in the Poissonized model. Define

$$\vartheta_q(\lambda) := \exp\left(-\frac{q}{q-1}\lambda\right). \tag{32}$$

Lemma 13 (Compound-Poisson coordinate law). *The random variables $Z_1, \ldots, Z_N$ are independent and identically distributed, with*

$$a_q(\lambda) := \mathbb{P}[Z_j = 0] = \frac{1 + Q\vartheta_q(\lambda)}{q}, \qquad \mathbb{P}[Z_j = z] = \frac{1 - \vartheta_q(\lambda)}{q} \quad (z \in \mathbb{F}_q^*). \tag{33}$$

Proof. Poisson splitting gives independent $\mathrm{Pois}(\lambda)$ numbers of impulses at the N coordinates. For any nontrivial additive character χ of $\mathbb{F}_q$, a uniform $U \in \mathbb{F}_q^*$ satisfies $\mathbb{E}[\chi(U)] = -1/Q$. Therefore the compound-Poisson Fourier coefficient is

$$\exp\left\{\lambda\left(-\frac{1}{Q} - 1\right)\right\} = \vartheta_q(\lambda).$$

The law is invariant under multiplication by $\mathbb{F}_q^*$. If its probability at zero is a and its probability at each nonzero value is r, then

$$a + Qr = 1, \qquad a - r = \vartheta_q(\lambda).$$

Solving these equations gives (33). □

Let $Y_j = Z_1 + \cdots + Z_j$ and define the status process $S_j = \mathbf{1}_{\{Y_j \neq 0\}}$. It is a two-state Markov chain. With

$$c_q(\lambda) := \frac{Q + \vartheta_q(\lambda)}{q}, \tag{34}$$

its transition matrix, in states zero/nonzero, is

$$\begin{pmatrix} a_q(\lambda) & 1 - a_q(\lambda) \\ (1 - \vartheta_q(\lambda))/q & c_q(\lambda) \end{pmatrix}. \tag{35}$$

Every transition ending in state zero has probability at most $a_q(\lambda)$, and every transition ending in state nonzero has probability at most $c_q(\lambda)$.

For $0 < \delta < 1/2$, define

$$B_q(\lambda, \delta) := \exp(\mathsf{h}(\delta)) a_q(\lambda)^{1-\delta} c_q(\lambda)^{\delta}. \tag{36}$$

Lemma 14 (Poissonized accumulator occupation). *In the Poissonized model,*

$$\mathbb{P}[\mathrm{wt}(ZA) \leq \delta N] \leq (N+1) B_q(\lambda, \delta)^N. \tag{37}$$

Proof. For brevity, write $a_q = a_q(\lambda)$ and $c_q = c_q(\lambda)$. Fix a status sequence with exactly h nonzero states. Each transition ending in zero has probability at most a_q, and each transition ending in nonzero has probability at most c_q. Hence the sequence has probability at most $a_q^{N-h} c_q^h$. Since there are $\binom{N}{h}$ such sequences,

$$\mathbb{P}[\mathrm{wt}(ZA) \le \delta N] \le \sum_{h=0}^{\lfloor \delta N \rfloor} \binom{N}{h} a_q^{N-h} c_q^h.$$

Using $\binom{N}{h} \le \exp(N\mathsf{h}(h/N))$, define

$$f(r) := \mathsf{h}(r) + (1-r)\log a_q + r \log c_q.$$

For $0 < r \le \delta$,

$$f'(r) = \log \frac{1-r}{r} + \log \frac{c_q}{a_q} \ge 0,$$

because $\delta < 1/2$ and $c_q \ge a_q$. Thus f is increasing on $[0, \delta]$, and every summand is at most $\exp(Nf(\delta))$. Since there are at most $N+1$ summands,

$$\mathbb{P}[\mathrm{wt}(ZA) \le \delta N] \le (N+1)\exp(N\mathsf{h}(\delta)) a_q^{(1-\delta)N} c_q^{\delta N} = (N+1) B_q(\lambda, \delta)^N.$$

□

Proposition 3 (Fixed-message Markov bound). *For every fixed nonzero message x of weight w, with $m = tw$ and $\lambda = m/N$,*

$$\mathbb{P}_E[\mathrm{wt}(xEA) \le \delta N] \le 3(N+1)\sqrt{m}\, \exp\left(\frac{2wt^2}{N}\right) B_q(\lambda, \delta)^N. \tag{38}$$

Proof. For $K \sim \mathrm{Pois}(m)$, Stirling's inequality gives $\mathbb{P}[K = m] \ge 1/(3\sqrt{m})$. Hence the with-replacement probability is at most $3\sqrt{m}$ times the Poissonized probability. Apply lemmas 12 and 14. □

Entropy closure. The function $B_q(\lambda, \delta)$ decreases in λ: the parameter $\vartheta_q(\lambda)$ decreases, while both a_q and c_q are increasing affine functions of ϑ_q. Moreover,

$$\lim_{\lambda \to \infty} \log B_q(\lambda, \delta) = \mathsf{h}(\delta) + \delta \log(q-1) - \log q. \tag{39}$$

Lemma 15 (Bounded-field high-weight closure). *Fix $R \in (0,1)$, $\beta > 0$, and a finite integer $q_* \ge 2$. Suppose $\delta > 0$ satisfies*

$$\exp(\mathsf{h}(\delta)) \left(1 - \frac{1-e^{-\beta}}{2}\right)^{1-\delta} < 1, \tag{40}$$

$$\mathsf{h}(\delta) + \delta \log(q-1) < (1-R)\log q \quad \textit{for every prime power } 2 \le q \le q_*. \tag{41}$$

Then there exists $\kappa > 0$ such that, uniformly over all prime powers $q \le q_*$,

$$\mathbb{P}\left[\exists x : \mathrm{wt}(x) > \frac{\beta N}{t}, \quad \mathrm{wt}(xEA) \le \delta N\right] \le e^{-\kappa N} \tag{42}$$

for all sufficiently large N.

Proof. For $w > \beta N/t$, put $\alpha = w/N$ and $R_N = n/N \le R$, and define

$$\Phi_{N,q}(\alpha) := R_N \mathsf{h}(\alpha/R_N) + \alpha \log Q + \log B_q(t\alpha, \delta).$$

By proposition 3, the entropy bound

$$\binom{n}{w} \le \exp(NR_N \mathsf{h}(\alpha/R_N)),$$

and a union bound over the $\binom{n}{w}Q^w$ messages of weight w, their total contribution is at most

$$\exp\{N\Phi_{N,q}(\alpha) + o(N)\}. \tag{43}$$

The $o(N)$ term is uniform over w and $q \le q_*$. Indeed, $2wt^2/N \le 2Rt^2 = o(N)$, while the remaining prefactor $3(N+1)\sqrt{tw}$ is polynomial.

We claim that $\Phi_{N,q}(\alpha)$ is bounded above by a negative constant, uniformly over $q \le q_*$ and $\beta/t < \alpha \le R_N$. Suppose otherwise. Then there exist $N_j \to \infty$, admissible α_j, and prime powers $q_j \le q_*$ such that

$$\limsup_{j\to\infty} \Phi_{N_j,q_j}(\alpha_j) \ge 0.$$

Since only finitely many fields occur and $0 \le \alpha_j \le R$, after passing to a subsequence we may assume that $q_j = q$ is fixed and $\alpha_j \to \alpha \in [0, R]$. Write $t_j = t(N_j)$.

If $\alpha = 0$, then both message-count terms in $\Phi_{N_j,q}(\alpha_j)$ tend to zero. Since $t_j\alpha_j > \beta$ and B_q decreases in its first argument,

$$B_q(t_j\alpha_j, \delta) \le B_q(\beta, \delta).$$

Moreover,

$$a_q(\beta) = 1 - \frac{Q}{q}\big(1 - e^{-q\beta/Q}\big) \le 1 - \frac{1 - e^{-\beta}}{2}, \qquad c_q(\beta) \le 1.$$

Consequently,

$$B_q(\beta, \delta) \le \exp(\mathsf{h}(\delta)) \left(1 - \frac{1 - e^{-\beta}}{2}\right)^{1-\delta} < 1$$

by (40). Thus the above limit superior is strictly negative, a contradiction.

If $\alpha > 0$, then $t_j\alpha_j \to \infty$. Also,

$$\begin{aligned} R_{N_j}\mathsf{h}(\alpha_j/R_{N_j}) + \alpha_j \log Q &= R_{N_j}\left[\mathsf{h}(\alpha_j/R_{N_j}) + \frac{\alpha_j}{R_{N_j}} \log Q\right] \\ &\le R_{N_j} \log q, \end{aligned} \tag{44}$$

because $\mathsf{h}(p) + p \log Q \le \log(1 + Q) = \log q$ for $p \in [0, 1]$. Using $R_{N_j} \to R$ and (39), we obtain

$$\limsup_{j\to\infty} \Phi_{N_j,q}(\alpha_j) \le R \log q + \mathsf{h}(\delta) + \delta \log Q - \log q < 0$$

by (41), again a contradiction.

Thus $\Phi_{N,q}(\alpha)$ has a uniform negative margin. Absorbing the uniform $o(N)$ term in (43), each admissible weight contributes at most $e^{-\kappa_0 N}$ for some $\kappa_0 > 0$. Summing over at most N weights proves (42). □

6 Proof of the Main Theorem

Proof (Proof of theorem 2). Fix $R \in (0,1)$ and $C > 0$. Set

$$\varepsilon := \frac{1}{4}, \qquad c_0 := 1 - 2\varepsilon = \frac{1}{2},$$

and choose $\beta > 0$ as in lemma 1. Let $\eta > 0$ be supplied by lemma 2, decreasing it if necessary.

Choose $\delta > 0$ sufficiently small that

$$0 < \delta < \min\left\{\frac{1}{2}, \frac{\eta}{2}, \frac{c_0}{2e}\right\}, \tag{45}$$

$$\exp(\mathsf{h}(\delta))\left(1 - \frac{1 - e^{-\beta}}{2}\right)^{1-\delta} < 1, \tag{46}$$

$$\frac{\mathsf{h}(\delta)}{\log 2} + \delta < \frac{1-R}{2}. \tag{47}$$

Such a δ exists because the left-hand side of (46) tends to $1 - (1 - e^{-\beta})/2 < 1$, while the left-hand side of (47) tends to zero as $\delta \downarrow 0$. Condition (47) implies, for every $q \geq 2$,

$$\mathsf{h}(\delta) + \delta \log(q-1) \leq \left(\frac{\mathsf{h}(\delta)}{\log 2} + \delta\right) \log q < (1-R) \log q. \tag{48}$$

Put $\zeta = \eta - \delta > 0$ and choose an integer $Q_* \geq 2$ satisfying $\zeta \log Q_* > 3 \log 2$. Finally, choose γ sufficiently large that proposition 1 contributes at most N^{-C-1} and the exceptional probability in lemma 1 is also at most N^{-C-1}.

Split by message weight at $w_0 = \lfloor \beta N/t \rfloor$. The range $1 \leq w \leq w_0$ contributes at most N^{-C-1} by proposition 1, uniformly in q.

For $w > w_0$, split by field size. If $Q = q - 1 \geq Q_*$, apply proposition 2. If $Q < Q_*$, then $q \leq Q_*$ and the possible fields form a finite set. Apply lemma 15 with $q_* = Q_*$. Its hypotheses are exactly (46) and (48). In either case, the high-weight contribution is $e^{-\Omega(N)}$. For all sufficiently large N, the total failure probability is at most N^{-C}.

Finally, if $d_*(EA) > 0$, then EA has full row rank. Since A is invertible, E has row rank n, and the ordinary minimum distance of the generated code equals $d_*(EA)$. □

7 Optimality of the Polynomial Tail

The following elementary obstruction explains why theorem 2 has the correct qualitative tail scale for logarithmic row weight.

Proposition 4 (Terminal-row obstruction). *Fix $0 < \delta < 1$, let $L = \lfloor \delta N \rfloor$, and suppose $t = o(N)$ with $t \le L$. Then, uniformly over all finite fields,*

$$\mathbb{P}[d_*(EA) \le L] \ge 1 - \left(1 - \frac{\binom{L}{t}}{\binom{N}{t}}\right)^n. \tag{49}$$

In particular, for $t = \lceil \gamma \log N \rceil$,

$$\frac{\binom{L}{t}}{\binom{N}{t}} = N^{-\gamma \log(1/\delta) + o(1)}. \tag{50}$$

If $\gamma \log(1/\delta) > 1$, the right-hand side of (49) is $N^{1-\gamma \log(1/\delta)+o(1)}$.

Proof. Let $I = \{N - L + 1, \dots, N\}$. If the support of row i lies inside I, then the weight-one message e_i has $e_i E$ supported inside I. Prefix accumulation cannot create support before the first nonzero coordinate, and I extends to the final coordinate. Since A is invertible, $e_i EA$ is nonzero; hence it is a nonzero vector supported inside I and has weight at most L.

The row events are independent, and each occurs with probability $\binom{L}{t}/\binom{N}{t}$, proving (49). Since $t = O(\log N)$,

$$\log \frac{\binom{L}{t}}{\binom{N}{t}} = t \log \delta + O(t^2/N),$$

which gives (50). When the exponent exceeds one, the single-row probability is $o(1/N)$ and $1 - (1-p)^n = np(1 + o(1))$, yielding the final assertion. □

8 Conclusion

We resolved the fixed-row-weight distance conjecture of Block et al. by proving that a single expand–accumulate component with $t = \Theta(\log N)$ has constant relative distance with arbitrarily strong inverse-polynomial confidence. The result strengthens the conjecture by holding uniformly over all finite fields, including fields whose size varies with the block length.

The proof combines three mechanisms: projective trace compression and random-gap estimates for sparse messages, a surplus of prefix-zero equations over large fields, and a Poissonized Markov-chain analysis over bounded fields. A terminal-row obstruction shows that, for a fixed logarithmic row-weight constant, the inverse-polynomial failure scale is qualitatively optimal. Concrete finite-length parameter analysis will be treated separately.

References

1. Anubhav Baweja, Giacomo Fenzi, Pratyush Mishra, and Tushar Mopuri. Field-agnostic SNARKs with small proofs via encode-repeat-accumulate (ERA) codes. Cryptology ePrint Archive, Paper 2026/864, 2026.
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.
3. Alexander R. Block, Zhiyong Fang, Jonathan Katz, Justin Thaler, Hendrik Waldner, and Yupeng Zhang. Field-agnostic SNARKs from expand-accumulate codes. In *Advances in Cryptology–CRYPTO 2024*, LNCS 14929, pages 276–307, 2024.
4. Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch, and Peter Scholl. Correlated pseudorandomness from expand-accumulate codes. In *Advances in Cryptology–CRYPTO 2022*, LNCS 13508, pages 603–633, 2022.
5. R. W. R. Darling, Mathew D. Penrose, Andrew R. Wade, and Sandy L. Zabell. Rank deficiency in sparse random GF[2] matrices. arXiv:1211.5455, revised 2018.
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, 1998.
7. Jörg Kliewer, Kamil S. Zigangirov, Christian Koller, and Daniel J. Costello, Jr. Coding theorems for repeat multiple accumulate codes. arXiv:0810.3422, 2008.
8. Arya Mazumdar, Venkat Chandar, and Gregory W. Wornell. Update-efficiency and local repairability limits for capacity approaching codes. *IEEE Journal on Selected Areas in Communications*, 32(5):976–988, 2014.
9. Srinivasan Raghuraman, Peter Rindal, and Titouan Tanguy. Expand-convolute codes for pseudorandom correlation generators from LPN. In *Advances in Cryptology–CRYPTO 2023*, LNCS 14084, pages 602–632, 2023.