Concrete Minimum Distance Bounds of Accumulate Codes over any Fixed Finite Field

Alexander R. Block[1], Himanshu Dongre[1], and Hendrik Waldner[2]

[1]University of Illinois at Chicago. alexander.r.block@gmail.com, hdongr2@uic.edu
[2]Helmholtz Center for Information Security (CISPA). hendrik.waldner@cispa.de

October 31, 2026

Abstract

We present a new concrete unified distance analysis of three accumulate-type error-correcting codes over any finite field, specifically the Repeat-Accumulate (RA), Repeat-Accumulate-Accumulate (RAA), and Expand-Accumulate (EA) codes. For message length k, codeword length n, finite field size $q > 2$, and target relative minimum distance δ, we give new upper bounds on the probability that each of the above codes do not have minimum distance at least δ. Our bounds are concrete and explicit: they do not rely on n or q being sufficiently large. We present (multiple) closed-form upper bounds on the failure probability for each code, and complement these bounds with asymptotic analyses that (nearly) match prior-known asymptotic bounds. Our technical core is a new fine-grained union bound for accumulate codes, which may be of independent interest, that examines inputs to the accumulator from a purely distributional perspective, deriving properties that are necessary for an input vector to accumulate to a targeted output weight, and analyzing the probability the input distribution achieves these properties. This fine-grained union bound generalizes the input-output weight enumerator approach used by many prior works.

Our concrete analysis shows that (1) the RA code has concretely small distance (a maximum of $\delta \leqslant 2.4979 \times 10^{-8}$ for $k = 2^{40}$ and $q = 2^{32}$, except with probability 2^{-60}), and (2) the RAA and EA codes achieve somewhat large concrete distance for smaller field sizes $q \leqslant 2^{32}$. Concretely, for rate $k/n = 1/4$, $k \in \{2^{40}, 2^{50}\}$, and $q = 2^8$, the RAA code has distance $\delta = 1/8$ except with 2^{-60} probability, whereas the rate $1/2$ and $1/4$ EA code for the same k and $q = 2^4$, respectively, has distance $\delta \leqslant 0.0689756$ and $\delta \leqslant 0.232955$ except with probability 2^{-60}. For fields of size $q \geqslant 2^{32}$, under our concrete bounds, all codes experience a sharp drop-off in attainable minimum distance, as our analysis targets (small) fixed q.

Contents

1 Introduction **2**
- 1.1 Our Results . . . 4
  - 1.1.1 RA Code . . . 4
  - 1.1.2 RAA Code . . . 6
  - 1.1.3 EA Code . . . 7
  - 1.1.4 Concrete Parameters. . . . 9
  - 1.1.5 Limitations of Our Analysis. . . . 11
- 1.2 Related Work . . . 11
- 1.3 AI Disclosure . . . 14

2 Preliminaries **14**

3 The Minimum Distance of Encode-Accumulate Codes **15**
- 3.1 The Generalized Repeat-Accumulate Code . . . 17
  - 3.1.1 Binomial-Moment Strategy . . . 21
  - 3.1.2 Chernoff Strategy . . . 23
  - 3.1.3 Comparison Between Binomial-Moment and Chernoff . . . 25
- 3.2 The Generalized RAA Code . . . 26
  - 3.2.1 Binomial-Moment Strategy . . . 26
  - 3.2.2 Chernoff Strategy . . . 30
- 3.3 The Expand-Accumulate Code . . . 32
- 3.4 Extensions to Other Turbo-like Codes . . . 35

4 Asymptotic Analysis of the RA, RAA, and EA Codes **35**
- 4.1 RA Code . . . 35
- 4.2 RAA Code . . . 37
  - 4.2.1 BM Strategy . . . 37
  - 4.2.2 Chernoff Strategy. . . . 42
- 4.3 EA Code . . . 44

5 More Parameter Comparisons and Generation **55**
- 5.1 Concrete Parameter Comparisons . . . 55

6 Conclusions and Open Problems **58**

References **59**

1 Introduction

Error-correcting codes (or *codes*) are a fundamental object for building cryptographic primitives, such as zkSNARKs [AHI+17, BCG+17, BBH+18, BCR+19, ZXZ+20, BCG20, BCL22, XZS22, GLS+23, BBH+19] and pseudorandom correlation generators (PCGs) (used for building secure multi-party computation protocols [BCG+18, BCG+22, RRT23, KPR+26, BCG+20, CRR21]). Code-based primitives are desirable for post-quantum security, concrete efficiency, and modularity: many systems can swap the underlying code freely depending on the application, allowing for improvements to cryptographic primitives by either constructing a new code or improving the analyses of existing codes, rather than designing a new system from scratch.

Briefly, a $[n, k, \delta]_q$ linear code C is a k-dimensional subspace of $\mathbb{F}_q^n$ such that the Hamming weight of every $c \in C \setminus \{\mathbf{0}^n\}$ is at least δn, and is uniquely specified by a generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$; we also let $[n, k]_q$ denote a linear code with unspecified or unknown minimum distance. Here, $k \leqslant n \in \mathbb{N}$ are the message length and codeword length, respectively, and $\delta \in (0, 1)$ is the relative minimum distance. For many cryptographic applications, the parameters of interest of a code C are the rate $\rho := k/n$, the relative distance δ, and the encoding time T_{Enc}, i.e., the time to compute $\mathbf{xG}$ for any $\mathbf{x} \in \mathbb{F}_q^k$. For code-based cryptographic primitives, ρ and T_{Enc} heavily influence the concrete efficiency, as these primitives often encode vectors of size ranging from millions to billions of symbols, whereas δ heavily influences the security since larger distance provides more resilience to linear tests [Cou20] and provide guarantees about proximity to the code [ABF26]. Finally, many primitives want these efficiencies from the code C irrespective of the underlying finite field $\mathbb{F}_q$; that is, the code C should be *field-agnostic*.

Due to this modularity, there has been recent attention and focus [GLS+23, BCG+22, RRT23, BFK+24, ZCF24, GNR26, QGZ26, KPR+26] on constructing and analyzing *cryptographic-friendly codes*: field-agnostic codes C with *concretely efficient* T_{Enc} and $\rho = \Theta(1)$ (ideally, $T_{\mathsf{Enc}} = cn$ for small constant $c \geqslant 1$ and $\rho \geqslant 1/2$) and *concretely large* $\delta = \Theta(1)$. This is in contrast to classical coding theory, for which these guarantees most often hold *asymptotically* (i.e., for large enough k). Indeed, while many asymptotically optimal codes exist with linear encoding time and constant relative distance, such codes exhibit poor concrete performance: either computing $\mathbf{xG}$ is concretely expensive [Spi96, DI14], or the concrete minimum distance is small [GLS+23, QGZ26, GNR26]. In applications of interest, computing $\mathbf{xG}$ for large $\mathbf{x}$ often becomes the key cryptographic bottleneck impacting overall performance; moreover, concretely small δ also increases the number of cryptographic operations required, again impacting overall performance. This has led to recent focus on concretely efficient codes.

Prior Work. To the best of our knowledge, the push for constructing cryptographic-friendly codes, began with the work of Golovnev et al. [GLS+23], where they constructed a concretely efficient linear-time encodable code (building off of Spielman's expander code [Spi96]) with concretely large constant relative distance, colloquially known as the Brakedown code. In the short time since [GLS+23] first appeared, an explosion of follow-up works building cryptographic-friendly codes has emerged. Broadly speaking, these works fall into two broad categories:[1]

1. works which revisit classical coding theory constructions and analyze them from the cryptographic-friendly perspective [GLS+23, DI14, AZ26, HGQ+26, AHK+26, BCF+25]; and
2. works which design new cryptographic-friendly codes by leveraging new ideas or borrowing from classical coding theory works [BCG+22, RRT23, KPR+26, BFM+26, BCC+23, LLX+26, Zei26, QGZ26, GNR26].

[1] We restrict our attention to works where the main contribution is constructing new cryptographic-friendly codes, or a concrete analysis of (asymptotically good) codes.

Remark 1.1. Like the codes we consider in this work, *all* prior works we focus on construct *randomized* codes. However, some applications use Reed-Solomon codes [RS60], which are *nearly* cryptographic-friendly codes. The main drawback being $O(n \log n)$ encoding time over *specific* fields (i.e., FFT-friendly fields), which degrades to $O(n^2)$ time over non-FFT-friendly fields. While some techniques exist to extend efficient encoding (i.e., $O(n \log n)$ time) of RS codes to non-FFT-friendly fields [BCK+23, BCK+22], to the best of our knowledge, these methods are: (1) not concretely efficient, (2) have not been implemented or deployed in practice, and (3) do not remove the $O(n \log n)$ encoding time.

Limitations of Prior Work. Despite this explosion of new codes and analyses in the cryptographic-friendly setting, prior work is still limited in three key aspects.

Finite Field Restrictions. Many prior works enforce different restrictions on the underlying finite fields, such as linear growth in the codeword length or requirements to be sufficiently large [Kha26b, GLS+23], or restrictions to prime-order [LLX+26] or binary (extension) fields [BCF+25]. These restrictions manifest either in the context of maintaining strictly linear-time encoding, or are an artifact of the underlying distance analysis (in order to obtain a vanishing failure probability). As such, these codes are not truly field-agnostic.

Analysis Limitations. Many prior works show that the constructed code has large (i.e., constant) minimum distance except with vanishing probability (on the order of $1/\text{poly}(n)$) [BCF+25, BFK+24, AZ26]. This is done via the standard union bound, and the upper bound is obtained via an input partition: for *large weight* inputs, the union bound often collapses to some negligible function in n, whereas for *small weight* inputs, a looser upper bound of $1/\text{poly}(n)$ is obtained. From our perspective, this type of analysis has two main drawbacks.

First, when choosing parameters, the negligible term is often simply *ignored*. This is problematic for smaller input and field sizes, as a negligible function only becomes small enough to ignore for *large enough inputs*. In practice, the inputs are often large enough, but nonetheless we view this as a limitation since it may exclude applications with small input sizes or small finite fields.

Second, the inverse-poly upper bounds obtained are often quite loose, leading to poor concrete parameter choices. To remedy this, works often compute the full union bound directly for smaller input sizes (e.g., up to $k = 2^{20}$) to choose parameters, then extrapolate to larger input sizes (e.g., $k = 2^{40}$). We stress this is an inherent limitation of computing the full union bound, which is often $\Omega(k^2)$ expensive arithmetic operations. This is also problematic as there is an implicit assumption that the code distance does not degrade as the input size increases *for concrete values*, whereas this is often only the case *asymptotically*. Thus, the parameters obtained for actual applications are often heuristics at best.

Asymptotic Focus. Despite the overarching goal of constructing cryptographic-friendly codes, prior work also has extensively focused on ensuring the constructed code is *asymptotically good* (i.e., $\delta = \Theta(1)$ asymptotically). While work often complements the asymptotic result with concrete values of δ, it leaves open an interesting question: are there cryptographic-friendly codes that are not asymptotically good, but *concretely good*, meaning for particular settings of k, ρ, and $\mathbb{F}$, can we show that δ is concretely large?

This Work: Accumulate Codes over Finite Fields. We attempt to address all three of these limitations in the context of *accumulate codes*. Such codes generally have the form $\mathbf{G} = \mathbf{MA}$, where $\mathbf{M} \in \mathbb{F}^{k \times n}$ is some (possibly not asymptotically good) code with super-efficient encoding, and $\mathbf{A} \in \mathbb{F}^{n \times n}$ is the accumulator matrix, which encodes $\mathbf{y}$ as $\mathbf{z} = \mathbf{yA}$ such that $z_i = \sum_{j=1}^{i} y_j$. Concretely, we examine the *repeat-accumulate* (RA) [DJM98, BMS09], the *repeat-accumulate-accumulate* (RAA) [DJM98, BMS09, KZC07], and the *expand-accumulate* (EA) [BCG+22] codes from the perspective of cryptographic-friendliness. The RA and RAA codes have strictly linear-time encoding algorithms, whereas the EA code has quasi-linear encoding

time, but is concretely more efficient than performing FFTs. All three codes have been studied extensively over the binary field [BMS09, DJM98, BCG+22, GM08, KZC07], and also gained recent attention to extend them to the finite field case [BFK+24, Kha26b, Kha26a, AZ26, AHK+26, LXL+26].

In light of the outlined limitations, we ask the following questions for each of the RA, RAA, and EA codes:

Can we obtain strong concrete distance guarantees
over any *fixed set of parameters* $(k, \rho, \mathbb{F})$*?*

1.1 Our Results

In this work, for all considered codes, we obtain *closed-form concrete upper bounds* with respect to any set of fixed parameters $(k, \rho, \mathbb{F}, \delta)$. That is, given any set of parameters for each code, we give an explicit closed-form upper bound on the failure probability. Moreover, we also obtain asymptotic bounds on the failure probability from our concrete bounds which match the best-known asymptotic bounds for each of these codes (when applicable). Finally, we give explicit parameter calculations for each code, and, in particular, *do not need to extrapolate/conjecture* distance parameters continue to hold for larger input sizes; e.g., we are able to generate explicit values of δ for $k = 2^{60}$ (and beyond).

1.1.1 RA Code

We analyze the generalized RA code over any finite field $\mathbb{F}$, which defines the matrix $\mathbf{M}$ as the product of three matrices $\mathbf{RTQ}$, where $\mathbf{R} \in \{0,1\}^{k \times n}$ is the *repeat matrix*, which copies an input $\mathbf{x}$ $r := 1/\rho \in \mathbb{N}$ times, $\mathbf{T} \in \mathbb{F}^{n \times n}$ is a random invertible diagonal matrix, and $\mathbf{Q} \in \{0,1\}^{n \times n}$ is a random permutation matrix. We let $\mathcal{C}_{\mathrm{RA}}$ denote the ensemble of all possible generator matrices $\mathbf{G} = \mathbf{RTQA}$ of the generalized RA code. Though this code is not known to be asymptotically good in the binary case [DJM98, KZC07], it has a super-efficient linear-time encoding algorithm, so we study whether this code can be *concretely good* for specific sets of parameters.

First, we obtain two closed-form upper bounds on the minimum distance of a uniformly sampled code in $\mathcal{C}_{\mathrm{RA}}$ when given a target failure probability τ.

Theorem 1.1 (Concrete RA Code Failure Probability)**.** *Let $q \geqslant 3$ be a prime power, $k, r \in \mathbb{Z}^+$ such that $r \geqslant 3$. Let $\mathcal{C}_{\mathrm{RA}}$ be the ensemble of $[rk, k]_q$ generalized Repeat-Accumulate codes. Then, for target failure probability $\tau \in (0,1)$, we have*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}} [\delta_{\min}(C) \leqslant \delta] \leqslant \tau$$

whenever:

1. *(Binomial-Moment)*

$$\delta \leqslant \left[\left(\frac{\sqrt{q-1}}{q-2} \right) \left(\sqrt{1 + (q-2)\eta_{\mathsf{bm}}} - 1 \right) \right]^2 \quad \textit{for} \quad \eta_{\mathsf{bm}} = \left[\frac{(1 + (q-2)\tau)^{1/k} - 1}{q-1} \right]^{1/r};$$

2. *(Chernoff)*

$$\delta \leqslant \frac{1 - \sqrt{1 - \eta_{\mathsf{ch}}}}{2} \quad \textit{for} \quad \eta_{\mathsf{ch}} = \left[\frac{(1 + (q-1)\tau)^{1/k} - 1}{q-1} \right]^{2/r}.$$

Equivalently, for any $\delta \in (0,1)$:

1. *(Binomial Moment)*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}}[\delta_{\min}(C) \leqslant \delta] \leqslant \frac{(1+\xi_{\mathsf{bm}})^k - 1}{q-2} \quad \textit{for} \quad \xi_{\mathsf{bm}} = (q-1)\left(2\sqrt{\frac{\delta}{q-1}} + \left(1 - \frac{1}{q-1}\right)\delta\right)^r;$$

2. *(Chernoff)*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}}[\delta_{\min}(C) \leqslant \delta] \leqslant \frac{(1+\xi_{\mathsf{ch}})^k - 1}{q-1} \quad \textit{for} \quad \xi_{\mathsf{ch}} = (q-1)(4\delta(1-\delta))^{r/2}.$$

Here and throughout this introduction, $\delta_{\min}(C)$ denotes the relative minimum distance of the code C. Note that in the above theorem, the Chernoff strategy inherently has $\delta < 1/2$, whereas for the Binomial-Moment, we have a larger upper bound on δ. Note that the upper bounds of the first portion of the theorem do not contradict the second part: for any $\delta \in (0,1)$, if δ is too large, then the bound is larger than 1 and thus the statements become trivial. As a corollary, we derive sufficient conditions on the relative distance δ for inverse polynomial failure probability of the RA code.

Corollary 1.1 (Inverse Polynomial Failure Probability of the RA Code)**.** *Let $q \geqslant 3$ be a prime power, $r \in \mathbb{Z}^+$ such that $r \geqslant 3$, and $k \in \mathbb{Z}^+$ be sufficiently large. Let $\mathcal{C}_{\mathrm{RA}}$ be the $[rk, k]_q$ ensemble of generalized Repeat-Accumulate codes. Then, for any $\beta > 0$ and $n = rk$, we have*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}}[\delta_{\min}(C) \leqslant \delta] = O(n^{-\beta})$$

whenever:

1. *(Binomial-Moment)*

$$\delta \leqslant \frac{r^{2/r}}{4}\left(1 - \frac{1}{q-1}\right)^{2/r}(q-1)n^{-2(\beta+1)/r}(1+o(1));$$

2. *(Chernoff)*

$$\delta \leqslant \frac{r^{2/r}}{4}n^{-2(\beta+1)/r}(1+o(1))$$

The above corollary also immediately yields a sufficient condition on δ to ensure that the failure probability vanishes as n goes to infinity. We confirm that for fixed q, the ensemble $\mathcal{C}_{\mathrm{RA}}$ is *not* asymptotically good. To the best of our knowledge, this result has not appeared formally in prior work.

Corollary 1.2. *For fixed $r, q > 2$, $\lim_{n\to\infty} \Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}}[\delta_{\min}(C) \leqslant \delta] = 0$, whenever $\delta = o_{q,r}(n^{-2/r})$. In particular, for any fixed arbitrarily small $\epsilon > 0$, we have $\lim_{n\to\infty} \Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}}[\delta_{min}(C) > n^{-2/r-\epsilon}] = 1$.*

This corollary generalizes the asymptotic bounds on δ for the binary case [KZC07]; in particular, the conditions on δ are identical ($\delta = o_r(n^{-2/r})$).

It is also known for the binary case that $\delta_{\min}(\mathcal{C}_{\mathrm{RA}}) \leqslant O_r(n^{-1/r})$ [BMS09]. We conjecture the same holds for generalized $\mathcal{C}_{\mathrm{RA}}$ over any fixed finite field. In particular, we believe the proof of [BMS09] extends to the finite field case.

Conjecture 1.1. *For fixed $r, q \in \mathbb{Z}_{\geqslant 3}$, all sufficiently large n, and any C in the $[rk, k]_q$ ensemble $\mathcal{C}_{\mathrm{RA}}$, we have $\delta_{\min}(C) \leqslant O_{q,r}(n^{-1/r})$.*

1.1.2 RAA Code

Next, we move to the generalized RAA code, which defines matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{G}_{\mathrm{RA}}\mathbf{TQ}$, where $\mathbf{G}_{\mathrm{RA}}$ is the generator matrix of $C_{\mathrm{RA}} \xleftarrow{\$} \mathcal{C}_{\mathrm{RA}}$; i.e., in runs the generalized RA code, then randomly scales and permutes the output before accumulating again. Thus, the generator matrix of the RAA code is $\mathbf{G} = \mathbf{MA}$ for $\mathbf{M}$ defined above. Let $\mathcal{C}_{\mathrm{RAA}}$ denote the ensemble of all possible generator matrices of the generalized RAA code. This code also has a super-efficient linear-time encoding algorithm (though not as efficient as the RA code), and is known to be asymptotically good for $r \geqslant 3$ over the binary field [KZK+08, KZC07, PS03], and recently over (suitably large) finite fields [AZ26, Kha26b, HGQ+26].

We study if $\mathcal{C}_{\mathrm{RAA}}$ can be concretely good. Towards this goal, we derive two closed-form concrete upper bounds on probability that the code does not have some minimum distance. That is, for fixed (k, r, q, δ), we give two explicit and closed-form upper bounds. Our first bound, obtained using Binomial-Moment upper bounds, is what we believe to be the tighter of the two concrete bounds in most cases, but has the restriction that $\delta \leqslant (1/4)(1 - 2/r) < 1/4$ (for all $r \geqslant 3$).

Theorem 1.2 (Concrete RAA Code Failure Probability 1)**.** *Let $k \in \mathbb{Z}^+$, fix $q, r \in \mathbb{Z}_{\geqslant 3}$ such that q is a prime power, and fix $\delta \in (0, (1/4)(1 - 2/r)]$ such that $4\delta k \geqslant 1$. Let $\mathcal{C}_{\mathrm{RAA}}$ be the ensemble of $[rk, k]_q$ generalized Repeat-Accumulate-Accumulate codes. Define the following parameters:*

$$L_{k,\delta} := \lfloor 4\delta k \rfloor; \quad \Lambda_{q,\delta} := 2\sqrt{\frac{\delta}{q-1}} + \left(1 - \frac{1}{q-1}\right)\delta; \quad \theta_{q,\delta} := \frac{\Lambda_{q,\delta}}{1 - \Lambda_{q,\delta}}$$

$$\Phi_q(x) := \frac{x(q-2)}{2(q-1)}\left[1 + \sqrt{1 + \frac{4(q-1)}{x(q-2)^2}}\right]; \quad \mathrm{t}^{(\ell)}_{k,r,q,\delta} := \binom{k}{\ell}(q-1)^\ell \Phi_q\left(\frac{\ell}{k}\theta_{q,\delta}\right)^{r\ell}$$

$$R_{k,r,q,\delta} := \begin{cases} 1 & L_{k,\delta} = 1 \\ [\mathrm{t}^{(L_{k,\delta})}_{k,r,q,\delta} / \mathrm{t}^{(1)}_{k,r,q,\delta}]^{\frac{1}{L_{k,\delta}-1}} & L_{k,\delta} \geqslant 2 \end{cases}; \quad \mathscr{G}_{k,r,q,\delta} := \begin{cases} \frac{1 - R^{L_{k,\delta}}_{k,r,q,\delta}}{1 - R_{k,r,q,\delta}} & R_{k,r,q,\delta} \neq 1 \\ L_{k,\delta}, & R_{k,r,q,\delta} = 1 \end{cases}$$

If $\Lambda_{q,\delta} < 1$, then

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RAA}}}[\delta_{\min}(C) \leqslant \delta] \leqslant \left(\frac{q-1}{q-2}\right)^2 \cdot k \cdot \Phi_q\left(\frac{\theta_{q,\delta}}{k}\right)^r \cdot \mathscr{G}_{k,r,q,\delta}.$$

As with the RA code, the above theorem allows us to derive the following asymptotic bound for the RAA code.

Corollary 1.3. *For every fixed prime power $q \geqslant 3$, fixed $r \in \mathbb{Z}_{\geqslant 3}$, and fixed $\delta \in (0, (1/4)(1 - 2/r)]$, for all sufficiently large k, the $[rk, k]_q$ RAA code ensemble $\mathcal{C}_{\mathrm{RAA}}$ satisfies*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RAA}}}[\delta_{\min}(C) \leqslant \delta] = O_{r,q,\delta}(k^{1-r/2})(q-1)^{-r/2}.$$

Recent work due to Khabbazian [Kha26b] shows that this bound is essentially tight: asymptotically, as k *and* q grow, Khabbazian shows that the failure probability is $\Omega(k^{1-\lceil r/2 \rceil}(q-1)^{-\lfloor r/2 \rfloor})$ for constant $r \geqslant 4$ and $q = \Omega(n)$.

While we believe the above concrete bound is fairly tight, it imposes a strong restriction of $\delta < 1/4$, which may be undesirable for many applications. To remedy this, we next state a looser concrete failure bound on the RAA code, but with $\delta < 1/2$, which we obtain via tight Chernoff bounds.

Theorem 1.3 (Concrete RAA Code Failure Probability 2)**.** *Let $k \in \mathbb{Z}^+$, fix $q, r \in \mathbb{Z}_{\geqslant 3}$ such that q is a prime power, and fix $\delta \in (0, 1/2)$. Let $\mathcal{C}_{\mathrm{RAA}}$ denote the $[rk, k]_q$ ensemble of generalized Repeat-Accumulate-Accumulate codes. Define the following parameters:*

$$m_\delta := \min(1, 4\delta);\ U_{k,\delta} := \lfloor m_\delta k \rfloor;\ \Lambda_\delta := 2\sqrt{\delta(1-\delta)};\ \theta_{q,\delta} := \frac{\Lambda_{q,\delta}}{1-\Lambda_{q,\delta}}$$

$$\Phi_q(x) := \tfrac{x(q-2)}{2(q-1)}\left[1 + \sqrt{1 + \tfrac{4(q-1)}{x(q-2)^2}}\right];\ \phi_{r,q,\delta} := (q-1)\Phi_q(\theta_{q,\delta})^r$$

$$Q_{r,q,\delta} := e^{r/2} m_\delta^{r/2-1} \phi_{r,q,\delta};\ \mathcal{S}_{k,r,q,\delta} := \begin{cases} \frac{1-Q_{r,q,\delta}^{U_{k,\delta}}}{1-Q_{r,q,\delta}} & Q_{r,q,\delta} \neq 1 \\ U_{k,\delta} & Q_{r,q,\delta} = 1 \end{cases}.$$

Then,

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RAA}}}[\delta_{\min}(C) \leqslant \delta] \leqslant \frac{(k^{1-r/2}\phi_{r,q,\delta}\mathcal{S}_{k,r,q,\delta})}{(q-2)}.$$

This theorem allows us to obtain the same asymptotic bound on the failure probability of the RAA code (with respect to a fixed q), but for larger δ.

Corollary 1.4. *For every fixed prime power $q \geqslant 3$, fixed $r \in \mathbb{Z}_{\geqslant 3}$, and fixed $\delta \in (0, 1/2)$, for all sufficiently large k, the $[rk, k]_q$ RAA code ensemble $\mathcal{C}_{\mathrm{RAA}}$ satisfies*

$$\Pr_{C \xleftarrow{\$} \mathcal{C}_{\mathrm{RAA}}}[\delta_{\min}(C) \leqslant \delta] = O_{r,q,\delta}(k^{1-r/2})(q-1)^{-1}.$$

1.1.3 EA Code

Finally, we obtain results on the original Expand-Accumulate code introduced by Boyle et al. [BCG+22]. This code defines the matrix $\mathbf{M}$ as a matrix $\mathbf{E} \in \mathbb{F}^{k \times n}$ formed by sampling each entry i.i.d. from the generalized Bernoulli distribution $\mathsf{Ber}_p(\mathbb{F}_q)$, where $X \leftarrow \mathsf{Ber}_p(\mathbb{F}_q)$ is defined as uniform over $\mathbb{F}_q \setminus \{0\}$ with probability p and 0 with probability $1-p$. We let $\mathcal{C}_{\mathrm{EA}}(p)$ denote the ensemble of EA codes with generalized Bernoulli probability p.

Theorem 1.4 (Concrete EA Code Failure Probability)**.** *Let $k \in \mathbb{Z}^+$, and fix $q \in \mathbb{Z}_{\geqslant 3}$, $\rho \in (0, 1)$ such that $k/\rho \in \mathbb{N}$, and $\delta \in (0, 1)$ such that q is a prime power. Let $p \in (0, 1)$ and let $\mathcal{C}_{\mathrm{EA}}(p)$ denote the $[k/\rho, k]_q$ ensemble of Expand-Accumulate Codes with respect to Bernoulli distribution $\mathsf{Ber}_p(\mathbb{F}_q)$. Define the following parameters:*

$$\Lambda_{q,\delta} := 2\sqrt{\frac{\delta}{q-1}} + \left(1 - \frac{1}{q-1}\right)\delta; \quad \mu_{q,\delta} := \left(1 - \frac{1}{q}\right)(1 - \Lambda_{q,\delta})$$

$$\eta_{p,q} := \left|1 - \frac{pq}{q-1}\right|; \quad \nu_{p,q,\delta} := 1 - \mu_{q,\delta}(1 - \eta_{p,q}).$$

For $\eta_{p,q} > 0$, define

$$\sigma_{k,p,q} := \left\lceil \frac{\ln(k(q-1))}{\ln(1/\eta_{p,q})} \right\rceil \quad \theta_{k,p,q} := \left[\frac{1 + (q-1)\eta_{p,q}^{\sigma_{k,p,q}}}{q} \right]^{\frac{k}{\sigma_{k,p,q}}}.$$

Then, if $\Lambda_{q,\delta} < 1$, we have

$$\Pr_{C \xleftarrow{\$} C_{\mathrm{EA}}(p)}[\delta_{\min}(C) \leqslant \delta] \leqslant \frac{1}{q-2} \cdot \begin{cases} (q^k - 1) \cdot (1-\mu_{q,\delta})^n & \eta_{p,q} = 0 \\ q^k(1-\mu_{q,\delta}(1-\theta_{k,p,q}))^n - (1-\mu_{q,\delta})^n + ek(q-1)\nu_{p,q,\delta}^n & \eta_{p,q} > 0 \end{cases}.$$

Moreover, if $\sigma_{k,p,q} = 1$, then

$$\Pr_{C \xleftarrow{\$} C_{\mathrm{EA}}(p)}[\delta_{\min}(C) \leqslant \delta] \leqslant \frac{(q^k - 1) \cdot (1-\mu_{q,\delta})^n + ek(q-1)\nu_{p,q,\delta}^n}{q-2}.$$

Finally, the above results also hold for $\Lambda_{q,\delta} = 2\sqrt{\delta(1-\delta)}$.

We also show the asymptotic failure probability of the EA code. However, one undesirable feature of our analysis is that asymptotically, we must have $\delta \leqslant O(q^{-\rho})$ for vanishing failure probability.

Corollary 1.5. *For every fixed prime power $q \geqslant 3$, fixed $\rho \in (0,1)$, and fixed $\delta \in (0, O(q^{-\rho}))$, for all sufficiently large k and $p = \Omega(\ln(n)/n) \in (0,1)$ for $n = k/\rho$, the $[n,k]_q$ EA code ensemble $C_{\mathrm{EA}}(p)$ satisfies the following.*

1. *If $p = \Theta(1)$, then*
$$\Pr_{C \xleftarrow{\$} C_{\mathrm{EA}}(p)}[\delta_{\min}(C) \leqslant \delta] = \exp(-\Theta_{r,q,\delta}(n)).$$
2. *If $p = \omega(\ln(n)/n)$, then for every constant $\beta > 0$, we have*
$$\Pr_{C \xleftarrow{\$} C_{\mathrm{EA}}(p)}[\delta_{\min}(C) \leqslant \delta] = O_{r,q,\delta}(n^{-\beta}).$$
3. *If $p = \gamma\frac{\ln(n)}{n}(1+o(1))$ for constant $\gamma > 0$, then*
$$\Pr_{C \xleftarrow{\$} C_{\mathrm{EA}}(p)}[\delta_{\min}(C) \leqslant \delta] \leqslant n^{-(\gamma(1-\Lambda_{q,\delta})-1)+o(1)}$$
for $\Lambda_{q,\delta} = 2\sqrt{\frac{\delta}{q-1}} + \left(1 - \frac{1}{q-1}\right)\delta$.

The above corollary states that for $p = \Theta(\ln(n)/n)$, the failure probability is inverse polynomial, and that for $p = \omega(\ln(n)/n)$, the failure probability is negligible in n. In particular, our failure probabilities asymptotically match the failure probabilities for the binary EA code [BCG+22].

As stated previously, our analysis has an undesirable condition of $\delta = O(q^{-\rho})$. While it is not clear if this upper bound is necessary for $p \in (\Omega(\ln(n)/n), o(1))$ to show any vanishing failure probability, we do believe that δ should have no such upper bound for $p = \Theta(1)$. Our intuition for this is due to the following observation: for $p = (q-1)/q$, the expansion matrix $\mathbf{E}$ is now a uniformly random matrix over $\mathbb{F}_q^{k \times n}$. This implies that with high probability, the code should meet the Gilbert-Varshamov bound [Gil52, Var57], which gives an upper bound on δ of $1 - 1/q$, so long as $\rho > 1 - H_q(\delta)$, where H_q is the q-ary entropy function. In fact, recent concurrent work [Rin26] proves this to be the case: for $p = \Theta(\ln(n)/n)$, it is shown that except with inverse polynomial failure probability, the EA code approaches the GV bound; see Section 1.2 for more discussion.

Moreover, our analysis says nothing asymptotically for $p = o(\ln(n)/n)$. For $p = O(1/n)$, which would yield a linear-time encodable code with high probability, our analysis is inconclusive.

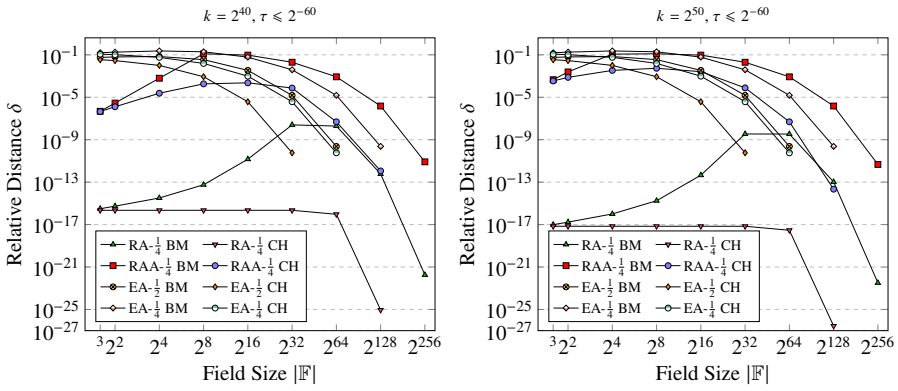

Figure 1: A sampling of relative distance calculations based on Theorems 1.1 to 1.4. k is the message length, τ is the targeted failure probability, the suffix "-ρ X" indicates the rate of the code analyzed and which strategy is used to compute the failure probability (Binomial-Moment or Chernoff). For the EA code, we plot $p = {}^{20\log(n)}/_{n}$ for $n = k/\rho$. Omitted RAA and EA data points indicate our script failed to find an admissible δ. For both RA-$\frac{1}{4}$ CH plots, $|\mathbb{F}| = 2^{256}$ is omitted since $\delta \leqslant 10^{-45}$.

1.1.4 Concrete Parameters.

We showcase the strengths of our concrete upper bounds by explicitly computing parameters for *large instance sizes*. When possible, we compare the generated parameters to prior work's *explicitly computed* parameters, not ones extrapolated due to prohibitive computation time or poor upper bound estimates. Though our computed distances may be worse than prior works', there are no assumptions made in the computation. Our scripts and generated data presented in this paper can be found at https://github.com/alexander-r-block/ra-raa-ea-parameter-gen/.

Our scripts implement the upper bounds on the failure probability given by Theorems 1.1 to 1.4. For the RA code, under Theorem 1.1, we directly calculate the maximum possible δ for a given failure probability τ and parameters r, k, q. For the RAA and EA codes, we implement a binary search within the interval $(0, \delta^\star)$ to find δ such that we are at or under the targeted failure probability, where $\delta^\star$ is a suitable upper bound on δ as given by our theorems. Our binary search then executes 50 iterations to find the largest δ admissible under target failure probability τ. Making this parameter larger gives a more accurate estimate of the largest possible admissible δ for a given set of parameters.

We ran all scripts for Theorems 1.1 to 1.4 (6 bounds total) on approximately 2200 sets of parameters: target failure probabilities 2^{-i} for $i \in \{16, 20, 40, 60, 80, 100\}$, message length $k = 2^i$ for $i \in \{5, 10, 15, \ldots, 60\}$, rate $\rho \in \{1/2, 1/4, 1/8, 1/16\}$ (or $1/3$ instead of $1/2$ for the RA and RAA codes), and $q = 3$ and $q = 2^i$ for $i \in \{2, 4, 8, 12, 16, 32, 64, 128, 256\}$. For the EA code, we discuss choices for the Bernoulli probability p below. All scripts are implemented in SageMath, and took at total of less than 2 minutes of total wall clock time to execute in SageMath 10.9 on Arch Linux on an AMD Ryzen 7 9800X3D using 16 cores.

We present a sampling of our concrete distance calculations using Theorems 1.1 to 1.4 in Figure 1. We examine target failure probability $\tau \leqslant 2^{-60}$ for message lengths $k = 2^{40}, 2^{50}$. Our choice of failure probability 2^{-60} is somewhat arbitrary, as many works in the literature target larger failure probability (e.g., 2^{-16} or 2^{-20}), or even smaller failure probability (e.g., 2^{-100}). We believe 2^{-60} is a strong middle ground since this is a statistical error for the code to not have minimum distance at least δ. However, within our script repository (see here), we include all generated parameters for all codes (as `.csv` files), under the assumptions stated previously.

$\log(k)$	$-\log(\text{fail prob})$			
	Ours $\delta = 0.02$	[AHK+26] $\delta = 0.05$	[BCF+25] $\delta = 0.2$	[AZ26] $\delta = 0.5$
15	40	16	7	11
20	60	20	10	20
25	80	24	13	28

Table 1: Comparison of failure probabilities for the RAA code with $r = 4$. For our parameters, we select $q = 2^{32}$, whereas for [BCF+25, AZ26], the parameters hold for $q \geqslant 2^{32}$. For [AHK+26], all parameters presented are extrapolated and not directly computed.

Let BM and CH denote Binomial-Moment and Chernoff, respectively, and let $q = |\mathbb{F}|$. For the RA code, as expected the distance calculations here are (more often than not) the worst of all tested parameters. Interestingly, both the BM strategy and the CH strategy experience sharp drop-offs for $q > 2^{64}$. Perhaps more interestingly, under the BH strategy, the RA code has larger minimum distance than rate $\frac{1}{2}$ EA code (under both strategies).

For the RAA code, under the BM strategy, a maximum of $\delta = 1/8$ is achieved at $q = 2^8$. Similar to the RA code, starting from $q = 2^{16}$, the distance of the RAA code under both the BM and CH strategies begins to degrade as q becomes large. The RAA code under the BM strategy has the largest δ for all $q \geqslant 2^{16}$.

For the EA code, we set $p \in \{\frac{1}{4}, \frac{1}{8}, \frac{5\log^2(n)}{n}, \frac{20\log(n)}{n}, \frac{32}{n}\}$ for the Bernoulli probability. As expected from our analysis, most parameters with $p = \frac{32}{n}$ were inadmissible. For Figure 1, interestingly, all computed δ for each of the remaining $p \neq \frac{32}{n}$ were nearly identical. A maximum of $\delta = 0.2329553292$ was computed for the rate $\frac{1}{4}$ EA code under the BM strategy for $k = 2^{50}$ and $q = 2^4$ (but for $k = 2^{40}$, the computed δ agrees for the first 9 digits), and a max. Finally, we remark that our numerical calculations reveal that we may need to perform a different kind of search for the EA code because for nearly all computed δ for $q \geqslant 2^{32}$, the failure probability is $\ll 2^{-100}$. For example, for the rate 1/2 EA code, the failure probability is 2^{-202} at $q = 2^{32}$, and then 2^{-1141} at $q = 2^{64}$.

See Figure 2 in Section 5 for more (q, δ) tradeoffs at $k \in \{2^{10}, 2^{20}, 2^{30}, 2^{60}\}$ under the same settings as above.

Comparison to Prior Concrete RAA Analyses. Compared with prior work, our analysis is still lacking. In Table 1, we compare our parameters to those of Brehm et al. [BCF+25] and Akhiani and Zhang [AZ26]. We take the parameters from [AZ26, Table 3] for comparison, but remark that: (1) [BCF+25] obtain stronger smaller failure probabilities for $\delta = 0.19$ under similar k and also provide a Mathematica script to validate parameters, and (2) [AZ26] do not provide the scripts they use to generate these parameters (as of this writing). We also compare with Abbaszadeh et al. [AHK+26] for completeness, but remark that all parameters that appear in Table 1 from Abbaszadeh et al. are extrapolated and not directly computed. When comparing to these works, our distance is an order of magnitude smaller than Brehm et al., but our failure probabilities are several orders of magnitude smaller (i.e., better).

We also briefly compare with Huang et al. [HGQ+26], who also give a concrete RAA code analysis. They obtain $\delta = 0.3$ at $r = 4$, $k = 2^{20}$, and $q = 2^{16} + 1$, with a failure probability of $\approx 2^{-37}$. Under the same parameters, we obtain $\delta = 0.09415$ with failure probability 2^{-40}. At the time of writing, [HGQ+26] do not provide scripts for computing their parameters.

We remark that all above prior works take a different approach their analyses than our work; see Section 3 for more details. Due to space constraints, we provide a comparison with non-RAA codes in Section 5.1.

1.1.5 Limitations of Our Analysis.

Our bounds represent a tradeoff: we obtain explicit, concrete upper bounds on the failure probabilities of the RA, RAA, and EA code. But, these bounds are worse than what prior work has obtained. This difference is somewhat fundamental: the approximations we make in our analysis (see Section 3 for complete details) allow us to leverage several combinatorial identities (e.g., the Binomial Theorem, negative Binomial generating function identity, etc.) to collapse large summations into easily computable forms. This relaxation in the failure probability bound allows us to explicitly compute large instance sizes very quickly, up to $k = 2^{60}$, at the cost of worse concrete distance δ. However, we believe we can tighten our bounds by combining our techniques with the techniques of prior work, and close the gap between upper bound tightness and efficiency of computing parameters.

1.2 Related Work

Here, we focus on minimum-distance bounds for fast linear codes relevant to the Ligero–Brakedown framework [AHI+17, GLS+23]. Additionally, we present an asymptotic overview of various codes in Table 2.

Turbo-like codes. Following turbo codes [BGT93, BM96b, BM96a], Divsalar, Jin, and McEliece [DJM98] studied repetition–interleaving–accumulation constructions. RA and RAA codes are repeat-multiple-accumulate (RMA) codes with one and two random interleavers and accumulators, respectively. The repetition factor for these codes is fixed as r, which gives $n = rk$ and $T_{\text{enc}} = O(n)$. The Binary RA code has $\delta = o(1)$ for constant r [DJM98, KZC07, KZK+08, BMS09]. Yang [Yan04] introduced weighted nonbinary RA; Kim, Cheun, and Lim [KCL11] studied its decoding performance via an approximate accumulator weight enumerator.

Additional accumulators (and random interleavers) yield binary ensembles with $\delta = \Theta(1)$ and tradeoffs approaching the GV bound as the number of accumulators increases [Pfi03, KZC07, KZK+08, RF09]; in particular, binary RAA code has constant relative distance for $\rho \leqslant 1/3$ [KZK+08].

Blaze [BCF+25] uses the RAA code in a polynomial commitment scheme [KZG10], with asymptotic and concrete bounds over characteristic two: after setup tests, $(k, \rho, \delta) = (2^{22}, 1/4, 0.19)$ with $\tau \lesssim 2^{-27.4}$. The accumulator application requires $O(n)$ additions. Generalized RAA also supports blind commitments over prime fields [HGQ+26]. Akhiani and Zhang [AZ26] strengthen the large-field bounds, reporting a partial failure bound of 2^{-20} at $(k, r, \delta) = (2^{20}, 4, 1/2)$.

For fixed $r \geqslant 4$, Khabbazian [Kha26b] proves $\tau = \Theta(n^{1-r})$ for generalized RAA code when $\delta \in (0, 1/2]$ and $q = \Omega(n)$. Without scaling (i.e., just permuting), $\tau = \Theta(n^{2-r})$ for $\text{char}(\mathbb{F}_q) > r$ and $\delta \in (0, 1/10]$. These exponents are asymptotically tight in the stated regimes; their tightness for concrete parameters remains open.

Expander-based codes and sketched codes. Spielman [Spi96] constructed asymptotically optimal constant-rate binary codes with $\delta = \Theta(1)$ and $O(n)$ encoding time via expander graphs. Druk–Ishai [DI14] improve this construction over binary and achieve minimum distance approaching the GV bound while maintaining $O(n)$ encoding time. Brakedown [GLS+23] further improves these constructions and extends them to any (suitably large) finite field, while giving a practically-efficient linear-time encoder. Haböck [Hab23] improves Brakedown's analysis, and further combines the Spielman and Brakedown analyses to improve the concrete encoding time and extend Brakedown's codes to smaller fields (as small as 32-bits).

Lightning [QGZ26] and Bolt [GNR26] introduce sketched codes, which is essentially a simplification of Brakedown's code that first obtains a shorter "sketch" of the input $\mathbf{x}$ then expands this sketch using a concretely

Code	T_{enc}	$\delta = d/n$	Field and Conditions
Reed–Solomon [RS60]	$O(n \log n)$	$1 - \rho + 1/n$	$\mathbb{F}_q$, $n \leqslant q$, FFT-friendly
Brakedown [GLS+23, Hab23]	$O(n)$	$\Omega(1)$	Fixed $\mathbb{F}_q$, $q \geqslant 2^{31}$
RA [DJM98, BMS09]	$O(n)$	$O(n^{-\rho})$	$\mathbb{F}_2$; $\rho \leqslant 1/3$
RAA / Blaze [KZK+08, BCF+25]	$O(n)$	$\Omega(1)$	$\mathbb{F}_{2^m}$; $\rho \leqslant 1/4$
Scaled RAA [AZ26, Kha26b]*, § 1.1	$O(n)$	$\Omega(1)$	$\mathbb{F}_q$, $\rho \leqslant 1/4$
Bernoulli EA [BCG+22]	$O(n \log n)^\dagger$	$\Omega(1)$	$\mathbb{F}_2$
Juxtaposed EA [BFK+24]‡	$O(n \log n)^\dagger$	$\Omega(1)$	$\mathbb{F}_q$
Fixed-weight EA [Kha26a]	$O(n \log n)$	$\Omega(1)$	$\mathbb{F}_q$
Flare Fixed-weight EA [LXL+26]⋆	$O(n \log n)$	$\geqslant 1 - \rho - \varepsilon$	$\mathbb{F}_q$, $q = \omega(n)$
Bernoulli and Fixed-weight EA/EC [Rin26]⋆	$O(n \log n)^\dagger$	$\geqslant 1 - \rho - \varepsilon$	$p = \Theta(\ln(n)/n)$ for Bernoulli
Block-accumulate [KPR+26]§	$O(n)$	$\Omega(1)$	$\mathbb{F}_2$
ERA [BFM+26]	$O(n)$¶	$\Omega(1)$	$\mathbb{F}_q$
Chosen-Block Block-accumulate [PR26]§	$O(n)$	$\Omega(1)$	$\mathbb{F}_2$
Chosen-Block Block-accumulate [Kha26c]‖	$O(n)$	$\Omega(1)$	$\mathbb{F}_q$, $q = \omega(n)$

Table 2: Asymptotic code guarantees. All of the above guarantees hold unconditionally, or have failure probabilities which approach 0 as $n \to \infty$.
* Requires $q \geqslant 2^{31}$ [AZ26] or $q = \Omega(n)$ [Kha26b].
† Expected encoding costs over the Bernoulli sampling.
‡ Concatenates fixed-weight and Bernoulli EA encodings of the same message, halving their common rate from $R < 1$ to $\rho = R/2 < 1/2$.
⋆ Assumes fixed gap $0 < \varepsilon < 1 - \rho$.
§ Assumes fixed local blocks with minimum distance $\delta_{\text{loc}} n_{\text{loc}} > 4$.
¶ Encoding includes a linear-time base code; assumes RAA rate $\rho \leqslant 1/4$.
‖ Assumes fixed local blocks with minimum distance $\delta_{\text{loc}} n_{\text{loc}} \geqslant 2$.

efficient code with large minimum distance. Overall, $\delta = \Theta(1)$ is achieved with linear-time encoding, though the concrete values of δ are small (see Section 5.1 for details).

Expand-accumulate (EA) codes. Boyle et al. [BCG+22] introduce binary EA codes for applications to PCGs. They prove binary $\delta = \Theta(1)$ for Bernoulli sampling with probability $p = \Omega(\log(n)/n)$, while conjecturing that the exact-weight variants (i.e., uniformly sample each row of the expansion matrix from the set of all weight t vectors in $\mathbb{F}_q^n$) also satisfy $\delta = \Theta(1)$. For fixed or expected (under Bernoulli sampling) row weight t, the encoding time of these codes is $T_{\text{enc}} = O(kt + n)$; $t = \Theta(\log n)$ gives $O(n \log n)$ encoding time.

Block et al. [BFK+24] derive an alternative $\mathbb{F}_q$ accumulator input–output weight enumerator (originally given in [RG11]) and prove $\delta = \Theta(1)$ for juxtaposed fixed-weight/Bernoulli EA components (i.e., $\mathbf{x} \mapsto (\mathbf{x}\mathbf{E}_1, \mathbf{x}\mathbf{E}_2)$, where $\mathbf{E}_1$ is the fixed-weight EA code and $\mathbf{E}_2$ is the Bernoulli EA code), giving field-agnostic SNARKs in the framework of Brakedown [GLS+23] with $O(n \log n)$ encoding time.

$\Pr[\delta < \delta_\rho n] \leqslant n^{-C}$ for every fixed $\rho \in (0, 1)$ and $C > 0$, using $t = \lceil \gamma(\rho, C) \log n \rceil$ and sufficiently large n. Li et al. [LXL+26] give another analysis of the fixed-row-weight EA code which approaches the Singleton bound: $\delta \geqslant 1 - \rho - \varepsilon$ for fixed $0 < \varepsilon < 1 - \rho$, $\log q = \omega(\log n)$, and row weight $t = \lceil \gamma \log n \rceil$ with a suitable constant γ.

In concurrent work, Rindal [Rin26] shows that asymptotically, the EA code approaches the GV bound for $p = \Theta(\ln(n)/n)$, both for binary and q-ary EA codes. Additionally, they give a similar asymptotic analysis for Expand-Convolute codes [RRT23], as well as exact/fixed-weight variants (e.g., a left-regular bipartite

expander). In both cases, the result extracts a common exact input-output enumerator framework for sparse random maps, similar to our framework.

Expand-convolute codes. Raghuraman, Rindal, and Tanguy introduce Expand-Convolute codes, borrowing ideas from general turbo-like codes, where the accumulator is instead replaced with a random convolution matrix, which encodes a vector $\mathbf{y}$ as $\mathbf{z} = \mathbf{y}\mathbf{C}$ such that $z_i = \sum_{j \in S_i} y_j$, where $S_i \subseteq [i]$ is a random subset of size at most m such that $i \in S_i$. They prove that for constant rate over the binary field, these codes satisfy $\delta = \Theta(1)$ for $m = O(\log(n))$, giving overall $O(n \log(n))$ encoding time.

Block-accumulate codes. Kolesnikov et al. [KPR+26] introduce binary block-accumulate codes for PCGs and zero-knowledge proofs, proving $\delta = \Theta(n)$ for an ensemble whose fixed-size local blocks have minimum distance greater than four. Such codes are defined via a block-diagonal matrix $\mathbf{G}$ such that each block is the generator matrix of some code with some guaranteed minimum distance (e.g., for block size $m_1 \times m_2$, each block code has distance at least $\delta m_2 \geqslant 4$). Then, $\mathbf{x} \mapsto \mathbf{x}\mathbf{G}\mathbf{M}$, where $\mathbf{M}$ is the generator matrix of the RAA (or, more specifically, the RMA) code. For constant block sizes, they achieve $T_{\text{enc}} = O(n)$, and their codes admit extremely efficient GPU implementations.

More recently, Peceny and Rindal [PR26] introduce and analyze the so-called Chosen-Block Block-Accumulate-Accumulate (BAA) codes, which replace the random block diagonals of Kolesnikov et al. [KPR+26] with a fixed linear code (e.g., Reed-Muller, BCH, or Golay codes for the binary case, Reed-Solomon codes for the q-ary case), and prove that these codes are asymptotically good (i.e., $\delta = \Omega(1)$) while maintaining $O(n)$ encoding time. Numerical calculations show that concretely, these codes are very close to the GV bound for certain sets of parameters. For large finite fields, Khabbazian [Kha26c] proves asymptotically that for large enough $q = \Omega(n)$ and every $d \geqslant 2$, there is a sufficiently small constant $\delta > 0$ such that the Chosen-Block BAA code over $\mathbb{F}_q$ has minimum distance δ except with probability $\Theta(n^{1-d} + n^{1-\lceil d/2 \rceil}(q-1)^{-\lfloor d/2 \rfloor})$, where d is the minimum distance of the chosen block diagonal code.

Encode-repeat-accumulate (ERA) codes. Baweja et al. [BFM+26] propose ERA codes, which first encode a message $\mathbf{x}$ using some outer code $\mathbf{C}$ with large rate and constant minimum distance, then encode the output with the RAA code. Key to their analysis is that with the generalized RAA encoder, they can share the random diagonal matrix $\mathbf{T}$ across accumulation steps, needing only independent permutation matrices. The total cost is $T_{\text{enc}} = T_0 + O(n) = O(n)$ for their linear-time initial encoder of cost T_0, and asymptotically their codes have $\delta = \Theta(1)$ for any constant-rate over any finite field.

Reed–Solomon and foldable codes. Reed–Solomon codes [RS60] have $d = n - k + 1$ and $T_{\text{enc}} = O(n \log n)$ on suitable FFT domains. FRI [BBH+18] and DEEP-FRI [BGK+20] improve proximity testing, leaving d unchanged. BaseFold [ZCF24] extends FRI to foldable linear codes and introduces random foldable (punctured Reed–Muller) codes with $T_{\text{enc}} = O(n \log n)$ over sufficiently large fields. Their distance bound depends on the base code, recursion depth, q, and τ; the full version [ZCF23, Table 1] reports $\delta_0 \approx 0.5044$ at $(k, \rho, k_0) = (2^{20}, 1/16, 2^5)$ over a roughly 31-bit field, with setup-failure probability $\tau \leqslant 2^{-128}$, where k_0 is the base dimension. This bound concerns the final recursively constructed code; its base-code distance does not carry over directly. BrakingBase [NST25] combines $O(k)$ commitment with polylogarithmic verification over sufficiently large fields; LCH-basis work [BSG26] improves polynomial arithmetic in characteristic two. Their contributions concern protocol and arithmetic costs, respectively.

Quasi-abelian codes. Bombar et al. [BCC+23] use QA codes for PCGs assuming the hardness of decoding (random) QA codes. Li et al. [LLX+26] give concrete distance bounds for rank-one QA codes over $\mathbb{F}_p[\mathbb{Z}_2^s]$, for sufficiently large primes p, and instantiate the Brakedown-PCS with these QA codes. Here $k = 2^s$, $n = ck$, and $\rho = 1/c$. Concretely, they prove that a random (systematic) $[n, k]_p$ QA code has at least δ with high probability, so long as $1 - H_p(\delta) - \frac{1+\log_p(k)}{c} > 0$, where H_p denotes the p-ary entropy function. In particular, by the GV bound, it is known that whenever $\delta \in (0, 1 - 1/q)$ codes of distance δ and rate $\rho > 1 - H_p(\delta)$ exist, showing that random QA codes are only $\log_p(k)/c$-far from the GV bound.

The QA code also has a nearly linear-time encoding algorithm via the Walsh–Hadamard algorithm. Walsh–Hadamard encoding uses $O(n \log(n))$ field additions and/or subtractions, and $O(n)$ multiplications, hence $T_{\text{enc}} = O(n \log n)$ at a fixed rate. Quasar [JLX+26] combines QA codes with BaseFold via code switching to reduce proof size and verification cost. We remark that [LLX+26, Theorem 1.1] does not establish asymptotic goodness at fixed c and p: its entropy condition fails as $s \to \infty$. Thus, these concrete parameters are not a fixed-field asymptotic-distance guarantee; i.e., it is unknown if the QA code is asymptotically good.

1.3 AI Disclosure

In preparing this manuscript, the authors used OpenAI Codex (GPT-6) to assist with drafting and editing, reviewing parameter-evaluation code, and LaTeX preparation. All AI-assisted material was reviewed, revised, and verified by the authors, who determined the research contributions and technical content and take full responsibility for the final manuscript.

2 Preliminaries

We let $\mathbb{N}$ denote the set of natural numbers, $\mathbb{Z}, \mathbb{Z}^+$ respectively denote the set of integers and positive integers, and for $a \in \mathbb{Z}$, $\mathbb{Z}_{\geqslant a}$ is all integers $\geqslant a$. For $n \in \mathbb{Z}^+$, we let $[n] := \{1, \ldots, n\}$, and for $a \leqslant b \in \mathbb{Z}$, we let $[a, b] := \{a, a + 1, \ldots, b\}$. When necessary, we let $[a, b]_{\mathbb{R}}$ denote the real interval from a to b. We let log denote the binary logarithm and ln denote the natural logarithm.

We let $\mathbb{F}_q$ denote a finite field of size $q = p^a$ for prime p and $a \in \mathbb{Z}^+$, and omit the subscript q when it is clear from context. We let boldface lowercase letters $\mathbf{x} \in \mathbb{F}^k$ denote row-vectors, and boldface uppercase letters $\mathbf{M} \in \mathbb{F}^{k \times n}$ denote matrices. We also let $\mathbf{0}^k$ denote the all-zero vector of length k. For a vector $\mathbf{x} \in \mathbb{F}^k$, we let $\mathsf{wt}(\mathbf{x})$ denote the Hamming weight of $\mathbf{x}$; i.e., the number of non-zero entries of $\mathbf{x}$.

Distributions. For a set S, we let $s \xleftarrow{\$} S$ denote uniformly sampling an element of S. For a distribution $\mathcal{D}$ over a set S, we let $s \leftarrow \mathcal{D}$ denote sampling an element of S according to the distribution $\mathcal{D}$. We let $\mathcal{D}^m$ denote the process of independently sampling from $\mathcal{D}$ m times.

The *hypergeometric distribution*, denoted by $\mathsf{Hypergeom}(N, K, m)$, is a distribution over N objects where K have a specific feature and m are sampled without replacement. Thus, for $X \leftarrow \mathsf{Hypergeom}(N, K, m)$, $\Pr[X = k]$ is the probability of sampling k objects with the specific feature from m objects without replacement from N total objects. The *generalized Bernoulli distribution over* $\mathbb{F}$, denoted by $\mathsf{Ber}_p(\mathbb{F})$ for probability $p \in (0, 1)$, which samples uniformly from $\mathbb{F} \setminus \{0\}$ with probability p and samples 0 with probability $1 - p$. Note that for $p = (q - 1)/q$, $\mathsf{Ber}_p(\mathbb{F})$ is simply the uniform distribution over $\mathbb{F}$. Finally, the *Binomial distribution*, denoted as $\mathsf{Bin}(N, p)$ for probability $p \in (0, 1)$, which samples N coins according to $\mathsf{Ber}_p(\mathbb{F}_2)$.

Coding Theory. All codes in this work are *linear codes*. A $[n, k]_q$ linear code C is a k-dimensional subspace of $\mathbb{F}_q^n$ and is uniquely defined by a generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ as $C := \{\mathbf{xG} \in \mathbb{F}^n : \mathbf{x} \in \mathbb{F}_q^k\}$. We say that k is the

message length, n is the codeword length, and define $\rho := k/n$ as the rate of the code. The minimum distance of a code C, denoted by $d_{\min}(C)$, is defined as $d_{\min}(C) := \min_{\mathbf{c}\in C\setminus\{\mathbf{0}^n\}}(\mathsf{wt}(\mathbf{c})) = \min_{\mathbf{x}\in\mathbb{F}^k\setminus\{\mathbf{0}^k\}}(\mathsf{wt}(\mathbf{xG}))$. We let $\delta_{\min}(C) := d_{\min}(C)/n$ denote the relative minimum distance of C.

3 The Minimum Distance of Encode-Accumulate Codes

The starting point of our analysis is to examine the Repeat-Accumulate code (and variants thereof) in a more general sense as "Encode-Accumulate" codes. Let $\mathbb{F}$ be a finite field of size q, let $k \in \mathbb{Z}^+$ be the message length, let $\rho \in (0,1)$ be rate of the code, and define $n := k/\rho$. Let $\mathbf{M} \in \mathbb{F}^{k\times n}$ be some linear code sampled from some distribution $\mathcal{D}$, and let $\mathbf{A} \in \{0,1\}^{n\times n} \subseteq \mathbb{F}^{n\times n}$ be the accumulator, defined as $\mathbf{z} = \mathbf{yA}$ such that $z_i = \sum_{j=1}^{i} y_j$. Then, in the most general sense, encode-accumulate codes have generator matrices $\mathbf{G} := \mathbf{MA}$, and the distance analysis proceeds as follows.

Let $\delta \in (0,1)$ be the target distance parameter. Then, nearly all distance analyses, to the best of our knowledge, use the union bound to analyze the probability $\mathbf{G}$ has distance at most δ, expressed as follows:

$$\Pr_{\mathbf{M}\leftarrow\mathcal{D}}[\exists \mathbf{x} \in \mathbb{F}^k \setminus \{\mathbf{0}^k\} \colon \mathsf{wt}(\mathbf{xMA}) \leqslant \delta n] \leqslant \sum_{\ell=1}^{k} \binom{k}{\ell}(q-1)^{\ell-1} \cdot \Pr_{\mathbf{M}\leftarrow\mathcal{D}}[\mathsf{wt}(\mathbf{xMA}) \leqslant \delta n | \mathsf{wt}(\mathbf{x}) = \ell],$$

where $\mathsf{wt}(\cdot)$ represents the Hamming weight.[2] Now, setting $\mathbf{y} := \mathbf{xM}$, the tail probability on the right-hand side is further broken down as

$$\Pr_{\mathbf{M}\leftarrow\mathcal{D}}[\mathsf{wt}(\mathbf{xMA}) \leqslant \delta n | \mathsf{wt}(\mathbf{x}) = \ell] = \sum_{w=1}^{2\lfloor \delta n\rfloor} \Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell] \Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{yA}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}) = w],$$

where the above summation over w above stops at $2\lfloor \delta n\rfloor$ due to the fact that $\mathsf{wt}(\mathbf{zA}) \geqslant \lceil \mathsf{wt}(\mathbf{z})/2\rceil$ for any $\mathbf{z}$ [BFK+24]. Finally, we can again break down the tail probability above as

$$\Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{yA}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}) = w] = \sum_{h=1}^{\lfloor \delta n\rfloor} \Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{yA}) = h | \mathsf{wt}(\mathbf{y}) = w].$$

Prior Analyses

At this point, prior analyses (e.g., [BFK+24, AHK+26, AZ26, BFM+26, BCF+25, HGQ+26]) directly utilize the structure of the distribution $\mathcal{D}$ of encoding matrices $\mathbf{M}$. As a concrete example, consider the generalized Repeat-Accumulate code, defined as $\mathbf{M} = \mathbf{RP}$, where $\mathbf{R} \in \{0,1\}^{k\times n}$ is defined as $\mathbf{xR} = (\mathbf{x}, \ldots, \mathbf{x}) \in (\mathbb{F}^k)^r$, for repetition factor $r := 1/\rho \in \mathbb{Z}^+$, and $\mathbf{P} = \mathbf{TQ}$ where $\mathbf{T} \in \mathbb{F}^{n\times n}$ is a random full-rank diagonal matrix and $\mathbf{Q} \in \{0,1\}^{n\times n}$ is a random permutation matrix; here and throughout, we assume $\{0,1\} \subseteq \mathbb{F}$. Directly taking advantage of the structure of $\mathbf{RP}$, to analyze $\Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{yA}) = h | \mathsf{wt}(\mathbf{y}) = w]$, prior work used the so-called *input-output weight enumerator (IOWE)* $A_{w,h}^{(n,q)}$, which is the number of vectors $\mathbf{y} \in \mathbb{F}^n$ of weight $\mathsf{wt}(\mathbf{y}) = w$ such that $\mathsf{wt}(\mathbf{yA}) = h$. Then this conditional probability is analyzed as $A_{w,h}^{(n,q)}/\left(\binom{n}{w}\cdot(q-1)^w\right)$.

It is known that the IOWE for the Accumulator code $\mathbf{A}$ over a finite field of size q satisfies [GR11, BFK+24]

$$A_{w,h}^{(n,q)} = \sum_{i=0}^{w-1} \binom{h-1}{\lceil \frac{w-i}{2}\rceil - 1}\binom{n-h}{\lfloor \frac{w-i}{2}\rfloor}\binom{h-\lceil \frac{w-i}{2}\rceil}{i}(q-1)^{\lceil \frac{w-i}{2}\rceil}(q-2)^i,$$

[2] Those familiar with the union bound will notice an $\ell - 1$ power of $(q-1)$ instead of ℓ. This is due to the fact that if $\mathsf{wt}(\mathbf{xM}) = w$ for any w and matrix $\mathbf{M}$, then $\mathsf{wt}(\mathbf{yM}) = w$ for any $\mathbf{y} = \alpha\mathbf{x}$ since $\mathbf{M}$ is a linear function, where $\alpha \in \mathbb{F}\setminus\{0\}$.

with the special case of $q = 2$ reducing to $A^{(n,2)}_{w,h} = \binom{h-1}{\lceil \frac{w}{2} \rceil - 1}\binom{n-h}{\lfloor \frac{w}{2} \rfloor}$ [DJM98]. Then, analyzing the union bound over a non-binary finite field reduces to manipulating the following summation

$$\sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{h=\lceil r\ell/2 \rceil}^{\lfloor \delta n \rfloor} \frac{A^{(n,q)}_{w,h}}{\binom{n}{w}(q-1)^w}.$$

Prior approaches now *stratify* the union bound by partitioning the input space into *low-weight* and *high-weight*: for some $\beta := \beta(k) > 0$, the low-weight partition is $\ell \leqslant \beta k$, while the high-weight partition is $\beta k \leqslant \ell \leqslant k$. For example, [BFK+24] sets $\beta = \gamma k/\log(k)$ for some constant γ, [HGQ+26] sets $\beta = 0.01k$, and [AZ26] sets $\beta = O(k/\log^2(k))$. Within the high-weight partition, a negligible upper bound is often obtained, followed by an inverse polynomial upper bound on the low-weight partition. Overall, this yields an inverse polynomial failure probability.

This approach does work and gives desired asymptotic bounds. However, the concrete upper bounds obtained by such an approach are often messy and quite loose. Thus, to generate parameters, the above union bound is often *directly computed* for small values of k (e.g., $k = 2^{12}$ in [BFK+24], $k = 2^{20}$ in [HGQ+26], and $k = 2^{25}$ in [AZ26, BCF+25]), since the summation is of size $\Omega(k^3)$, and computing large binomials and exponentials is concretely expensive, even on modern hardware.

We also view this approach as non-generalizable and somewhat wasteful: the IOWE method fixes $\mathbf{y}$ to be a uniformly random weight-w vector and asks what is the probability that this random vector is in the "bad set" (i.e., a vector counted by $A^{(n,q)}_{w,h}$). This approach cannot generalize to *other* distributions of $\mathbf{y}$ that aren't uniform over some space of fixed-weight vectors, nor does it take advantage of any structure the distribution of $\mathbf{y}$ may have in understanding the random variable $\mathsf{wt}(\mathbf{yA})$. Finally, although all known asymptotic bounds on these codes (that we are aware of) utilize the IOWE method, this does not imply it is the best method for deriving better *concrete* bounds.

Our Approach

For our more general and fine-grained analysis of accumulate codes, we forgo taking explicit advantage of the structure of the distribution $\mathcal{D}$ (of matrices $\mathbf{M}$) *at this stage*. Instead, our approach, in a nutshell, breaks down the structure of $\mathbf{y}$ that is needed in order to satisfy $\mathsf{wt}(\mathbf{yA}) = h$, then analyzes the probability $\mathbf{y}$ satisfies this structure for a given distribution $\mathcal{D}$.

First, conditioning on the event that $\mathsf{wt}(\mathbf{y}) = w$, we can uniquely write $\mathbf{y}$ as $\mathbf{y} = \mathbf{0}^{G_0}\alpha_1\mathbf{0}^{G_1}\alpha_2 \cdots \alpha_w\mathbf{0}^{G_w}$, where $\alpha_1, \ldots, \alpha_w \in \mathbb{F} \setminus \{0\}$ are the w non-zeros of $\mathbf{y}$, and $G_0, \ldots, G_w \in \mathbb{N}$ denote the number of zeros in the $w+1$ *gaps* determined by the non-zero positions. As $\mathsf{wt}(\mathbf{y}) = w$, the gap variables satisfy

$$\sum_{i=0}^{w} G_i = n - w. \tag{3.1}$$

Weight Counting. Given the above unique representation of $\mathbf{y}$, we can compute $\mathbf{z} = \mathbf{yA}$ using $z_i = \sum_{j \leqslant i} y_j$ and observe the following structure of $\mathbf{z}$:

$$\begin{aligned}
z_1 = z_2 = \cdots = z_{G_0} &= 0,\\
z_{G_0+1} = z_{G_0+2} = \cdots = z_{G_0+G_1+1} &= \alpha_1,\\
z_{G_0+G_1+2} = z_{G_0+G_1+3} = \cdots = z_{G_0+G_1+G_2+2} &= \alpha_1 + \alpha_2,\\
z_{G_0+G_1+G_2+3} = z_{G_0+G_1+G_2+4} = \cdots = z_{G_0+G_1+G_2+G_3+3} &= \alpha_1 + \alpha_2 + \alpha_3,
\end{aligned}$$

$$\vdots$$

$$z_{G_0+\cdots+G_{w-1}+w} = z_{G_0+\cdots+G_{w-1}+w+1} = \cdots = z_n = \alpha_1 + \alpha_2 + \cdots + \alpha_w.$$

Therefore, we can express the weight of $\mathbf{z}$ as partial sums S_i, defined for all $i \in [w]$ as $S_i := \sum_{j=1}^{i} \alpha_j \in \mathbb{F}$. By the previous derivation of $\mathbf{z}$, we have that $z_j = 0$ for $j \in [G_0]$. Next, α_1 is encountered, and the accumulator outputs the symbol $S_1 = \alpha_1$ for $1 + G_1$ consecutive coordinates; i.e., $z_j = S_1$ for $j \in \{1 + G_0, \ldots, 1 + G_0 + G_1\}$. Generalizing, after the ith non-zero input α_i is encountered, the accumulator outputs the symbol S_i for exactly $1 + G_i$ consecutive coordinates of $\mathbf{z}$; explicitly, $z_j = S_i$ for $j \in \{i + \sum_{j=0}^{i-1} G_j, \ldots, i + \sum_{j=0}^{i-1} G_j + G_i\}$.

This implies we can express $\mathsf{wt}(\mathbf{z})$ as the following sum:

$$\mathsf{wt}(\mathbf{z}) = \sum_{i=1}^{w} \mathbb{I}(S_i \neq 0)(1 + G_i) = \underbrace{\sum_{i=1}^{w} \mathbb{I}(S_i \neq 0)}_{:= C_w} + \underbrace{\sum_{i=1}^{w} \mathbb{I}(S_i \neq 0) \cdot G_i}_{:= T_w},$$

where $\mathbb{I}(\cdot) \in \{0, 1\}$ is the indicator function. With $\mathsf{wt}(\mathbf{z}) = C_w + T_w$, we have a clearer picture of when $\mathsf{wt}(\mathbf{z}) = h$: $h = C_w + T_w$. Since T_w depends on C_w, for the above equality to hold, if $C_w = c$, then $T_w = h - c$ must occur. Taking advantage of the fact that $\mathsf{wt}(\mathbf{z}) \geqslant \lceil w/2 \rceil$ and that $C_w \leqslant w$, we have

$$\Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{z}) = h | \mathsf{wt}(\mathbf{y}) = w] = \sum_{c=\lceil w/2 \rceil}^{\min(h,w)} \Pr_{\mathbf{M}}[C_w = c] \cdot \Pr_{\mathbf{M}}[T_w = h - c | C_w = c].$$

A Fine-Grained Accumulator Union Bound. Thus, we can rewrite the union bound of *any* encode-accumulate code as

$$\begin{aligned}
&\Pr_{\mathbf{M}}[\exists \mathbf{x} \in \mathbb{F}^k \setminus \{\mathbf{0}^k\} \colon \mathsf{wt}(\mathbf{xMA}) \leqslant \delta n] \\
&\leqslant \sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{w=1}^{2\lfloor \delta n \rfloor} \Pr_{\mathbf{M}}[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell] \sum_{h=\lceil w/2 \rceil}^{\lfloor \delta n \rfloor} \left(\sum_{c=\lceil w/2 \rceil}^{\min(h,w)} \Pr_{\mathbf{M}}[C_w = c] \Pr_{\mathbf{M}}[T_w = h - c | C_w = c] \right).
\end{aligned} \tag{3.2}$$

We view Eq. (3.2) as a cleaner and more fine-grained union bound, specifically for encode-accumulate codes, which (1) can fully take advantage of the properties of the distribution of $\mathbf{M} \leftarrow \mathcal{D}$, and (2) readily extends to distributions for which the vector $\mathbf{y}$ is not simply uniform over all weight-w vectors in $\mathbb{F}^n$. Now, the union bound analysis above boils down to analyzing $\Pr[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell]$, $\Pr[C_w = c]$, and $\Pr[T_w = h - c | C_w = c]$, giving a bound that only requires one to analyze these three random variables, which may be simpler than trying to bound many products of binomials and exponentials (as in the IOWE analysis). We demonstrate the versatility of this union bound for the RA, RAA, and EA codes in the remainder of this section.

3.1 The Generalized Repeat-Accumulate Code

In light of Eq. (3.2), we analyze the distance of the generalized RA code as follows. First, recall the definition of the generalized RA code. For repetition parameter $r \in \mathbb{Z}_{\geqslant 3}$, define $\mathbf{M} = \mathbf{RP}$, where $\mathbf{R} \in \{0, 1\}^{k \times n}$ is defined as $\mathbf{xR} = (\mathbf{x}, \ldots, \mathbf{x}) \in (\mathbb{F}^k)^r$, and $\mathbf{P} = \mathbf{TQ}$ where $\mathbf{T} \in \mathbb{F}^{n \times n}$ is a random full-rank diagonal matrix and

$\mathbf{Q} \in \{0,1\}^{n\times n}$ is a random permutation matrix. Now, by definition of **RP**, for any $\mathbf{x} \in \mathbb{F}^k \setminus \{\mathbf{0}^k\}$ and $\mathbf{y} = \mathbf{xRP}$, $\Pr[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell] = 1$ if and only if $w = r\ell$ (and is 0 otherwise). Then, we can rewrite Eq. (3.2) as

$$\sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{h=\lceil w/2 \rceil}^{\lfloor \delta n \rfloor} \sum_{c=\lceil w/2 \rceil}^{\min(h,w)} \Pr_{\mathbf{P}}[C_w = c] \Pr_{\mathbf{P}}[T_w = h - c | C_w = c], \tag{3.3}$$

where $w := r\ell$. Now, we can take advantage of the structure of $\mathbf{P} = \mathbf{TQ}$.

Recall that $C_w = \sum_{i=1}^{w} \mathbb{I}(S_i \neq 0)$ and that $S_i = \sum_{j=1}^{i} \alpha_j$, where $(\alpha_1, \ldots, \alpha_w)$ are the non-zeros of $\mathbf{y}$. Since $\mathbf{T}$ is a random diagonal matrix, this implies that $(\alpha_1, \ldots, \alpha_w)$ is uniformly distributed over $(\mathbb{F} \setminus \{0\})^w$. Therefore, we can analyze C_w as a discrete-time, two-state Markov chain with state space $\{0,1\}$, where state 1 indicates $S_i \neq 0$ while state 0 indicate $S_i = 0$, with the following structure. While in state 1 at time i, the chain transitions to 0 with probability $p := 1/(q-1)$ (indicating $\alpha_{i+1} = -S_i \neq 0$), and remains in 1 with probability $1 - p$ (indicating $S_{i+1} \neq 0$). However, while in state 0 ($S_i = 0$), the chain transitions to state 1 with probability 1. The transition matrix for this Markov chain is thus given by $Q := \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$. With the Markov chain defined, C_w is now the random variable representing the amount of time the chain remains in the 1 state. Using standard combinatorial arguments, we show the following.

Proposition 3.1. *Let $q > 2$. For every integer c such that $\lceil w/2 \rceil \leqslant c \leqslant w$:*

$$\Pr[C_w = c] = \frac{cq - w}{c(q-2)} \binom{c}{w-c} \left(\frac{1}{q-1}\right)^{w-c} \left(\frac{q-2}{q-1}\right)^{2c-w}.$$

Moreover, $\Pr[C_w = c] = 0$ *otherwise.*

Proof. First, for $\mathbf{y} = \mathbf{xRP}$ non-zero $\mathbf{x} \in \mathbb{F}^k$ of weight ℓ, let $(\alpha_1, \ldots, \alpha_w)$ denote the non-zeros of $\mathbf{y}$, where $w = r\ell$ for repetition factor r. Let $q > 2$ denote the cardinality of $\mathbb{F}$. Recall the definition of C_w as

$$C_w = \sum_{i=1}^{w} \mathbb{I}(S_i \neq 0) \qquad S_i = \sum_{j=1}^{i} \alpha_j.$$

Now, we can formally define the markov chain $\mathcal{V}$ representing C_w as follows. Define $V_i = \mathbb{I}(S_i \neq 0)$ for all $i \in [w]$, and set $\mathcal{V} = (V_i)_{i=1}^{w}$ as the Markov chain. This is clearly a discrete-time Markov chain; moreover, the state-space is $\{0,1\}$ by definition of the indicator function $\mathbb{I}$.

Now, $V_i = 0$ if and only if $S_i = 0$. Moreover, if $S_i = 0$, then by definition, $S_{i+1} = S_i + \alpha_{i+1} = \alpha_{i+1} \neq 0$. Therefore,

$$\Pr[V_{i+1} = 1 \mid V_i = 0] = 1.$$

Next, $V_i = 1$ if and only if $S_i \neq 0$. By definition of S_{i+1}, we have that $S_{i+1} = 0$ if and only if $\alpha_{i+1} = -S_i$. This happens exactly with probability $p := 1/(q-1)$. Thus,

$$\Pr[V_{i+1} = 0 \mid V_i = 1] = p.$$

This immediately implies

$$\Pr[V_{i+1} = 1 \mid V_i = 1] = 1 - p.$$

Therefore, the transition matrix of $\mathcal{V}$ is given by

$$Q := \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}.$$

Now, by the above derivation, we have

$$C_w = \sum_{i=1}^{w} V_i.$$

Suppose that a realization of the Markov chain $\mathcal{V}$ satisfies $C_w = c$. Then, the (w-tuple) sample path $(V_1, \ldots, V_w)$ contains exactly c 1s and $w - c$ 0s. Observe that every 0 is isolated. As mention before, $V_i = 0 \iff S_i = 0$, and therefore

$$S_{i-1} = -\alpha_i \neq 0, \quad \text{and} \quad S_{i+1} = \alpha_{i+1} \neq 0.$$

Hence, every 0 is both preceded and followed by a 1, whenever those neighboring indices exist. Since the chain always starts in state $V_1 = 1$, each 0 must be immediately preceded by a 1. Consequently, the support of C_w is

$$\left\lceil \frac{w}{2} \right\rceil \leqslant C_w \leqslant w.$$

We now count all admissible sample paths having exactly c 1s. There are two possible terminal states, and hence two cases:

1. $V_w = 0$: Since the path ends in 0, the final 1 necessarily precedes a 0. Among the remaining $c - 1$ 1s, $w - c - 1$ must also precede 0s. Hence, the number of such paths is $\binom{c-1}{w-c-1}$. Each such path contains
   - $w - c$ transitions of type $1 \to 0$,
   - $2c - w$ transitions of type $1 \to 1$,

   and therefore occurs with probability $p^{w-c}(1-p)^{2c-w}$.
2. $V_w = 1$: Now the final 1 is not followed by a 0. Therefore exactly $w - c$ of the first $c - 1$ 1s precede 0s. Hence, there are $\binom{c-1}{w-c}$ such paths. Each such path contains
   - $w - c$ transitions of type $1 \to 0$,
   - $2c - w - 1$ transitions of type $1 \to 1$,

   and therefore occurs with probability $p^{w-c}(1-p)^{2c-w-1}$.

The above two cases are mutually exclusive and exhaustive. Thus,

$$\Pr[C_w = c] = \binom{c-1}{w-c-1} p^{w-c}(1-p)^{2c-w} + \binom{c-1}{w-c} p^{w-c}(1-p)^{2c-w-1}.$$

We now simplify this expression into the required form. Using the identities

$$\binom{c-1}{w-c-1} = \frac{w-c}{c}\binom{c}{w-c}, \quad \text{and} \quad \binom{c-1}{w-c} = \frac{2c-w}{c}\binom{c}{w-c},$$

we obtain

$$\Pr[C_w = c] = \frac{w-c}{c} \Pr[\mathsf{Bin}(c,p) = w-c] + \Pr[\mathsf{Bin}(c-1,p) = w-c].$$

Furthermore,

$$\Pr[\mathsf{Bin}(c-1,p) = w-c] = \frac{2c-w}{c(1-p)} \Pr[\mathsf{Bin}(c,p) = w-c].$$

Substituting this identity gives

$$\begin{aligned}\Pr[C_w = c] &= \frac{1}{c}\left(w - c + \frac{2c-w}{1-p}\right) \Pr[\mathsf{Bin}(c,p) = w-c] \\ &= \frac{1}{c}\left(\frac{c(1+p) - wp}{1-p}\right) \Pr[\mathsf{Bin}(c,p) = w-c].\end{aligned}$$

Finally, substituting $p = \frac{1}{q-1}$ and unpacking the binomial PMF yields

$$\Pr[C_w = c] = \frac{cq-w}{c(q-2)} \binom{c}{w-c} \left(\frac{1}{q-1}\right)^{w-c} \left(\frac{q-2}{q-1}\right)^{2c-w}.$$

This completes the proof. □

Next, we analyze T_w and $\Pr[T_w = h - c | C_w = c]$. Observe that the random permutation $\mathbf{Q}$ distributes the support of $\mathbf{yT}$ uniformly over all w-subsets of $[n]$; equivalently, the gap vector $(G_0, \ldots, G_w)$ is uniformly distributed over weak compositions of $n - w$ into $w + 1$ non-negative parts. This fact allows us to exactly characterize $\Pr[T_w = h - c | C_w = c]$ as follows.

Proposition 3.2. *For every integer t such that $0 \leqslant t \leqslant n - w$:*

$$\Pr[T_w = t | C_w = c] = \frac{\binom{t+c-1}{c-1}\binom{n-t-c}{w-c}}{\binom{n}{w}},$$

where $\binom{a}{b} = 0$ whenever $a, b < 0$ or $a < b$, and $\Pr[T_w = t | C_w = c] = 0$ otherwise.

Proof. Since each gap variable is nonnegative and all sum to $n - w$ by Eq. (3.1), the sum of a subset of them must lie between 0 and $n - w$. Consequently, the support of T_w is

$$0 \leqslant T_w \leqslant n - w. \tag{3.4}$$

Now, c selected-gap variables must sum to t. The number of weak compositions of t into c nonnegative parts is $\binom{t+c-1}{c-1}$. The remaining $w + 1 - c$ gap variables must sum to $n - w - t$. By counting weak compositions of $n - w - t$ into $w + 1 - c$ nonnegative parts, this may occur in $\binom{n-t-c}{w-c}$ ways. Since the gap vector is uniformly distributed over the $\binom{n}{w}$ weak compositions of $n - w$ into $w + 1$ nonnegative parts, the stated probability follows immediately. □

Proposition 3.2 now allows us to rewrite the outer summations of Eq. (3.3) as

$$\sum_{c=\lceil w/2 \rceil}^{\min(w, \lfloor \delta n \rfloor)} \Pr[C_w = c] \sum_{h=c}^{\lfloor \delta n \rfloor} \Pr[T_w = h - c | C_w = c], \tag{3.5}$$

where Proposition 3.2 allows us to remove the dependence on h from the c summation since $\Pr[T_w = t | C_w = c] = 0$ whenever $c > w$. In this form, we show that the inner summation of Eq. (3.5) is distributed as a hypergeometric tail bound.

Proposition 3.3. *Let* $\mathscr{H}_w \leftarrow \mathsf{Hypergeom}(n, \lfloor \delta n \rfloor, w)$. *Then, for every integer* c,

$$\sum_{h=c}^{\lfloor \delta n \rfloor} \frac{\binom{h-1}{c-1}\binom{n-h}{w-c}}{\binom{n}{w}} = \Pr[\mathscr{H}_w \geqslant c].$$

This allows us to now cleanly rewrite Eq. (3.3) as

$$\text{Eq. (3.3)} \leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{c=\lceil w/2 \rceil}^{w} \Pr[C_w = c] \Pr[\mathscr{H}_w \geqslant c]. \tag{3.6}$$

Moreover, Proposition 3.3 is especially powerful since we can use it to directly apply well-studied and tight tail bounds on the hypergeometric distribution. Theorem 1.1 then follows by using these tail bounds and performing algebraic manipulations to obtain the final bounds. We employ two strategies: a *Binomial-Moment (BM) strategy* and a *Chernoff strategy*. We first focus on the Binomial-Moment strategy, followed by the Chernoff Strategy.

3.1.1 Binomial-Moment Strategy

We use the following BM hypergeometric tail bound.

Lemma 3.1 (BM bound)**.** *For every* $c \in [w]$, $\Pr[\mathscr{H}_w \geqslant c] \leqslant \binom{w}{c}\delta^c$.

Proof. Let $\mathfrak{d} := \lfloor \delta n \rfloor$. We apply Markov's inequality to the cth binomial moment of $\mathscr{H}_w$. Since $\binom{\mathscr{H}_w}{c} \geqslant 1$, whenever $\mathscr{H}_w \geqslant c$, we have

$$\mathbb{I}(\mathscr{H}_w \geqslant c) \leqslant \binom{\mathscr{H}_w}{c}.$$

Hence, taking expectations, we have

$$\begin{aligned}
\Pr[\mathscr{H}_w \geqslant c] &\leqslant \mathbb{E}\left[\binom{\mathscr{H}_w}{c}\right] = \sum_{h=c}^{\min(w,\mathfrak{d})} \binom{h}{c} \frac{\binom{\mathfrak{d}}{h}\binom{n-\mathfrak{d}}{w-h}}{\binom{n}{w}} \\
&= \frac{1}{\binom{n}{w}} \sum_{h=c}^{\min(w,\mathfrak{d})} \binom{\mathfrak{d}}{c}\binom{\mathfrak{d}-c}{h-c}\binom{n-\mathfrak{d}}{w-h} \\
&= \frac{\binom{\mathfrak{d}}{c}}{\binom{n}{w}} \sum_{s=0}^{\min(w-c,\mathfrak{d}-c)} \binom{\mathfrak{d}-c}{s}\binom{n-\mathfrak{d}}{w-c-s} \\
&\leqslant \frac{\binom{\mathfrak{d}}{c}}{\binom{n}{w}}\binom{n-c}{w-c} && \text{(Vandermonde Identity)} \\
&= \binom{w}{c}\frac{\binom{\mathfrak{d}}{c}}{\binom{n}{c}} = \binom{w}{c}\frac{(\mathfrak{d})_c}{(n)_c},
\end{aligned}$$

where $(a)_c := a(a-1)(a-2)\cdots(a-c+1)$ denotes the falling factorial. Therefore,

$$\Pr[\mathscr{H}_w \geqslant c] \leqslant \binom{w}{c}\frac{(\mathfrak{d})_c}{(n)_c} = \binom{w}{c}\prod_{s=0}^{c-1}\frac{\mathfrak{d}-s}{n-s} \leqslant \binom{w}{c}\left(\frac{\mathfrak{d}}{n}\right)^c \leqslant \binom{w}{c}\delta^c. \qquad \square$$

$\square$

We now prove the Binomial-Moment portion of Theorem 1.1.

Proof of Theorem 1.1, Binomial-Moment. We first prove the equivalent form of the theorem without the target τ, and then proceed to derive the targeted version. For the BM strategy, using Proposition 3.1 and Lemma 3.1, we have

$$\text{Eq. (3.6)} \leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell}(q-1)^{\ell-1} \sum_{c=\lceil \frac{w}{2} \rceil}^{w} \tfrac{cq-w}{c(q-2)} \binom{c}{w-c} \left(\tfrac{1}{q-1}\right)^{w-c} \left(\tfrac{q-2}{q-1}\right)^{2c-w} \binom{w}{c} \delta^c. \tag{3.7}$$

Set $u := w - c$. Then, for $c = w - u$ and $u \in [0, \lfloor \frac{w}{2} \rfloor]$, we have $\binom{w}{c}\binom{c}{w-c} = \frac{w!}{u!^2(w-2u)!}$. Furthermore, $\frac{cq-w}{c(q-2)} = 1 + \frac{w-2u}{(w-u)(q-2)}$. Now, we temporarily restrict our focus to the inner c-sum of Eq. (3.7) and rewrite it in terms of u:

$$\begin{aligned}
\mathfrak{S}_{\ell,\delta} &:= \sum_{u=0}^{\lfloor w/2 \rfloor} \left(1 + \frac{w-2u}{(w-u)(q-2)}\right) \frac{w!}{u!^2(w-2u)!} \left(\frac{1}{q-1}\right)^u \left(\frac{q-2}{q-1}\right)^{w-2u} \delta^{w-u} \\
&= \sum_{u=0}^{\lfloor w/2 \rfloor} \left(1 + \frac{w-2u}{(w-u)(q-2)}\right) \frac{w!}{u!^2(w-2u)!} \left(\frac{\delta}{q-1}\right)^u \left(\frac{q-2}{q-1}\delta\right)^{w-2u}.
\end{aligned}$$

For $u \in [0, \lfloor \frac{w}{2} \rfloor]$, we have $1 \leqslant 1 + \frac{w-2u}{(w-u)(q-2)} \leqslant \frac{q-1}{q-2}$. Hence,

$$\mathfrak{S}_{\ell,\delta} \leqslant \frac{q-1}{q-2} \sum_{u=0}^{\lfloor w/2 \rfloor} \frac{w!}{u!^2(w-2u)!} \left(\frac{\delta}{q-1}\right)^u \left(\frac{q-2}{q-1}\delta\right)^{w-2u}.$$

Set $\sigma := \frac{\delta}{q-1}$ and $\gamma := \frac{q-2}{q-1}\delta$. Then,

$$\mathfrak{S}_{\ell,\delta} \leqslant \frac{q-1}{q-2} \sum_{u=0}^{\lfloor w/2 \rfloor} \frac{w!}{u!^2(w-2u)!} \sigma^u \gamma^{w-2u}.$$

Comparing the above u-sum with the following trinomial expansion

$$\left(\gamma + \sqrt{\sigma} + \sqrt{\sigma}\right)^w = \sum_{\substack{u_1+u_2+u_3=w \\ u_j \geqslant 0}} \frac{w!}{u_1!u_2!u_3!} \sigma^{\frac{u_1+u_2}{2}} \gamma^{u_3},$$

we have that the u-sum consists of all terms of the trinomial expansion for which $u_1 = u_2 = u$. Since all terms of the trinomial expansion are nonnegative, we have

$$\mathfrak{S}_{\ell,\delta} \leqslant \tfrac{q-1}{q-2}\left(\gamma + 2\sqrt{\sigma}\right)^w = \tfrac{q-1}{q-2}\left(2\sqrt{\tfrac{\delta}{q-1}} + \tfrac{q-2}{q-1}\delta\right)^w.$$

Define $\Lambda^{\mathsf{bm}}_{q,\delta} := 2\sqrt{\frac{\delta}{q-1}} + \frac{q-2}{q-1}\delta$. Substituting the above inequality into Eq. (3.7) with $w = r\ell$ and $\Lambda^{\mathsf{bm}}_{q,\delta}$ yields

$$\text{Eq. (3.2)} \leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell}(q-1)^{\ell-1} \left[\tfrac{q-1}{q-2}(\Lambda^{\mathsf{bm}}_{q,\delta})^{r\ell}\right]$$

$$= \tfrac{1}{q-2} \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} \left[(q-1)(\Lambda^{\mathsf{bm}}_{q,\delta})^r\right]^\ell = \tfrac{1}{q-2} \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} \xi_{\mathsf{bm}}^\ell \leqslant \tfrac{(1+\xi_{\mathsf{bm}})^k - 1}{q-2},$$

where $\xi_{\mathsf{bm}} := (q-1)(\Lambda^{\mathsf{bm}}_{q,\delta})^r$, and the final upper bound is obtained by extending the sum to k and simplifying via the Binomial theorem. This proves the claim.

Now, for target failure probability $\tau \in (0,1)$, by the above bound, it is sufficient that $((1+\xi_{\mathsf{bm}})^k - 1)/(q-2) \leqslant \tau$, which is equivalent to $(1+\xi_{\mathsf{bm}})^k \leqslant 1 + (q-2)\tau$. Simplifying again and expanding the definition of ξ_{bm}, this is equivalent to $(q-1)(\Lambda^{\mathsf{bm}}_{q,\delta})^r \leqslant [1+(q-2)\tau]^{1/k} - 1$. From this final inequality, define $\eta_{\mathsf{bm}} := \left[([1+(q-2)\tau]^{1/k} - 1)/(q-1)\right]^{1/r}$. This yields the inequality

$$\Lambda^{\mathsf{bm}}_{q,\delta} = 2\sqrt{\delta/(q-1)} + ((q-2)/(q-1))\delta \leqslant \eta_{\mathsf{bm}}.$$

For $\chi = \sqrt{\delta}$, this inequality becomes $((q-2)/(q-1))\chi^2 + (2/\sqrt{q-1})\chi - \eta_{\mathsf{bm}} \leqslant 0$. Solving this yields $\delta \leqslant [(\sqrt{q-1}/(q-2))(\sqrt{1+(q-2)\eta_{\mathsf{bm}}} - 1)]^2$. □ □

From the above proof, we derive the following corollary which will be useful when analyzing the RAA code.

Corollary 3.1. *For* $\mathbf{y} \in \mathbb{F}_q^n$*, generalized permutation* $\mathbf{P} = \mathbf{TQ} \in \mathbb{F}_q^{n\times n}$*,* $\mathbf{z} = \mathbf{yPA}$*, and* $\Lambda^{\mathsf{bm}}_{q,\delta} := 2\sqrt{\delta/(q-1)} + (1 - 1/(q-1))\delta$*, we have* $\Pr_{\mathbf{P}}[\mathsf{wt}(\mathbf{z}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}) = w] \leqslant \frac{q-1}{q-2}(\Lambda^{\mathsf{bm}}_{q,\delta})^w$.

3.1.2 Chernoff Strategy

First, we prove the following lemma bounding the hypergeometric tail via a Chernoff bound.

Lemma 3.2 (Hypergeometric Chernoff bound)**.** *For every* $c \in [w/2, w]$ *and* $\delta \in (0, 1/2)$*,*

$$\Pr[\mathscr{H}_w \geqslant c] \leqslant \exp\left[-w \cdot D_{\mathsf{KL}}\left(\frac{c}{w}\middle\|\delta\right)\right] \leqslant [4\delta(1-\delta)]^{w/2},$$

where $D_{\mathsf{KL}}(a\|b) := a\ln(\frac{a}{b}) + (1-a)\ln(\frac{1-a}{1-b})$ *denotes the Kullback-Leibler divergence.*

Proof. Let $\mathfrak{d} := \lfloor \delta n \rfloor$. By the standard KL divergence form of the Chernoff bound on the hypergeometric tail, we have

$$\Pr[\mathscr{H}_w \geqslant c] \leqslant \exp\left[-w \cdot D_{\mathsf{KL}}\left(\frac{c}{w}\middle\|\frac{\mathfrak{d}}{n}\right)\right]. \tag{3.8}$$

For fixed a, the KL divergence $D_{\mathsf{KL}}(a\|b)$ decreases with b, as long as $b < a$. Also, for every $c \in \left\{\lceil \frac{w}{2} \rceil, \ldots, w\right\}$, and by the definitions of $\mathfrak{d}$ and δ, we have the following chain of inequalities: $\mathfrak{d}/n \leqslant \delta < 1/2 \leqslant c/w$. Therefore,

$$D_{\mathsf{KL}}\left(\frac{c}{w}\middle\|\frac{\mathfrak{d}}{n}\right) \geqslant D_{\mathsf{KL}}\left(\frac{c}{w}\middle\|\delta\right).$$

Now, for fixed b, $D_{\mathsf{KL}}(a\|b)$ increases with a so long as $a > b$. Since $\delta < 1/2 \leqslant c/w$, this gives us

$$D_{\mathsf{KL}}\left(\frac{c}{w}\middle\|\delta\right) \geqslant D_{\mathsf{KL}}\left(\frac{1}{2}\middle\|\delta\right) = -\ln(4\delta(1-\delta))/2.$$

Substituting the above inequality into Eq. (3.8), we get the required result. □ □

Now, we prove the Chernoff portion of Theorem 1.1.

Proof of Theorem 1.1, Chernoff Strategy. Next, we prove the Chernoff portion of the theorem. Now, using Proposition 3.1 and Lemma 3.2, we can upper bound Eq. (3.6) as

$$\leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} (q-1)^{\ell-1} [4\delta(1-\delta)]^{w/2} \sum_{c=\lceil \frac{w}{2} \rceil}^{w} \Pr[C_w = c].$$

Noting that the entire support of C_w is in the set $\{\lceil w/2 \rceil, \ldots, w\}$, we have that $\sum_{c=\lceil w/2 \rceil}^{w} \Pr[C_w = c] = 1$. Therefore, the above inner sum collapses to 1, yielding

$$\Pr[\text{failure}] \leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} (q-1)^{\ell-1} [4\delta(1-\delta)]^{w/2}.$$

Substituting $w = r\ell$ then yields

$$\Pr[\text{failure}] \leqslant \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} (q-1)^{\ell-1} [4\delta(1-\delta)]^{\frac{r\ell}{2}} = \frac{1}{q-1} \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} \left[(q-1)[4\delta(1-\delta)]^{\frac{r}{2}} \right]^{\ell}$$
$$= \frac{1}{q-1} \sum_{\ell=1}^{\lfloor 2\delta k \rfloor} \binom{k}{\ell} \xi_{\mathsf{ch}}^{\ell},$$

where $\xi_{\mathsf{ch}} := (q-1)[4\delta(1-\delta)]^{r/2}$. We again upper bound the above ℓ-sum by removing the sum truncation and letting the sum run up to $\ell = k$; that is,

$$\Pr[\text{failure}] \leqslant \frac{1}{q-1} \sum_{\ell=1}^{k} \binom{k}{\ell} \xi_{\mathsf{ch}}^{\ell}.$$

Finally, collapsing the above as a binomial expansion yields the required result:

$$\Pr[\text{failure}] \leqslant [(1+\xi_{\mathsf{ch}})^k - 1]/(q-1).$$

Finally, for target $\tau \in (0,1)$, we can analogously derive the stated bounds on δ by solving the inequality $(1+\xi_{\mathsf{ch}})^k \leqslant (q-1)\tau + 1$ as in the BM case. □ □

As in the Binomial-Moment case (i.e., the BM proof of Theorem 1.1), we derive the following corollary which will be used when analyzing the RAA code.

Corollary 3.2. *For* $\mathbf{y} \in \mathbb{F}_q^n$*, generalized permutation* $\mathbf{P} = \mathbf{TQ} \in \mathbb{F}_q^{n \times n}$*,* $\mathbf{z} = \mathbf{yPA}$*,* $\delta \in (0, 1/2)$*, and* $\Lambda_{q,\delta}^{\mathsf{ch}} := 2\sqrt{\delta(1-\delta)}$*, we have*

$$\Pr_{\mathbf{P}}[\mathsf{wt}(\mathbf{z}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}) = w] \leqslant (\Lambda_{q,\delta}^{\mathsf{ch}})^w.$$

3.1.3 Comparison Between Binomial-Moment and Chernoff

A natural question to ask is under what circumstances is the BM bound better or worse than the Chernoff bound. Below, we characterize this tradeoff.

Proposition 3.4 (BM versus Ch)**.** *From Theorem 1.1, define*

$$B_{\mathsf{bm}} := \frac{(1+\xi_{\mathsf{bm}})^k - 1}{q-2}; \quad B_{\mathsf{ch}} := \frac{(1+\xi_{\mathsf{ch}})^k - 1}{q-1}.$$

Also define $\Lambda^{\mathsf{bm}}_{q,\delta} = 2\sqrt{\delta/(q-1)} + (1-1/(q-1))\delta$ *and* $\Lambda^{\mathsf{ch}}_{\delta} := 2\sqrt{\delta(1-\delta)}$*. Then, we have the sufficient conditions*

$$\Lambda^{\mathsf{bm}}_{q,\delta} \leqslant \Lambda^{\mathsf{ch}}_{\delta} \implies B_{\mathsf{bm}} \leqslant B_{\mathsf{ch}}; \qquad \Lambda^{\mathsf{bm}}_{q,\delta} \geqslant \left(\frac{q-1}{q-2}\right)^{1/r} \Lambda^{\mathsf{ch}}_{\delta} \implies B_{\mathsf{ch}} \leqslant B_{\mathsf{bm}}.$$

Remark 3.1 (Intermediate-regime comparison). Outside the parameter regimes identified in the proposition, namely when $\Lambda^{\mathsf{ch}}_{\delta} < \Lambda^{\mathsf{bm}}_{q,\delta} < (\frac{q-1}{q-2})^{1/r}\Lambda^{\mathsf{ch}}_{\delta}$, neither bound is guaranteed to dominate the other. In this intermediate regime, the sharper estimate is obtained by comparing B_{bm} and B_{ch} directly.

Proof. Define the function

$$f(\xi) = \frac{(1+\xi)^k - 1}{\xi}, \qquad \xi > 0.$$

Using the binomial theorem,

$$f(\xi) = \sum_{\ell=1}^{k} \binom{k}{\ell} \xi^{\ell-1},$$

so f is increasing on $(0,\infty)$. Hence, whenever $\xi_{\mathsf{bm}} \leqslant \xi_{\mathsf{ch}}$, we have, by the monotonicity of f,

$$\frac{(1+\xi_{\mathsf{bm}})^k - 1}{(1+\xi_{\mathsf{ch}})^k - 1} \leqslant \frac{\xi_{\mathsf{bm}}}{\xi_{\mathsf{ch}}}.$$

Therefore, $B_{\mathsf{bm}} \leqslant B_{\mathsf{ch}}$ is guaranteed whenever

$$\frac{\xi_{\mathsf{bm}}}{\xi_{\mathsf{ch}}} \leqslant \frac{q-2}{q-1}. \tag{3.9}$$

Substituting the definitions of ξ_{bm} and ξ_{ch}, this condition becomes

$$\frac{q-2}{q-1}\left(\frac{\Lambda^{\mathsf{bm}}_{q,\delta}}{\Lambda^{\mathsf{ch}}_{\delta}}\right)^r \leqslant \frac{q-2}{q-1} \implies B_{\mathsf{bm}} \leqslant B_{\mathsf{ch}}, \tag{3.10}$$

which simplifies to the first proposed implication.

For establishing the second proposed implication, observe that

$$\xi_{\mathsf{ch}} \leqslant \xi_{\mathsf{bm}} \iff (q-1)\left(\Lambda^{\mathsf{ch}}_{\delta}\right)^r \leqslant (q-2)\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^r$$

$$\iff \Lambda^{\mathsf{bm}}_{q,\delta} \geqslant \left(\frac{q-1}{q-2}\right)^{1/r} \Lambda^{\mathsf{ch}}_{\delta}.$$

Also, if $\xi_{\mathsf{ch}} \leqslant \xi_{\mathsf{bm}}$, then the numerator of B_{ch} is smaller than that of B_{bm}, and, moreover, the denominator of B_{ch}—$(q-1)$—is larger than that of B_{bm}—$(q-2)$. Hence, if $\xi_{\mathsf{ch}} \leqslant \xi_{\mathsf{bm}}$, then $B_{\mathsf{ch}} \leqslant B_{\mathsf{bm}}$. Therefore,

$$\Lambda^{\mathsf{bm}}_{q,\delta} \geqslant \left(\frac{q-1}{q-2}\right)^{1/r} \Lambda^{\mathsf{ch}}_{\delta} \iff \xi_{\mathsf{ch}} \leqslant \xi_{\mathsf{bm}} \implies B_{\mathsf{ch}} \leqslant B_{\mathsf{bm}}.$$

This completes the proof. □ □

3.2 The Generalized RAA Code

The Generalized Repeat-Accumulate-Accumulate (RAA) code is defined analogously as the RA code, except there is a second "permute-accumulate" step. This corresponds to the matrix $\mathbf{M} = \mathbf{R}\mathbf{P}_1\mathbf{A}\mathbf{P}_2$ in our generalized view of "encode-accumulate", where $\mathbf{R}$ is again the $k \times n$ repeat matrix for $n = rk$, and $\mathbf{P}_i = \mathbf{T}_i\mathbf{Q}_i$ for $i \in [2]$ are, respectively, uniformly random full-rank diagonal and permutation matrices. Define $\mathbf{y}_1 = \mathbf{x}\mathbf{R}\mathbf{P}_1$, $\mathbf{z}_1 = \mathbf{y}_1\mathbf{A}$, $\mathbf{y}_2 = \mathbf{z}_1\mathbf{P}_2$, and $\mathbf{z} = \mathbf{y}_2\mathbf{A}$. Then, for $w := r\ell$, we can rewrite Eq. (3.2) as

$$\sum_{\ell=1}^{\min(k,\lfloor 4\delta k\rfloor)} \binom{k}{\ell}(q-1)^{\ell-1} \sum_{h=\lceil w/2\rceil}^{2\lfloor \delta n\rfloor} \Pr_{\mathbf{P}_2}[\mathsf{wt}(\mathbf{z}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}_2) = h] \Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{z}_1) = h | \mathsf{wt}(\mathbf{y}_1) = w]. \quad (3.11)$$

Here, we truncate the sums since by the definition of the accumulator, we have $\mathsf{wt}(\mathbf{z}) \geqslant \lceil \mathsf{wt}(\mathbf{z}_1)/2\rceil \geqslant \lceil r\mathsf{wt}(\mathbf{x})/2\rceil$. In this form, we can leverage our prior RA code analysis (Corollary 3.1 and Propositions 3.1 and 3.2) to prove Theorems 1.2 and 1.3.

3.2.1 Binomial-Moment Strategy

As before, we begin with the Binomial-Moment Strategy.

Proof of Theorem 1.2. First, define $\Lambda^{\mathsf{bm}}_{q,\delta} := 2\sqrt{\delta/(q-1)} + (1-1/(q-1))\delta$. Then, by Corollary 3.1, we can upper bound the inner h-sum above as

$$\begin{aligned}
&\sum_{h=\lceil w/2\rceil}^{2\lfloor \delta n\rfloor} \Pr[\mathsf{wt}(\mathbf{z}_1) = h | \mathsf{wt}(\mathbf{y}_1) = w] \Pr[\mathsf{wt}(\mathbf{z}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}_2) = h] \\
&\leqslant \sum_{h=\lceil w/2\rceil}^{2\lfloor \delta n\rfloor} \Pr[\mathsf{wt}(\mathbf{z}_1) = h | \mathsf{wt}(\mathbf{y}_1) = w] \frac{q-1}{q-2}(\Lambda^{\mathsf{bm}}_{q,\delta})^h \leqslant \frac{q-1}{q-2} \mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}[(\Lambda^{\mathsf{bm}}_{q,\delta})^{\mathsf{wt}(\mathbf{z}_1)}].
\end{aligned}$$

Now, recall the definition of the random variable $C_w = \sum_{i=1}^{w} \mathbb{I}(S_i \neq 0)$, where $S_i = \sum_{j=1}^{i} \alpha_j$ vector $(\alpha_1, \ldots, \alpha_w)$ representing the non-zeros of $\mathbf{y}_1$, and assume that $\Lambda^{\mathsf{bm}}_{q,\delta} < 1$. By the law of total expectation, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}\left[\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^{\mathsf{wt}(\mathbf{z}_1)}\right] &= \mathop{\mathbb{E}}_{C_w}\left[\mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}\left[\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^{\mathsf{wt}(\mathbf{z}_1)} \middle| C_w\right]\right] \\
&= \sum_{c=\lceil w/2\rceil}^{w} \Pr[C_w = c] \mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}\left[\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^{\mathsf{wt}(\mathbf{z}_1)} \middle| C_w = c\right]. \quad (3.12)
\end{aligned}$$

Expanding the inner conditional expectation yields

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}\left[\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^{\mathsf{wt}(\mathbf{z}_1)} \,\middle|\, C_w = c\right] = \sum_{h=c}^{n-w+c} \Pr[\mathsf{wt}(\mathbf{z}_1) = h \mid C_w = c]\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h \\
&= \sum_{h=c}^{n-w+c} \Pr[T_w = h-c \mid C_w = c]\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h = \sum_{h=c}^{n-w+c} \frac{\binom{h-1}{c-1}\binom{n-h}{w-c}}{\binom{n}{w}}\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h \\
&= \frac{1}{\binom{n}{w}} \sum_{h=c}^{n-w+c} \binom{h-1}{c-1}\binom{n-h}{w-c}\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h \leqslant \frac{\binom{n-c}{w-c}}{\binom{n}{w}} \sum_{h=c}^{n-w+c} \binom{h-1}{c-1}\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h \\
&\leqslant \left(\frac{w}{n}\right)^c \sum_{h=c}^{n-w+c} \binom{h-1}{c-1}\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h \leqslant \left(\frac{w}{n}\right)^c \sum_{h=c}^{\infty} \binom{h-1}{c-1}\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^h = \left(\frac{w}{n}\theta^{\mathsf{bm}}_{q,\delta}\right)^c,
\end{aligned}$$

where the first inequality follows from $\binom{n-h}{w-c} \leqslant \binom{n-h}{w-c} \leqslant \binom{n-c}{w-c}$ since $h \geqslant c$, the second inequality follows since $\binom{n-c}{n-w}/\binom{n}{w} = \prod_{j=0}^{c-1} \frac{w-j}{n-j} \leqslant (w/n)^c$, and the last equality follows by the negative-binomial generating function identity with $\theta^{\mathsf{bm}}_{q,\delta} := \Lambda^{\mathsf{bm}}_{q,\delta}/(1-\Lambda^{\mathsf{bm}}_{q,\delta})$, which holds since $\Lambda^{\mathsf{bm}}_{q,\delta} < 1$. Now, substituting this into Eq. (3.12), we obtain

$$\mathop{\mathbb{E}}_{\mathsf{wt}(\mathbf{z}_1)}\left[\left(\Lambda^{\mathsf{bm}}_{q,\delta}\right)^{\mathsf{wt}(\mathbf{z}_1)}\right] \leqslant \sum_{c=\lceil w/2\rceil}^{w} \Pr[C_w = c]\left(\frac{w}{n}\theta^{\mathsf{bm}}_{q,\delta}\right)^c = \mathop{\mathbb{E}}_{C_w}\left[\left(\frac{w}{n}\theta^{\mathsf{bm}}_{q,\delta}\right)^{C_w}\right].$$

Now, define $\theta := (w/n)\theta^{\mathsf{bm}}_{q,\delta}$. By Proposition 3.1, rewrite the expectation above as

$$\begin{aligned}
\mathop{\mathbb{E}}_{C_w}\left[\theta^{C_w}\right] &= \sum_{c=\lceil \frac{w}{2}\rceil}^{w} \frac{cq-w}{c(q-2)}\binom{c}{w-c}\left(\frac{1}{q-1}\right)^{w-c}\left(\frac{q-2}{q-1}\right)^{2c-w}\theta^c \\
&= \sum_{c=\lceil \frac{w}{2}\rceil}^{w} \frac{cq-w}{c(q-2)}\binom{c}{w-c}\left(\frac{\theta}{q-1}\right)^{w-c}\left(\frac{q-2}{q-1}\theta\right)^{2c-w} \leqslant \frac{q-1}{q-2}\sum_{u=0}^{\lfloor w/2\rfloor}\binom{w-u}{u}\sigma^u\gamma^{w-2u},
\end{aligned}$$

where $u := w-c$, $\sigma := \theta/(q-1)$, and $\gamma := ((q-2)/(q-1))\theta$, and the inequality follows since $c \leqslant w$ implies $(cq-w)/(c(q-2)) \leqslant (q-1)/(q-2)$.

Define polynomial

$$F_w(\sigma,\gamma) := \sum_{u=0}^{\lfloor w/2\rfloor}\binom{w-u}{u}\sigma^u\gamma^{w-2u}.$$

Then, this polynomial satisfies the recursion $F_0 = 1$, $F_1 = \gamma$, and $F_w = \gamma F_{w-1} + \sigma F_{w-2}$. Letting ν_0 denote the positive root of the characteristic equation $\nu^2 = \gamma\nu + \sigma$, we have $\nu_0 := (\gamma + \sqrt{\gamma^2 + 4\sigma})/2$.

Claim 3.1. *$F_w(\sigma,\gamma) \leqslant \nu_0^w$ for every $w \geqslant 0$.*

Proof of Claim 3.1. For the base cases, this holds since $F_0(\sigma,\gamma) = 1 = \nu_0^0$ and $F_1(\sigma,\gamma) = \gamma \leqslant \nu_0$. By induction, suppose the claim holds for F_{w-1} and F_{w-2}. Then, using the recurrence for F_w, we have

$$\begin{aligned}
F_w(\sigma,\gamma) &= \gamma F_{w-1}(\sigma,\gamma) + \sigma F_{w-2}(\sigma,\gamma) \leqslant \gamma\nu_0^{w-1} + \sigma\nu_0^{w-2} \\
&= \nu_0^{w-2}(\gamma\nu_0 + \sigma) = \nu_0^w,
\end{aligned}$$

where the last inequality follows from the characteristic equation $\nu_0^2 = \gamma\nu_0 + \sigma$. □

Therefore, we have $\mathbb{E}_{C_w}[\theta^{C_w}] \leqslant \frac{q-1}{q-2}\nu_0^w = \frac{q-1}{q-2}\Phi_q(\theta)^w$ for $\Phi_q(x)$ defined in Theorem 1.2 (obtained by expanding ν_0 by expanding the definitions of σ and γ and treating θ as a variable).

Now, the δ-regime assumed by the theorem ensures $\delta < \frac{1}{4}$. Hence, we have $\min\{k, \lfloor 4\delta k \rfloor\} = \lfloor 4\delta k \rfloor$ and that $\Lambda^{\mathsf{bm}}_{q,\delta} < 1$. Let $L_{k,\delta} := \lfloor 4\delta k \rfloor$. Then, putting everything together from the proof so far, we have

$$\text{Eq. (3.11)} \leqslant \frac{q-1}{(q-2)^2} \sum_{\ell=1}^{L_{k,\delta}} \binom{k}{\ell} (q-1)^\ell \Phi_q\left(\frac{w}{n}\theta^{\mathsf{bm}}_{q,\delta}\right)^w.$$

Setting $w = r\ell$ yields

$$\text{Eq. (3.11)} \leqslant \frac{q-1}{(q-2)^2} \sum_{\ell=1}^{L_{k,\delta}} \binom{k}{\ell} (q-1)^\ell \Phi_q\left(\frac{\ell}{k}\theta^{\mathsf{bm}}_{q,\delta}\right)^{r\ell} = \frac{q-1}{(q-2)^2} \sum_{\ell=1}^{L_{k,\delta}} \mathrm{t}^{(\ell)}_{k,r,q,\delta}. \tag{3.13}$$

If $L_{k,\delta} = 0$, then the sum is vacuous, and the failure probability is 0. Now, define $R_{k,r,q,\delta}$ as 1 if $L_{k,\delta} = 1$, and otherwise $R_{k,r,q,\delta} := (\mathrm{t}^{(L_{k,\delta})}_{k,r,q,\delta}/\mathrm{t}^{(1)}_{k,r,q,\delta})^{\frac{1}{L_{k,\delta}-1}}$ if $L_{k,\delta} \geqslant 2$. To proceed with the proof, we need the following lemma.

Lemma 3.3. *For $\delta \leqslant (1/4)(1-2/r)$, the sequence $\mathcal{T} := (\mathrm{t}^{(\ell)}_{k,r,q,\delta})_{\ell=1}^{L_{k,\delta}}$ is log-convex.*

Proof of Lemma 3.3. For any t_ℓ, extend $\log(\mathrm{t}_\ell)$ to real $\lambda \in [1, L_{k,\delta}]$ as

$$\mathscr{L}(\lambda) := \log\Gamma(k+1) - \log\Gamma(\lambda+1) - \log\Gamma(k-\lambda+1) + \lambda\log(q-1) + r\lambda\log\Phi_q\left(\frac{\lambda}{k}\theta^{\mathsf{bm}}_{q,\delta}\right), \tag{3.14}$$

where Γ denotes the Gamma function. Then, for each integer $\ell \in \{1, \ldots, L_{k,\delta}\}$, we have $\mathscr{L}(\ell) = \log(\mathrm{t}_\ell)$.

Now, to continue the proof, we first prove the following lemma.

Lemma 3.4. *For every $x \geqslant 0$, $\Phi_q(x) = \sqrt{\frac{x}{q-1}}\exp\left[\sinh^{-1}\left(\frac{q-2}{2\sqrt{q-1}}\sqrt{x}\right)\right]$.*

Proof of Lemma 3.4. Let

$$\eta_q := \frac{q-2}{2\sqrt{q-1}}.$$

Using $e^{\sinh^{-1}(x)} = x + \sqrt{1+x^2}$, we obtain

$$\begin{aligned}
\sqrt{\frac{x}{q-1}}e^{\sinh^{-1}(\eta_q\sqrt{x})} &= \sqrt{\frac{x}{q-1}}\left(\eta_q\sqrt{x} + \sqrt{1+\eta_q^2 x}\right) \\
&= \sqrt{\frac{x}{q-1}}\left(\frac{q-2}{2\sqrt{q-1}}\sqrt{x} + \sqrt{1+\frac{x(q-2)^2}{4(q-1)}}\right) \\
&= \sqrt{\frac{x}{q-1}}\left(\frac{q-2}{2\sqrt{q-1}}\sqrt{x} + \frac{q-2}{2\sqrt{q-1}}\sqrt{x}\sqrt{1+\frac{4(q-1)}{x(q-2)^2}}\right) \\
&= \frac{x(q-2)}{2(q-1)}\left(1 + \sqrt{1+\frac{4(q-1)}{x(q-2)^2}}\right) \\
&= \Phi_q(x).
\end{aligned}$$

□

By Lemma 3.4,

$$r\lambda \log \Phi_q\left(\frac{\lambda}{k}\theta^{\mathsf{bm}}_{q,\delta}\right) = \frac{r}{2}\lambda\log\lambda + \beta_{k,q,\delta}\lambda + r\lambda \sinh^{-1}\left(\eta_q\sqrt{\frac{\lambda}{k}\theta^{\mathsf{bm}}_{q,\delta}}\right),$$

where $\beta_{k,q,\delta}$ is a constant independent of λ. Let $\mathfrak{v}_\lambda := \frac{q-2}{2\sqrt{q-1}}\sqrt{\frac{\lambda}{k}\theta^{\mathsf{bm}}_{q,\delta}}$. Twice differentiating Eq. (3.14) gives

$$\begin{aligned}\mathscr{L}''(\lambda) &= -\psi_1(\lambda+1) - \psi_1(k-\lambda+1) + \frac{r}{2\lambda} + \frac{r\mathfrak{v}_\lambda\left(3+2\mathfrak{v}_\lambda^2\right)}{4\lambda\left(1+\mathfrak{v}_\lambda^2\right)^{3/2}} \\ &= -\psi_1(\lambda+1) - \psi_1(k-\lambda+1) + \frac{r}{2\lambda} + \frac{r}{4\lambda}\varphi(\mathfrak{v}_\lambda), \end{aligned} \tag{3.15}$$

where $\varphi(x) := (x(3+2x^2))/(1+x^2)^{3/2}$, and ψ_1 denotes the trigamma function $\psi_1(z) := \frac{d^2}{dz^2}\ln\Gamma(z) = \sum_{j=0}^{\infty}\frac{1}{(z+j)^2}$. We use only the elementary inequality $\psi_1(z+1) < \frac{1}{z}$ for $z \in \mathbb{R}^+$. Applying this inequality to Eq. (3.15), we have

$$\mathscr{L}''(\lambda) > \frac{1}{\lambda}\left(\frac{r}{2} - 1 + \frac{r}{4}\varphi(\mathfrak{v}_\lambda)\right) - \frac{1}{k-\lambda}. \tag{3.16}$$

The function φ is increasing on $[0,\infty)$ since $\varphi'(x) = \frac{3}{(1+x^2)^{5/2}} > 0$. Hence, since $\mathfrak{v}_\lambda > 0$, we have $\varphi(\mathfrak{v}_\lambda) > \varphi(0) = 0$. Putting this in Eq. (3.16), we obtain

$$\mathscr{L}''(\lambda) > \frac{r-2}{2\lambda} - \frac{1}{k-\lambda}.$$

The RHS is nonnegative precisely when $(r-2)(k-\lambda) \geqslant 2\lambda \iff \lambda \leqslant \frac{r-2}{r}k$. Define $\lambda^\star := \frac{r-2}{r}k$. Thus, $\mathscr{L}''(\lambda) \geqslant 0$ for $1 \leqslant \lambda \leqslant \lambda^\star$. It remains to consider the range $\lambda^\star \leqslant \lambda \leqslant L_{k,\delta}$. In this range, $\mathfrak{v}_\lambda \geqslant \frac{q-2}{2\sqrt{q-1}}\sqrt{\frac{\lambda^\star}{k}\theta^{\mathrm{BM}}_{q,\delta}} = \mathfrak{v}^\star_{r,q,\delta}$. By the monotonicity of φ, $\varphi(\mathfrak{v}_\lambda) \geqslant \varphi(\mathfrak{v}^\star_{r,q,\delta})$. Consequently, Eq. (3.16) gives $\mathscr{L}''(\lambda) > \frac{\zeta_{r,q,\delta}}{\lambda} - \frac{1}{k-\lambda}$, where $\zeta_{r,q,\delta} := (r/2) - 1 + (r/4)\varphi(\mathfrak{v}^\star_{r,q,\delta})$. Furthermore, $\lambda \leqslant L_{k,\delta} \leqslant 4\delta k$, so $\lambda - 4\delta\lambda \leqslant 4\delta k - 4\delta\lambda \iff \frac{\lambda}{k-\lambda} \leqslant \frac{4\delta}{1-4\delta}$. But, under the δ-regime assumed for this lemma, we have $\frac{4\delta}{1-4\delta} \leqslant \zeta_{r,q,\delta}$. Thus, $\frac{\lambda}{k-\lambda} \leqslant \zeta_{r,q,\delta} \iff \frac{1}{k-\lambda} \leqslant \frac{\zeta_{r,q,\delta}}{\lambda}$. Therefore, we have $L''(\lambda) \geqslant 0$ whenever $\lambda^\star \leqslant \lambda \leqslant \mathcal{L}_{k,\delta}$. Therefore, $\mathscr{L}$ is convex throughout $[1, L_{k,\delta}]$. Thus, the sequence $(\log \mathfrak{t}_\ell)_{\ell=1}^{L_{k,\delta}}$ is convex. Therefore, the sequence $(\mathfrak{t}_\ell)_{\ell=1}^{L_{k,\delta}}$ is log-convex whenever $\delta \leqslant \zeta_{r,q,\delta}/(4(1+\zeta_{r,q,\delta}))$.

Finally, we show that $(1/4)(1-2/r) \leqslant \zeta_{r,q,\delta}/(4(1+\zeta_{r,q,\delta}))$. Notice that $\varphi(x) \geqslant 0$ whenever $x \geqslant 0$. Moreover, $\mathfrak{v}^\star_{r,q,\delta} > 0$ by definition. Now, the above desired condition is equivalent to showing $r \leqslant 2(1+\zeta_{r,q,\delta})$. By definition, $1+\zeta_{r,q,\delta} = \frac{r}{2} + \frac{r}{4}\varphi(\mathfrak{v}^\star_{r,q,\delta}) = \frac{r}{2}(1+\frac{1}{2}\varphi(\mathfrak{v}^\star_{r,q,\delta}))$. Therefore, $2(1+\zeta_{r,q,\delta}) = r(1+\frac{1}{2}\varphi(\mathfrak{v}^\star_{r,q,\delta})) \geqslant r$ holds since $\varphi(\mathfrak{v}^\star_{r,q,\delta}) \geqslant 0$. This completes the proof. □

By Lemma 3.3, the sequence $(\mathfrak{t}^{(\ell)}_{k,r,q,\delta})_{\ell=1}^{L_{k,\delta}}$ is log-convex. Equivalently, the function $\mathscr{L} : [1, L_{k,\delta}]_{\mathbb{R}} \to \mathbb{R}$ defined in Eq. (3.14), whose restriction to the integers satisfies $\mathscr{L}(\ell) = \log \mathfrak{t}^{(\ell)}_{k,r,q,\delta}$ for $\ell \in [L_{k,\delta}]$, is convex. Hence, its graph lies below the chord joining its values at 1 and $L_{k,\delta}$. Thus, the chord bound on convex functions yields for every $\lambda \in [1, L_{k,\delta}]_{\mathbb{R}}$:

$$\mathscr{L}(\lambda) \leqslant \frac{L_{k,\delta}-\lambda}{L_{k,\delta}-1}\mathscr{L}(1) + \frac{\lambda-1}{L_{k,\delta}-1}\mathscr{L}(L_{k,\delta}).$$

Setting $\lambda = \ell \in \mathbb{Z}$ and exponentiating the above inequality, we have

$$\begin{aligned}\mathrm{t}^{(\ell)}_{k,r,q,\delta} \leqslant \left(\mathrm{t}^{(1)}_{k,r,q,\delta}\right)^{\frac{L_{k,\delta}-\ell}{L_{k,\delta}-1}} \left(\mathrm{t}^{(L_{k,\delta})}_{k,r,q,\delta}\right)^{\frac{\ell-1}{L_{k,\delta}-1}} &= \mathrm{t}^{(1)}_{k,r,q,\delta} \left(\frac{\mathrm{t}^{(L_{k,\delta})}_{k,r,q,\delta}}{\mathrm{t}^{(1)}_{k,r,q,\delta}}\right)^{\frac{\ell-1}{L_{k,\delta}-1}} \\ &= \mathrm{t}^{(1)}_{k,r,q,\delta} R^{\ell-1}_{k,r,q,\delta}.\end{aligned}$$

Substituting this into Eq. (3.13) gives

$$\text{Eq. (3.11)} \leqslant \frac{q-1}{(q-2)^2} \sum_{\ell=1}^{L_{k,\delta}} \mathrm{t}^{(1)}_{k,r,q,\delta} R^{\ell-1}_{k,r,q,\delta} = \frac{q-1}{(q-2)^2} \mathrm{t}^{(1)}_{k,r,q,\delta} \sum_{\ell'=0}^{L_{k,\delta}-1} R^{\ell'}_{k,r,q,\delta}. \tag{3.17}$$

The sum in the above RHS is the sum of the first $L_{k,\delta}$ terms of the geometric progression with first term 1 and common ratio $R_{k,r,q,\delta}$. Define this sum as

$$\mathcal{G}_{k,r,q,\delta} := \begin{cases} \frac{1-R^{L_{k,\delta}}_{k,r,q,\delta}}{1-R_{k,r,q,\delta}}, & R_{k,r,q,\delta} \neq 1, \\ L_{k,\delta}, & R_{k,r,q,\delta} = 1 \end{cases}.$$

Substituting this into Eq. (3.17) yields the desired upper bound

$$\Pr[\text{failure}] \leqslant \frac{q-1}{(q-2)^2} \mathrm{t}^{(1)}_{k,r,q,\delta} \mathcal{G}_{k,r,q,\delta} = \left(\frac{q-1}{q-2}\right)^2 \cdot k \cdot \Phi_q\left(\frac{\theta_{q,\delta}}{k}\right)^r \cdot \mathcal{G}_{k,r,q,\delta}. \qquad \square$$

3.2.2 Chernoff Strategy

We now prove the Chernoff Strategy. In order to prove Theorem 1.3, we first establish an auxiliary lemma.

Lemma 3.5. *For all $0 < a \leqslant 1$ and $b > 0$, the function Φ_q satisfies*

$$\Phi_q(ab) \leqslant \sqrt{a}\Phi_q(b).$$

Proof.

$$\begin{aligned}\Phi_q(ab) &= \frac{ab(q-2)}{2(q-1)}\left(1 + \sqrt{1 + \frac{4(q-1)}{ab(q-2)^2}}\right) \\ &= \frac{\sqrt{a}b(q-2)}{2(q-1)}\left(\sqrt{a} + \sqrt{\sqrt{a} + \frac{4(q-1)}{b(q-2)^2}}\right) \\ &\leqslant \frac{\sqrt{a}b(q-2)}{2(q-1)}\left(1 + \sqrt{1 + \frac{4(q-1)}{b(q-2)^2}}\right) && (\because a \leqslant 1) \\ &= \sqrt{a}\Phi_q(b).\end{aligned}$$

$\square$

Proof of Theorem 1.3. Since $\delta < \frac{1}{2}$, $\Lambda_\delta^{\mathsf{ch}} < 1$. Therefore, we can replace $\Lambda_{q,\delta}^{\mathsf{bm}}$ in the proof of Theorem 1.2 with $\Lambda_\delta^{\mathsf{ch}}$ and obtain the same exact bounds, with $\theta_{q,\delta}^{\mathsf{bm}}$ now replaced with $\theta_\delta^{\mathsf{ch}} = \Lambda_\delta^{\mathsf{ch}}/(1-\Lambda_\delta^{\mathsf{ch}})$, and obtain the following upper bound:

$$\text{Eq. (3.11)} \leqslant \frac{1}{q-2}\sum_{\ell=1}^{U_{k,\delta}} \binom{k}{\ell}(q-1)^\ell \Phi_q\left(\frac{w}{n}\theta_\delta^{\mathsf{ch}}\right)^w.$$

Putting $w = r\ell$, we have

$$\text{Eq. (3.11)} \leqslant \frac{1}{q-2}\sum_{\ell=1}^{U_{k,\delta}} \binom{k}{\ell}(q-1)^\ell \Phi_q\left(\frac{\ell}{k}\theta_\delta^{\mathsf{ch}}\right)^{r\ell},$$

Now, using Lemma 3.5 with $a = \frac{\ell}{k} < 1$ and $b = \theta_\delta^{\mathsf{ch}}$, we have

$$\begin{aligned}\Pr[\text{failure}] &\leqslant \frac{1}{q-2}\sum_{\ell=1}^{U_{k,\delta}} \binom{k}{\ell}\left(\frac{\ell}{k}\right)^{\frac{r\ell}{2}}\left[(q-1)\Phi_q\left(\theta_\delta^{\mathsf{ch}}\right)^r\right]^\ell \\ &= \frac{1}{q-2}\sum_{\ell=1}^{U_{k,\delta}} \binom{k}{\ell}\left(\frac{\ell}{k}\right)^{\frac{r\ell}{2}} \phi_{r,q,\delta}^\ell.\end{aligned}$$

For all $\ell \in \{1,\dots,U_{k,\delta}\}$, let

$$\mathfrak{s}_{k,r,q,\delta}^{(\ell)} := \binom{k}{\ell}\left(\frac{\ell}{k}\right)^{\frac{r\ell}{2}} \phi_{r,q,\delta}^\ell.$$

Then, for all $\ell \in \{1,\dots,U_{k,\delta}-1\}$,

$$\begin{aligned}\frac{\mathfrak{s}_{k,r,q,\delta}^{(\ell+1)}}{\mathfrak{s}_{k,r,q,\delta}^{(\ell)}} &= \frac{\binom{k}{\ell+1}\left(\frac{\ell+1}{k}\right)^{\frac{r(\ell+1)}{2}}\phi_{r,q,\delta}^{\ell+1}}{\binom{k}{\ell}\left(\frac{\ell}{k}\right)^{\frac{r\ell}{2}}\phi_{r,q,\delta}^\ell} \\ &= \frac{k-\ell}{\ell+1}\left(\frac{\ell+1}{k}\right)^{\frac{r}{2}}\left[\left(1+\frac{1}{\ell}\right)^\ell\right]^{\frac{r}{2}}\phi_{r,q,\delta} \\ &= \left(1-\frac{\ell}{k}\right)\left(\frac{\ell+1}{k}\right)^{\frac{r}{2}-1}\left[\left(1+\frac{1}{\ell}\right)^\ell\right]^{\frac{r}{2}}\phi_{r,q,\delta}.\end{aligned}$$

Now, we use the following three estimates to bound the above ratio:

$$\left(1+\frac{1}{\ell}\right)^\ell < e, \qquad \frac{\ell+1}{k} \leqslant \frac{U_{k,\delta}}{k} \leqslant m_\delta, \qquad 1-\frac{\ell}{k} \leqslant 1.$$

Therefore, since $r > 2$,

$$\frac{\mathfrak{s}_{k,r,q,\delta}^{(\ell+1)}}{\mathfrak{s}_{k,r,q,\delta}^{(\ell)}} \leqslant e^{\frac{r}{2}} m_\delta^{\frac{r}{2}-1}\phi_{r,q,\delta} =: Q_{r,q,\delta}.$$

Iterating, we obtain

$$\begin{aligned}\mathfrak{s}^{(\ell)}_{k,r,q,\delta} &\leqslant \mathfrak{s}^{(1)}_{k,r,q,\delta} Q^{\ell-1}_{r,q,\delta} \\ &= k^{1-\frac{r}{2}} \phi_{r,q,\delta} Q^{\ell-1}_{r,q,\delta}.\end{aligned}$$

Thus, the failure probability satisfies

$$\begin{aligned}\Pr[\text{failure}] &\leqslant \frac{1}{q-2} \sum_{\ell=1}^{U_{k,\delta}} \mathfrak{s}^{(\ell)}_{k,r,q,\delta} \\ &\leqslant \frac{k^{1-\frac{r}{2}} \phi_{r,q,\delta}}{q-2} \sum_{\ell=0}^{U_{k,\delta}-1} Q^{\ell}_{r,q,\delta}.\end{aligned}$$

The remaining sum is a geometric sum. Hence, if we define

$$\mathcal{S}_{k,r,q,\delta} := \sum_{\ell=0}^{U_{k,\delta}-1} Q^{\ell}_{r,q,\delta} = \begin{cases} \frac{1-Q^{U_{k,\delta}}_{r,q,\delta}}{1-Q_{r,q,\delta}}, & Q_{r,q,\delta} \neq 1, \\ U_{k,\delta}, & Q_{r,q,\delta} = 1, \end{cases}$$

then, we get the closed-form bound

$$\Pr[\text{failure}] \leqslant \frac{k^{1-r/2} \phi_{r,q,\delta} \mathcal{S}_{k,r,q,\delta}}{q-2}. \qquad \square$$

3.3 The Expand-Accumulate Code

We now turn to the original expand-accumulate (EA) code, due to Boyle et al. [BCG+22]. Let $\rho \in (0,1)$ denote the rate, $k \in \mathbb{Z}^+$ be the message length, $n = k/\rho$ be the codeword length, and let $p \in (0,1)$. Then, the EA code defines $\mathbf{M} = \mathbf{E} \in \mathbb{F}_q^{k \times n}$, where each entry of $\mathbf{E}$ is i.i.d. sampled from $\mathsf{Ber}_p(\mathbb{F}_q)$. Now, for input $\mathbf{x} \in \mathbb{F}_q^k$, we are interested in bounding the following probability: $\Pr_{\mathbf{E}}[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell]$. This probability follows the Binomial distribution with probability $\kappa_\ell := (1 - 1/q) \cdot (1 - (1 - p \cdot (q/(q-1)))^\ell)$, n trials, and exactly w successes. Therefore,

$$\Pr_{\mathbf{E}}[\mathsf{wt}(\mathbf{y}) = w | \mathsf{wt}(\mathbf{x}) = \ell] = \binom{n}{w} \kappa_\ell^w (1 - \kappa_\ell)^{n-w}.$$

Now, we analyze the random variables C_w and T_w for $\mathbf{y} = \mathbf{x}\mathbf{E}$, conditioned on $\mathsf{wt}(\mathbf{y}) = w$. Observe that T_w is identically distributed as before, i.e.,

$$\Pr[T_w = h - c | C_w = c] = \frac{\binom{h-1}{c-1}\binom{n-h}{w-c}}{\binom{n}{w}}.$$

This is because that for any permutation $\mathbf{Q} \in \{0,1\}^{n \times n}$, the distributions $\mathbf{x}\mathbf{E}$ and $\mathbf{x}\mathbf{E}\mathbf{Q}$ are identically distributed. Thus,

$$\Pr[\mathsf{wt}(\mathbf{y}\mathbf{A}) \leqslant \delta n | \mathsf{wt}(\mathbf{y}) = w] \leqslant \sum_{h=\lceil \frac{w}{2} \rceil}^{\lfloor \delta n \rfloor} \sum_{c=\lceil \frac{w}{2} \rceil}^{w} \Pr[C_w = c] \Pr[T_w = h - c | C_w = c]$$

$$= \sum_{c=\lceil w/2 \rceil}^{w} \Pr[C_w = c] \sum_{h=c}^{\lfloor \delta n \rfloor} \frac{\binom{h-1}{c-1} \cdot \binom{n-h}{w-c}}{\binom{n}{w}} = \sum_{c=\lceil w/2 \rceil}^{\min(w, 2\lfloor \delta n \rfloor)} \Pr[C_w = c] \Pr[\mathscr{H}_w \geqslant c].$$

Moving to C_w, since $\mathbf{y} = \mathbf{x}\mathbf{E}$ is Binomially distributed over $\mathbb{F}_q$ with probability κ_ℓ, it is also distributed as $\mathsf{Ber}_{\kappa_\ell}(\mathbb{F}_q)^n$. Thus, conditioned on $\mathsf{wt}(\mathbf{y}) = w$, this implies that the non-zeros of $\mathbf{y}$ are uniformly distributed over $\mathbb{F}_q \setminus \{0\}$. This implies that C_w is also identically distributed as before, which means we can apply Proposition 3.1. All together, we can rewrite and upper bound Eq. (3.2) as

$$\sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{w=1}^{2\lfloor \delta n \rfloor} \binom{n}{w} \kappa_\ell^w (1-\kappa_\ell)^{n-w} \sum_{c=\lceil w/2 \rceil}^{w} \Pr[C_w = c] \Pr[\mathscr{H}_w \geqslant c].$$

Define $\Lambda^{\mathsf{bm}}_{q,\delta} = 2\sqrt{\delta/(q-1)} + (1 - 1/(q-1))\delta$ as before. Then, by Lemma 3.1 and the proof of Theorem 1.1 in Section 3.1, we have $\Pr[\mathscr{H}_w \geqslant c] \leqslant \binom{w}{c}\delta^c$, and thus upper bound $\sum_{c=\lceil w/2 \rceil}^{w} \Pr[C_w = c]\binom{w}{c}\delta^c \leqslant ((q-1)/(q-2))(\Lambda^{\mathsf{bm}}_{q,\delta})^w$. Thus, we can rewrite the above union bound as

$$\leqslant \sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^{\ell-1} \sum_{w=1}^{2\lfloor \delta n \rfloor} \binom{n}{w} \kappa_\ell^w (1-\kappa_\ell)^{n-w} \left(\frac{q-1}{q-2}\right) (\Lambda^{\mathsf{bm}}_{\delta,q})^w.$$

Focusing on the inner w-sum, we can upper bound it as

$$\begin{aligned} &\sum_{w=1}^{2\lfloor \delta n \rfloor} \binom{n}{w} \kappa_\ell^w (1-\kappa_\ell)^{n-w} \left(\tfrac{q-1}{q-2}\right) (\Lambda^{\mathsf{bm}}_{\delta,q})^w \leqslant \tfrac{q-1}{q-2} \sum_{w=0}^{n} \binom{n}{w} (\kappa_\ell \Lambda^{\mathsf{bm}}_{q,\delta})^w (1-\kappa_\ell)^{n-w} \\ =&\tfrac{q-1}{q-2}(1 - \kappa_\ell + \kappa_\ell \Lambda^{\mathsf{bm}}_{q,\delta})^n = \tfrac{q-1}{q-2}[1 - (1 - \Lambda^{\mathsf{bm}}_{q,\delta})\kappa_\ell]^n. \end{aligned}$$

Combining the above with Eq. (3.2), we now analyze

$$\mathcal{S}_k := \frac{1}{q-2} \sum_{\ell=1}^{k} \binom{k}{\ell} (q-1)^\ell [1 - (1 - \Lambda^{\mathsf{bm}}_{q,\delta})\kappa_\ell]^n.$$

Proof of Theorem 1.4. Assume that $\Lambda^{\mathsf{bm}}_{q,\delta} < 1$. Then, for $\mu_{q,\delta}$, $\eta_{p,q}$, and $\nu_{p,q,\delta}$ defined in Theorem 1.4, observe that $1-(1-\Lambda^{\mathsf{bm}}_{q,\delta})\kappa_\ell = 1-\mu_{q,\delta}(1-\eta^\ell_{p,q})$. Now, if $\eta_{p,q} = 0$, we have $\mathcal{S}_k = \frac{(1-\mu_{q,\delta})^n}{q-2} \sum_{\ell=1}^{k} \binom{k}{\ell}(q-1)^\ell = \frac{q^k-1}{q-2}(1-\mu_{q,\delta})^n$, showing the case for $\eta_{p,q} = 0$. We now assume that $\eta_{p,q} > 0$ and remark that for $p > 0$, $\eta_{p,q} < 1$. Then, the binomial expansion of the final factor of $\mathcal{S}_k$ gives us

$$\left(1 - \mu_{q,\delta} + \mu_{q,\delta}\eta^\ell_{p,q}\right)^n = \sum_{j=0}^{n} \binom{n}{j} (1-\mu_{q,\delta})^{n-j} \mu^j_{q,\delta} \eta^{j\ell}_{p,q}.$$

Substituting into the definition of $\mathcal{S}_k$ yields

$$\begin{aligned} \mathcal{S}_k &= \frac{1}{q-2} \sum_{j=0}^{n} \binom{n}{j} (1-\mu_{q,\delta})^{n-j} \mu^j_{q,\delta} \sum_{\ell=1}^{k} \binom{k}{\ell} \left((q-1)\eta^j_{p,q}\right)^\ell \\ &= \frac{1}{q-2} \sum_{j=0}^{n} \binom{n}{j} (1-\mu_{q,\delta})^{n-j} \mu^j_{q,\delta} \left[\left(1 + (q-1)\eta^j_{p,q}\right)^k - 1\right]. \end{aligned} \tag{3.18}$$

Thus, if $J \leftarrow \mathsf{Bin}(n, \mu_{q,\delta})$, then $\mathcal{S}_k = \frac{1}{q-2}\mathbb{E}_{J\sim\mathsf{Bin}(n,\mu_{q,\delta})}[(1+(q-1)\eta_{p,q}^J)^k - 1]$. Now, we partition the expectation in Eq. (3.18) into three regimes: $J = 0$, $1 \leqslant J < \sigma_{k,p,q}$, and $\sigma_{k,p,q} \leqslant J \leqslant n$. For $J = 0$, the contribution is exactly

$$\mathop{\mathbb{E}}_{J\sim\mathsf{Bin}(n,\mu_{q,\delta})}\left[\left(\left(1+(q-1)\eta_{p,q}^J\right)^k - 1\right)\mathbb{I}(J=0)\right] = \left(q^k-1\right)(1-\mu_{q,\delta})^n. \tag{3.19}$$

For analyzing the regime $1 \leqslant J < \sigma_{k,p,q}$, define the function $f(x) := \ln\left(1 + (q-1)\eta_{p,q}^x\right)$. Twice differentiating yields $f''(x) = \frac{(q-1)(\ln\eta_{p,q})^2\eta_{p,q}^x}{\left(1+(q-1)\eta_{p,q}^x\right)^2} \geqslant 0$. Hence, f is convex. The chord inequality applied on $[0, \sigma_{k,p,q}]_{\mathbb{R}}$ gives

$$f(x) \leqslant [1 - \tfrac{x}{\sigma_{k,p,q}}]\ln q + \tfrac{x}{\sigma_{k,p,q}}\ln\left(1+(q-1)\eta_{p,q}^{\sigma_{k,p,q}}\right) = \ln\left[q\left[\frac{1+(q-1)\eta_{p,q}^{\sigma_{k,p,q}}}{q}\right]^{\frac{x}{\sigma_{k,p,q}}}\right],$$

and therefore $1 + (q-1)\eta_{p,q}^x \leqslant q\left(\frac{1+(q-1)\eta_{p,q}^{\sigma_{k,p,q}}}{q}\right)^{x/\sigma_{k,p,q}}$, which implies $\left(1 + (q-1)\eta_{p,q}^x\right)^k \leqslant q^k\theta_{k,p,q}^x$. It follows that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{J\sim\mathsf{Bin}(n,\mu_{q,\delta})}\left[\left(\left(1+(q-1)\eta_{p,q}^J\right)^k - 1\right)\mathbb{I}(1 \leqslant J < \sigma_{k,p,q})\right] \\
&\quad\leqslant q^k\,\mathbb{E}\left[\theta_{k,p,q}^J\mathbb{I}(1 \leqslant J < \sigma_{k,p,q})\right] \leqslant q^k\left(\mathbb{E}\left[\theta_{k,p,q}^J\right] - \Pr[J=0]\right) \\
&\quad= q^k\left((1-\mu_{q,\delta}+\mu_{q,\delta}\theta_{k,p,q})^n - (1-\mu_{q,\delta})^n\right).
\end{aligned} \tag{3.20}$$

For $\sigma_{k,p,q} \leqslant J \leqslant n$, by the definition of $\sigma_{k,p,q}$, we have $k(q-1)\eta_{p,q}^J \leqslant k(q-1)\eta_{p,q}^{\sigma_{k,p,q}} \leqslant 1$. Further, using the inequalities $(1+x)^k \leqslant e^{kx}$ and $e^x - 1 \leqslant ex$ for $x \in [0,1]$, we obtain $\left(1+(q-1)\eta_{p,q}^J\right)^k - 1 \leqslant ek(q-1)\eta_{p,q}^J$. Consequently,

$$\begin{aligned}
\mathop{\mathbb{E}}_{J\sim\mathsf{Bin}(n,\mu_{q,\delta})}\left[\left(\left(1+(q-1)\eta_{p,q}^J\right)^k - 1\right)\mathbb{I}(J \geqslant \sigma_{k,p,q})\right] &\leqslant ek(q-1)\mathbb{E}\left[\eta_{p,q}^J\right] \\
&= ek(q-1)\nu_{p,q,\delta}^n.
\end{aligned} \tag{3.21}$$

Finally, from Eqs. (3.18) to (3.21), we have

$$\begin{aligned}
\text{Eq. (3.2)} &\leqslant \frac{(q^k-1)(1-\mu_{q,\delta})^n + q^k\left(\left(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_{k,p,q}\right)^n - (1-\mu_{q,\delta})^n\right) + ek(q-1)\nu_{p,q,\delta}^n}{q-2} \\
&= \frac{q^k\left(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_{k,p,q}\right)^n - (1-\mu_{q,\delta})^n + ek(q-1)\nu_{p,q,\delta}^n}{q-2}.
\end{aligned}$$

Finally, if $\sigma_{k,p,q} \leqslant 1$, the middle regime $1 \leqslant J < \sigma_{k,p,q}$ collapses, leaving only the cases $J = 0$ and $1 = \sigma_{k,p,q} \leqslant J \leqslant n$, which gives the bound for this case. □ □

Remark 3.2. For the above proof, replacing $\Lambda_{q,\delta}^{\mathsf{bm}}$ with $\Lambda_\delta^{\mathsf{ch}} := 2\sqrt{\delta(1-\delta)}$ such that $\Lambda_\delta^{\mathsf{ch}} < 1$ results in an identical proof for the Chernoff version of Theorem 1.4 for $\delta \in (0, 1/2)$.

3.4 Extensions to Other Turbo-like Codes

Here, we briefly highlight how to apply our analysis to some other Turbo-like codes in the literature. We note that (1) this list is not extensive, but just a sampling of some turbo-like codes, and (2) we leave a formal treatment of these codes and others in our framework as important future work.

Encode-Repeat-Accumulate Codes. For repetition factor r, message length k, and target codeword length n, the ERA code of Baweja et al. [BFM+26] leverages an outer code $C_{\text{out}} \colon \mathbb{F}^k \to \mathbb{F}^{n/r}$ such that C_{out} has large minimum distance. Then, the ERA code encodes any $\mathbf{x} \in \mathbb{F}^k$ by first obtaining $\mathbf{y}' = C_{\text{out}}(\mathbf{x})$, then encoding $\mathbf{y}'$ using the RAA code. One key efficiency improvement of the ERA code is that $\mathbf{P}_1$ and $\mathbf{P}_2$ only need independent permutation matrices $\mathbf{Q}_1, \mathbf{Q}_2$, and share the same random full-rank diagonal matrix $\mathbf{T}$.

The analysis of the ERA code in our framework is relatively straightforward. First, let $\mathbf{G}$ denote the generator matrix of C_{out} and let δ_{out} be its minimum distance. Then, set $\mathbf{M} = \mathbf{GRP}_1\mathbf{AP}_2$. This allows us to analyze the ERA code under Eq. (3.2). Let $\mathbf{y}_1 = \mathbf{xGR}$, $\mathbf{z}_1 = \mathbf{y}_1\mathbf{A}$, $\mathbf{y}_2 = \mathbf{z}_1\mathbf{P}_2$, and $\mathbf{z}_2 = \mathbf{y}_2\mathbf{A}$. Now, by definition of C_{out} and $\mathbf{R}$, $\mathsf{wt}(\mathbf{y}_1) \geqslant \delta_{\text{out}} n$.

We can re-write the inner w-sum of Eq. (3.2) as

$$= \sum_{w'=\delta_{\text{out}} n}^{4\lfloor \delta n \rfloor} \sum_{w=\lceil w'/2 \rceil}^{2\lfloor \delta n \rfloor} \Pr[\mathsf{wt}(\mathbf{z}_1) = w | \mathsf{wt}(\mathbf{y}_1) = w'] \Pr[\mathsf{wt}(\mathbf{y}_1) = w' | \mathsf{wt}(\mathbf{x}) = \ell].$$

Then, the analysis proceeds by analyzing the above quantities with respect to the code C_{out}, nearly identically to the RAA code, except with the following two important differences: (1) under the ERA code, the permutations $\mathbf{P}_1$ and $\mathbf{P}_2$ are not independent, which must be carefully handled in the context of the above bound; and (2) the output weight distribution of C_{out} must be understood (or can be estimated) in order to understand the probability $\mathsf{wt}(\mathbf{y}_1) = w'$.

Expand-Convolute Codes. Expand-convolute codes are defined identically to our expand-accumulate codes of Section 3.3, except now the accumulator matrix $\mathbf{A}$ is replaced with a random convolution matrix $\mathbf{C}$ with state-size m. That is, for every $i \in [n]$, there is a random subset $\Psi_i \subseteq [i]$ of size at most m which also contains i such that $\mathbf{z} = \mathbf{yC}$ satisfies $z_i = \sum_{j \in \Psi_i} y_j$. In our framework, the random convolution changes the distribution of the random variables C_w and T_w, since now the partial sums Ψ_i depend on a random $\leqslant m$-sized subset S_i of the non-zeros of $\mathbf{y}$. In particular, the gap vector we analyzed for T_w also changes depending on the subset Ψ_i. However, once these random variables are analyzed, our framework admits an analysis for non-binary expand-convolute codes.

4 Asymptotic Analysis of the RA, RAA, and EA Codes

In this section, we prove the asymptotic theorems presented in this work. In particular, we prove Corollaries 1.1, 1.2 and 1.5.

4.1 RA Code

Proof of Corollary 1.1. Let $\tau = n^{-\beta}$ for $\beta > 0$. Then, from Theorem 1.1, for the BM case, consider η_{bm}:

$$\eta_{\mathsf{bm}} = \left(\frac{[1 + (q-2)\tau]^{1/k} - 1}{q-1} \right)^{1/r} = \left(\frac{[1 + (q-2)n^{-\beta}]^{1/k} - 1}{q-1} \right)^{1/r}.$$

Since k, and thus n, are sufficiently large, we apply the Binomial approximation $(1+x)^r = 1 + rx(1+o(1))$ and obtain

$$\left(\frac{[1+(q-2)n^{-\beta}]^{1/k}-1}{q-1}\right)^{1/r} = \left(\frac{q-2}{q-1}\frac{n^{-\beta}}{k}(1+o(1))\right)^{1/r}$$
$$= \left(\frac{r(q-2)}{q-1}n^{-\beta-1}(1+o(1))\right)^{1/r} = \left[\frac{r(q-2)}{q-1}\right]^{1/r} n^{-(\beta+1)/r}(1+o(1)).$$

Therefore, by Theorem 1.1,

$$\begin{aligned}
\delta &\leqslant \frac{q-1}{(q-2)^2}\left(\sqrt{1+(q-2)\eta_{\mathsf{bm}}}-1\right)^2 \\
&= \frac{q-1}{(q-2)^2}\left(\sqrt{1+(q-2)\left[\frac{r(q-2)}{q-1}\right]^{1/r} n^{-(\beta+1)/r}(1+o(1))}-1\right)^2 \\
&= \left[\frac{\sqrt{q-1}}{q-2}\left(\frac{q-2}{2}\left[\frac{r(q-2)}{q-1}\right]^{1/r} n^{-(\beta+1)/r}(1+o(1))\right)(1+o(1))\right]^2 \\
&= \frac{r^{2/r}}{4}(q-2)^{2/r}(q-1)^{1-2/r}n^{-2(\beta+1)/r}(1+o(1)).
\end{aligned}$$

Now, similarly for the Chernoff case, consider η_{ch}:

$$\eta_{\mathsf{ch}} = \left(\frac{[1+(q-1)\tau]^{1/k}-1}{q-1}\right)^{2/r} = \left(\frac{[1+(q-1)n^{-\beta}]^{1/k}-1}{q-1}\right)^{2/r}.$$

Again applying the Binomial approximation, we have

$$\left(\frac{[1+(q-1)n^{-\beta}]^{1/k}-1}{q-1}\right)^{2/r} = \left(\frac{n^{-\beta}}{k}(1+o(1))\right)^{2/r}$$
$$= \left(rn^{-\beta-1}(1+o(1))\right)^{2/r} = r^{2/r}n^{-2(\beta+1)/r}(1+o(1)).$$

Therefore,

$$\begin{aligned}
\delta &\leqslant \frac{(1-\sqrt{1-\eta_{\mathsf{ch}}})}{2} \\
&= \frac{(1-\sqrt{1-r^{2/r}n^{-2(\beta+1)/r}(1+o(1))})}{2} = \frac{r^{2/r}}{4}n^{-2(\beta+1)/r}(1+o(1)).
\end{aligned}$$

□

Proof of Corollary 1.2. For fixed $r, q > 2$,

$$\delta = o\left(n^{-2/r}\right) \implies \left(\Lambda^{\mathrm{BM}}_{q,\delta}\right)^r = o\left(n^{-1}\right)$$
$$\implies \frac{\left[1+(q-1)\left(\Lambda^{\mathrm{BM}}_{q,\delta}\right)^r\right]^{n/r}-1}{q-2} = o(1).$$

Substituting the above result in Theorem 1.1 gives the desired bound. A direct corollary of this result is the following, which completes the proof of Corollary 1.2. □

Corollary 4.1. *In the ensemble of length-n RA codes with repetition factor $r > 2$, finite field size $q > 2$, and $n \to \infty$, almost all codes have minimum distance lower bounded by*

$$d_{min}^{\mathsf{RA}} > n^{\frac{r-2}{r}-\varepsilon},$$

where ε can be made arbitrarily small.

Proof. By definition,

$$\text{failure} \iff d_{\min}^{\mathsf{RA}} \leqslant \delta n.$$

Thus, by the preceding proof,

$$\lim_{n\to\infty} \Pr\left[d_{\min}^{\mathsf{RA}} \leqslant n^{\frac{r-2}{r}-\varepsilon}\right] = 0,$$

where ε can be made arbitrarily small. □

4.2 RAA Code

We now use our bounds to establish linear minimum-distance growth of the RAA ensemble. Throughout this subsection, let $r, q > 2$ be fixed, and let $n = rk \to \infty$.

4.2.1 BM Strategy

In this section, we prove Corollary 1.3. Recall the quantities defined in the statement of Theorem 1.2.

Lemma 4.1 (Limit of $R_{k,r,q,\delta}$)**.** *Define*

$$\mathscr{R}_{r,q,\delta} := \exp\left(\frac{H(4\delta)}{4\delta}\right)(q-1)\Phi_q\left(4\delta\theta_{q,\delta}^{\mathsf{bm}}\right)^r, \tag{4.1}$$

where

$$H(x) := -x\ln x - (1-x)\ln(1-x)$$

denotes the binary entropy function. If $\Lambda_{q,\delta}^{\mathsf{bm}} < 1$, then

$$\lim_{n\to\infty} R_{k,r,q,\delta} = \mathscr{R}_{r,q,\delta}.$$

Proof. We have

$$\mathrm{t}_{k,r,q,\delta}^{(L_{k,\delta})} := \binom{k}{L_{k,\delta}}(q-1)^{L_{k,\delta}}\Phi_q\left(\frac{L_{k,\delta}}{k}\theta_{q,\delta}^{\mathsf{bm}}\right)^{rL_{k,\delta}}.$$

Taking logarithms and dividing by $L_{k,\delta}$,

$$\frac{1}{L_{k,\delta}}\ln \mathrm{t}_{k,r,q,\delta}^{(L_{k,\delta})} = \frac{1}{L_{k,\delta}}\ln\binom{k}{L_{k,\delta}} + \ln(q-1) + r\ln\Phi_q\left(\frac{L_{k,\delta}}{k}\theta_{q,\delta}^{\mathsf{bm}}\right). \tag{4.2}$$

By the standard entropy asymptotics for the binomial coefficient,

$$\ln \binom{k}{L_{k,\delta}} = kH\left(\frac{L_{k,\delta}}{k}\right) + O(\ln k).$$

Consequently,

$$\frac{1}{L_{k,\delta}} \ln \binom{k}{L_{k,\delta}} = \frac{k}{L_{k,\delta}} H\left(\frac{L_{k,\delta}}{k}\right) + O\left(\frac{\ln k}{k}\right).$$

Moreover, we have

$$\lim_{k\to\infty} \frac{L_{k,\delta}}{k} = 4\delta, \quad \lim_{k\to\infty} \frac{k}{L_{k,\delta}} = \frac{1}{4\delta}, \quad \lim_{k\to\infty} \frac{\ln k}{k} = 0.$$

Hence, from the two above equations, we get

$$\lim_{k\to\infty} \frac{1}{L_{k,\delta}} \ln \binom{k}{L_{k,\delta}} = \frac{H(4\delta)}{4\delta}.$$

Since Φ_q is continuous on $(0,\infty)$,

$$\lim_{k\to\infty} \Phi_q\left(\frac{L_{k,\delta}}{k}\theta^{\mathsf{bm}}_{q,\delta}\right) = \Phi_q\left(4\delta\theta^{\mathsf{bm}}_{q,\delta}\right).$$

Thus, it follows from the two above equations and Eq. (4.2) that

$$\lim_{k\to\infty} \frac{1}{L_{k,\delta}} \ln \mathfrak{t}^{(L_{k,\delta})}_{k,r,q,\delta} = \ln \mathscr{R}_{r,q,\delta}.$$

On the other hand,

$$\mathfrak{t}^{(1)}_{k,r,q,\delta} = k(q-1)\Phi_q\left(\frac{\theta^{\mathsf{bm}}_{q,\delta}}{k}\right)^r,$$

and as $x \to 0$,

$$\Phi_q(x) = \sqrt{\frac{x}{q-1}} + O(x).$$

Therefore,

$$\mathfrak{t}^{(1)}_{k,r,q,\delta} = (q-1)^{1-\frac{r}{2}}\left(\theta^{\mathsf{bm}}_{q,\delta}\right)^{\frac{r}{2}} k^{1-\frac{r}{2}}\left(1 + O\left(\frac{1}{\sqrt{k}}\right)\right), \tag{4.3}$$

and hence

$$\ln \mathfrak{t}^{(1)}_{k,r,q,\delta} = O(\ln k).$$

Moreover, since $L_{k,\delta} = \Theta(k)$,

$$\lim_{k\to\infty} \frac{1}{L_{k,\delta}} \ln \mathfrak{t}^{(1)}_{k,r,q,\delta} = 0. \tag{4.4}$$

Now, by the definition of $R_{k,r,q,\delta}$,

$$\ln R_{k,r,q,\delta} = \frac{\ln \mathrm{t}_{k,r,q,\delta}^{(L_{k,\delta})} - \ln \mathrm{t}_{k,r,q,\delta}^{(1)}}{L_{k,\delta} - 1}.$$

Since $\lim_{k\to\infty} \frac{L_{k,\delta}}{L_{k,\delta}-1} = 1$, Eq. (4.2) and Eq. (4.4) give

$$\lim_{k\to\infty} \ln R_{k,r,q,\delta} = \ln \mathscr{R}_{r,q,\delta}.$$

Exponentiating both sides completes the proof. □

Theorem 4.1 (Sufficient condition for vanishing failure probability)**.** *Let $r, q > 2$ be fixed. Let δ be a fixed constant satisfying*

$$\delta \leqslant \frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})}, \quad \Lambda_{q,\delta}^{\mathsf{bm}} < 1, \quad \textit{and} \quad \mathscr{R}_{r,q,\delta} < 1,$$

where $\mathscr{R}_{r,q,\delta}$ is defined in Eq. (4.1). Then,

$$\lim_{n\to\infty} \Pr[\textit{failure}] = 0.$$

More concretely, as $n \to \infty$,

$$\Pr[\textit{failure}] \leqslant \frac{(q-1)^{2-\frac{r}{2}}}{(q-2)^2} \frac{\left(\theta_{q,\delta}^{\mathsf{bm}}\right)^{\frac{r}{2}}}{1-\mathscr{R}_{r,q,\delta}} k^{1-\frac{r}{2}}(1+o(1)).$$

Remark 4.1 (Existence of an admissible constant δ). We show here that the δ-regime assumed by the above theorem is nonempty. As $\delta \to 0$

$$\Lambda_{q,\delta}^{\mathsf{bm}} = 2\sqrt{\frac{\delta}{q-1}} + O(\delta),$$

and hence

$$\theta_{q,\delta}^{\mathsf{bm}} = 2\sqrt{\frac{\delta}{q-1}} + O(\delta).$$

Moreover,

$$H(4\delta) = 4\delta \ln \frac{1}{4\delta} + 4\delta + O\left(\delta^2\right),$$

so that

$$\exp\left(\frac{H(4\delta)}{4\delta}\right) = \frac{e}{4\delta}(1+O(\delta)).$$

Also, as $x \to 0$,

$$\Phi_q(x) = \sqrt{\frac{x}{q-1}} + O(x).$$

Since

$$4\delta\theta^{\mathsf{bm}}_{q,\delta} = \frac{8}{\sqrt{q-1}}\delta^{\frac{3}{2}} + O\Big(\delta^2\Big),$$

we obtain

$$\Phi_q\Big(4\delta\theta^{\mathsf{bm}}_{q,\delta}\Big) = 2\sqrt{2}(q-1)^{-\frac{3}{4}}\delta^{\frac{3}{4}}\Big(1 + O\Big(\delta^{\frac{1}{4}}\Big)\Big).$$

Therefore,

$$\mathscr{R}_{r,q,\delta} = 2^{\frac{3r}{2}-2}e(q-1)^{1-\frac{3r}{4}}\delta^{\frac{3r}{4}-1}(1+o(1)).$$

Since $r > 2$, $\frac{3r}{4} - 1 > 0$, and consequently

$$\mathscr{R}_{r,q,\delta} \to 0 \quad \text{as} \quad \delta \to 0. \tag{4.5}$$

Furthermore,

$$\Lambda^{\mathsf{bm}}_{q,\delta} \to 0 \quad \text{as} \quad \delta \to 0. \tag{4.6}$$

It remains to verify that the "curvature" condition

$$\delta \leqslant \frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})}$$

is satisfied for every sufficiently small δ. Since $\theta^{\mathsf{bm}}_{q,\delta} = \Theta(\sqrt{\delta})$,

$$\mathfrak{v}^{\star}_{r,q,\delta} = O(\delta^{1/4}) \to 0 \quad \text{as} \quad \delta \to 0.$$

Moreover, as $x \to 0$,

$$\varphi(x) = \frac{x(3+2x^2)}{(1+x^2)^{3/2}} = 3x + O\Big(x^3\Big).$$

It follows that

$$\varphi\Big(\mathfrak{v}^{\star}_{r,q,\delta}\Big) = O\Big(\delta^{1/4}\Big),$$

and hence

$$\zeta_{r,q,\delta} = \frac{r-2}{2} + O\Big(\delta^{1/4}\Big).$$

Therefore,

$$\begin{aligned}\frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})} &= \frac{\frac{r-2}{2} + O\big(\delta^{1/4}\big)}{4\big(\frac{r}{2} + O\big(\delta^{1/4}\big)\big)} \\ &= \frac{r-2}{4r} + O\Big(\delta^{1/4}\Big).\end{aligned}$$

In particular,

$$\lim_{\delta\to 0}\frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})}=\frac{r-2}{4r}>0$$

because $r>2$. Finally, observe that

$$\zeta_{r,q,\delta}=\frac{r-2}{2}+\frac{r}{4}\varphi\left(\mathfrak{v}^{\star}_{r,q,\delta}\right)\geqslant\frac{r-2}{2},$$

since $\varphi(x)\geqslant 0$ for all $x\geqslant 0$. Moreover, since the function

$$x\mapsto\frac{x}{4(1+x)}$$

is increasing on $[0,\infty)$, it follows that

$$\frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})}\geqslant\frac{r-2}{4r}.$$

Consequently, every $0<\delta\leqslant\frac{r-2}{4r}$ satisfies the curvature condition. Thus, the curvature condition holds for all sufficiently small positive δ. Combining this result with Eq. (4.5) and Eq. (4.6), we conclude that, for every fixed $r,q>2$, there exists a constant

$$\delta_0=\delta_0(r,q)>0$$

such that every fixed $0<\delta<\delta_0$ satisfies the δ-regime assumed by the preceding theorem.

Proof of Theorem 4.1. By Lemma 4.1,

$$\lim_{k\to\infty}R_{k,r,q,\delta}=\mathscr{R}_{r,q,\delta}<1.$$

Hence, if we choose any constant ϱ satisfying $\mathscr{R}_{r,q,\delta}<\varrho<1$, then there exists $k_0\in\mathbb{N}$ such that $R_{k,r,q,\delta}\leqslant\varrho$ for every $k\geqslant k_0$. Consequently, for every $k\geqslant k_0$,

$$\mathscr{G}_{k,r,q,\delta}=\sum_{j=0}^{L_{k,\delta}-1}R^{j}_{k,r,q,\delta}\leqslant\sum_{j=0}^{\infty}\varrho^{j}=\frac{1}{1-\varrho}.$$

Therefore, from Theorem 1.2,

$$\Pr[\text{failure}]\leqslant\frac{q-1}{(q-2)^2}\frac{\mathfrak{t}^{(1)}_{k,r,q,\delta}}{1-\varrho}. \tag{4.7}$$

Further, from Eq. (4.3),

$$\mathfrak{t}^{(1)}_{k,r,q,\delta}=(q-1)^{1-\frac{r}{2}}\left(\theta^{\mathsf{bm}}_{q,\delta}\right)^{\frac{r}{2}}k^{1-\frac{r}{2}}\left(1+O\left(\frac{1}{\sqrt{k}}\right)\right).$$

Hence,

$$\Pr[\text{failure}]=O\left(k^{1-r/2}\right).$$

Since $r > 2$, $1 - r/2 < 0$, and thus

$$\lim_{n\to\infty} \Pr[\text{failure}] = 0.$$

Finally, because ϱ may be arbitrarily close to $\mathscr{R}_{r,q,\delta}$, Eq. (4.7) yields

$$\Pr[\text{failure}] \leqslant \frac{(q-1)^{2-\frac{r}{2}}}{(q-2)^2} \frac{\left(\theta^{\mathsf{bm}}_{q,\delta}\right)^{\frac{r}{2}}}{1-\mathscr{R}_{r,q,\delta}} k^{1-\frac{r}{2}}(1+o(1)).$$

□

Corollary 4.2 (Linear minimum distance of almost all RAA codes)**.** *For fixed $r, q > 2$, and a fixed constant $\delta \in \left(0, \frac{1}{4}\right)$ satisfying*

$$\delta \leqslant \frac{\zeta_{r,q,\delta}}{4(1+\zeta_{r,q,\delta})}, \quad \Lambda^{\mathsf{bm}}_{q,\delta} < 1, \quad \textit{and} \quad \mathscr{R}_{r,q,\delta} < 1,$$

almost all codes in the ensemble of length-n RAA codes satisfy

$$d^{\mathsf{RAA}}_{min} > \delta n.$$

Thus, the RAA ensemble has linear minimum-distance growth with high probability.

Proof. By definition,

$$\text{failure} \iff d^{\mathsf{RAA}}_{\min} \leqslant \delta n.$$

The claim therefore follows immediately from Theorem 4.1. □

All together, this section proves Corollary 1.3.

4.2.2 Chernoff Strategy.

In this section, we prove Corollary 1.4. Recall the quantities defined in the statement of Theorem 1.3.

Theorem 4.2 (Sufficient condition for vanishing failure probability using the Chernoff strategy)**.** *Let $r, q > 2$ be fixed. Let $\delta \in \left(0, \frac{1}{2}\right)$ be a fixed constant satisfying $Q_{r,q,\delta} < 1$. Then,*

$$\lim_{n\to\infty} \Pr[\textit{failure}] = 0.$$

More concretely, for every $k \geqslant 1$,

$$\Pr[\textit{failure}] \leqslant \frac{\phi_{r,q,\delta}}{(q-2)(1-Q_{r,q,\delta})} k^{1-\frac{r}{2}}.$$

In particular,

$$\Pr[\textit{failure}] = O_{r,q,\delta}\left(k^{1-\frac{r}{2}}\right).$$

Proof. By Theorem 1.3,

$$\Pr[\text{failure}] \leqslant \frac{k^{1-\frac{r}{2}}\phi_{r,q,\delta}\mathcal{S}_{k,r,q,\delta}}{q-2}.$$

Since $Q_{r,q,\delta} < 1$, we have for every $k \geqslant 1$,

$$\mathcal{S}_{k,r,q,\delta} := \sum_{\ell=0}^{U_{k,\delta}-1} Q_{r,q,\delta}^{\ell} \leqslant \sum_{\ell=0}^{\infty} Q_{r,q,\delta}^{\ell} = \frac{1}{1-Q_{r,q,\delta}}.$$

Consequently,

$$\Pr[\text{failure}] \leqslant \frac{\phi_{r,q,\delta}}{(q-2)(1-Q_{r,q,\delta})} k^{1-\frac{r}{2}}.$$

Since $r > 2$, $1 - \frac{r}{2} < 0$, and therefore

$$\lim_{n\to\infty} \Pr[\text{failure}] = 0.$$

□

Remark 4.2 (Existence of an admissible constant δ). We show here that the δ-regime assumed for the above theorem is nonempty. As $\delta \to 0$,

$$\Lambda_{\delta}^{\mathsf{ch}} := 2\sqrt{\delta(1-\delta)} = 2\sqrt{\delta} + O\left(\delta^{\frac{3}{2}}\right),$$

and hence

$$\theta_{\delta}^{\mathsf{ch}} := \frac{\Lambda_{\delta}^{\mathsf{ch}}}{1-\Lambda_{\delta}^{\mathsf{ch}}} = 2\sqrt{\delta} + O(\delta).$$

Moreover, since

$$\Phi_q(x) = \sqrt{\frac{x}{q-1}} + O_q(x) \quad \text{as} \quad x \to 0,$$

we obtain

$$\Phi_q\left(\theta_{\delta}^{\mathsf{ch}}\right) = \sqrt{\frac{2}{q-1}}\delta^{1/4}\left(1 + O_q\left(\delta^{1/4}\right)\right).$$

Therefore,

$$\phi_{r,q,\delta} := (q-1)\Phi_q\left(\theta_{\delta}^{\mathsf{ch}}\right)^r = 2^{r/2}(q-1)^{1-r/2}\delta^{r/4}\left(1 + O_{r,q}\left(\delta^{r/4}\right)\right).$$

Now, for sufficiently small δ, $m_{\delta} = 4\delta$. Hence,

$$\begin{aligned} Q_{r,q,\delta} &= e^{r/2}(4\delta)^{r/2-1}\phi_{r,q,\delta} \\ &= 2^{3r/2-2}e^{r/2}(q-1)^{1-r/2}\delta^{3r/4-1}\left(1 + O_{r,q}\left(\delta^{1/4}\right)\right). \end{aligned}$$

Since $r > 2$, $\frac{3r}{4} - 1 > 0$, and therefore

$$Q_{r,q,\delta} \to 0 \quad \text{as} \quad \delta \to 0.$$

Thus, for every fixed $r, q > 2$, there exists constant

$$\delta_0 = \delta_0(r, q) > 0$$

such that every fixed $0 < \delta < \delta_0$ satisfies $Q_{r,q,\delta} < 1$.

Corollary 4.3 (Linear minimum distance of almost all RAA codes)**.** *For fixed $r, q > 2$, and a fixed constant $\delta \in \left(0, \frac{1}{2}\right)$ satisfying $Q_{r,q,\delta} < 1$, almost all codes in the ensemble of length-n RAA codes satisfy*

$$d_{\min}^{\mathsf{RAA}} > \delta n.$$

Thus, the RAA ensemble has linear minimum-distance growth with high probability.

Proof. By definition,

$$\text{failure} \iff d_{\min}^{\mathsf{RAA}} \leqslant \delta n.$$

The claim therefore follows immediately from Theorem 4.2. □

All together, this section proves Corollary 1.4.

4.3 EA Code

In this section, we prove Corollary 1.5. Recall the parameters as given in Theorem 1.4.

Corollary 4.4 (Sufficient condition for vanishing failure probability for fixed p)**.** *Let prime power $q > 2$, $r \in \mathbb{Z}_{\geqslant 3}$, and $p \in (0, 1)$ be fixed. Define*

$$\delta^\star := \left(\frac{\sqrt{1 + (q-2)q^{1-1/r}} - \sqrt{q-1}}{q-2} \right)^2.$$

Then, for every $0 < \delta < \delta^\star$, the $[rk, k]_q$ EA code ensemble satisfies

$$\lim_{n\to\infty} \Pr[\textit{failure}] = 0.$$

In order to prove this, we first establish an auxiliary lemma along with a corollary.

Lemma 4.2. *Define*

$$\mathcal{U}_1 := q^k \left(1 - \mu_{q,\delta} + \mu_{q,\delta}\theta_{k,p,q}\right)^n,$$
$$\mathcal{L} := \lim_{n\to\infty} \theta_{k,p,q}.$$

Then,

$$\lim_{n\to\infty} \mathcal{U}_1 = 0$$

whenever

$$q\left(1 - \mu_{q,\delta} + \mu_{q,\delta}\mathcal{L}\right)^r < 1.$$

Proof. Define the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = q\big(1 - \mu_{q,\delta} + \mu_{q,\delta} x\big)^r.$$

Now we use the result that if a sequence x_k approaches x, and a function g is continuous at x, then $g(x_k)$ approaches $g(x)$. Since f defined above is a polynomial in x, it is continuous over $\mathbb{R}$. Moreover, since $\theta_{k,p,q} \to 0$, we have that

$$\lim_{k\to\infty} q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\theta_{k,p,q}\big)^r = \lim_{k\to\infty} f\big(\theta_{k,p,q}\big) = f(\mathcal{L}) = q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\mathcal{L}\big)^r < 1.$$

Hence, the sequence $a_k := q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\theta_{k,p,q}\big)^r$ converges to a number strictly less than 1. Therefore, for some constant $\epsilon \in \big(q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\mathcal{L}\big)^r, 1\big)$, we have

$$0 \leqslant a_k < \epsilon < 1 \iff 0 \leqslant a_k^k < \epsilon^k < 1$$

for all sufficiently large k. Thus, by the sandwich theorem,

$$\lim_{n\to\infty} \mathscr{U}_1 = \lim_{k\to\infty} a_k^k = 0$$

since

$$0 \leqslant \lim_{k\to\infty} a_k^k \leqslant \lim_{k\to\infty} \epsilon^k = 0. \qquad \square$$

Corollary 4.5. *If $p \in (0,1)$ is fixed, then*

$$\lim_{n\to\infty} \mathscr{U}_1 = 0$$

whenever

$$q\big(1 - \mu_{q,\delta}\big)^r < 1.$$

Proof. In order to prove this, it suffices to show that if $p \in (0,1)$ is fixed, then

$$\lim_{k\to\infty} \theta_{k,p,q} = 0.$$

Once this is shown, the result follows from the preceding lemma. Hence, the remaining proof is dedicated to proving the above limit.

By the definition of $\tau_{k,p,q}$,

$$\tau_{k,p,q} = \Theta(\ln k) \implies \lim_{k\to\infty} \frac{k}{\tau_{k,p,q}} = \infty,$$

and

$$\eta_{p,q}^{\tau_{k,p,q}} \leqslant \frac{1}{k(q-1)} \iff 1 \leqslant 1 + (q-1)\eta_{p,q}^{\tau_{k,p,q}} \leqslant 1 + \frac{1}{k}.$$

Therefore,

$$\lim_{k\to\infty} \theta_{k,p,q} = \lim_{k\to\infty} \left(\frac{1 + (q-1)\eta_{p,q}^{\tau_{k,p,q}}}{q} \right)^{k/\tau_{k,p,q}} = 0. \qquad \square$$

Proof of Corollary 4.4. Let

$$\begin{aligned}\mathcal{U}_1 &:= q^k\big(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_{k,p,q}\big)^n,\\ \mathcal{U}_2 &:= (1-\mu_{q,\delta})^n,\\ \mathcal{U}_3 &:= ek(q-1)\nu_{p,q,\delta}^n.\end{aligned}$$

Then, for $\lim_{n\to\infty}\Pr[\text{failure}]=0$, it suffices from Theorem 1.4 that

$$\lim_{n\to\infty}\mathcal{S}_k=0 \iff \lim_{n\to\infty}\mathcal{U}_1-\mathcal{U}_2+\mathcal{U}_3=0.$$

Furthermore, since

$$\lim_{n\to\infty}\mathcal{U}_3=\lim_{k\to\infty}ek(q-1)\nu_{p,q,\delta}^{rk}=0 \qquad (\because \nu_{p,q,\delta}<1),$$

and

$$\mathcal{U}_2=o(\mathcal{U}_1),$$

it suffices that

$$\lim_{n\to\infty}\mathcal{U}_1=\lim_{k\to\infty}\big[q(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_{k,p,q})^r\big]^k=0.$$

Now, observe that the condition $\delta<\delta^\star$ is precisely equivalent to $q(1-\mu_{q,\delta})^r<1$. Moreover, by Corollary 4.5, we have

$$q(1-\mu_{q,\delta})^r \implies \lim_{k\to\infty}\mathcal{U}_1=0.$$

Therefore, finally,

$$0<\delta<\delta^\star \iff q(1-\mu_{q,\delta})^r \implies \lim_{k\to\infty}\mathcal{U}_1=0 \implies \lim_{n\to\infty}\Pr[\text{failure}]=0.$$

□

We next allow p to depend on the blocklength: $p=p_n$.

Theorem 4.3 (Sufficient condition for vanishing failure probability for $p_n=o(1)$)**.** *Suppose*

$$p_n=o(1), \qquad p_n>0.$$

Define

$$\begin{aligned}\Lambda_{q,\delta} &:= 2\sqrt{\frac{\delta}{q-1}}+\frac{q-2}{q-1}\delta,\\ \eta_n &:= \left|1-\frac{p_nq}{q-1}\right|,\\ \tau_n &:= \left\lceil\frac{\ln(k(q-1))}{-\ln\eta_n}\right\rceil,\end{aligned}$$

$$\theta_n := \left(\frac{1+(q-1)\eta_n^{\tau_n}}{q}\right)^{k/\tau_n},$$
$$\nu_n := 1-(1-\Lambda_{q,\delta})p_n,$$
$$B_{r,q} := \frac{q^{1-\frac{1}{r}}-1}{q-1}.$$

Then, the sufficient condition for vanishing failure probability is given as follows:

1. *If* $p_n = \omega\left(\frac{\ln n}{n}\right)$, *then*

$$\Lambda_{q,\delta} < B_{r,q} \implies \lim_{n\to\infty} \Pr[\textit{failure}] = 0.$$

2. *If* $p_n = \gamma\frac{\ln n}{n}(1+o(1))$ *for a constant* $\gamma \in (0,\infty)$, *then*

$$\Lambda_{q,\delta} < \min\left\{\frac{B_{r,q}-\beta_{r,q}}{1-\beta_{r,q}}, 1-\frac{1}{\gamma}\right\} \implies \lim_{n\to\infty} \Pr[\textit{failure}] = 0,$$

   where $\beta_{r,q} := q^{-\frac{\gamma q}{r(q-1)}}$.
3. *If* $p_n = o\left(\frac{\ln n}{n}\right)$, *then no sufficient condition exists; that is, our analysis cannot establish a vanishing failure probability.*

In order to prove this, we first establish two auxiliary lemmas.

Lemma 4.3 (Asymptotics for $p_n = o(1)$)**.** *In addition to the quantities defined in the above theorem statement, define*

$$\lambda_n := \frac{np_n}{\ln n}.$$

Then,

$$\frac{\tau_n}{n} = \frac{q-1}{q}\frac{1}{\lambda_n}(1+o(1)). \tag{4.8}$$

Moreover,

$$\lambda_n \to \infty \implies \frac{\tau_n}{n} \to 0, \quad \theta_n \to 0,$$
$$\lambda_n \to \gamma \in (0,\infty) \implies \frac{\tau_n}{n} \to \frac{q-1}{\gamma q}, \quad \theta_n \to q^{-\frac{\gamma q}{r(q-1)}},$$
$$\lambda_n \to 0 \implies \frac{\tau_n}{n} \to \infty, \quad \theta_n \to 1.$$

Proof. Since $p_n \to 0$, for all sufficiently large n, $p_n < \frac{q-1}{q}$, and hence

$$\eta_n = 1-\frac{q}{q-1}p_n.$$

Using the fact that

$$-\ln(1-x) = x + O\left(x^2\right) \qquad (x \to 0),$$

we obtain

$$-\ln \eta_n = \frac{q}{q-1} p_n + O\left(p_n^2\right) = \frac{q}{q-1} p_n (1 + o(1)).$$

Moreover, since $k = \frac{n}{r}$,

$$\ln(k(q-1)) = \ln n + O(1) = \ln n (1 + o(1)).$$

Therefore,

$$\begin{aligned} \tau_n &= \frac{\ln(k(q-1))}{-\ln \eta_n} (1 + o(1)) \\ &= \frac{q-1}{q} \frac{\ln n}{p_n} (1 + o(1)). \end{aligned}$$

Dividing by n gives

$$\frac{\tau_n}{n} = \frac{q-1}{q} \frac{\ln n}{n p_n} (1 + o(1)) = \frac{q-1}{q} \frac{1}{\lambda_n} (1 + o(1)),$$

which proves Eq. (4.8). It remains to determine θ_n. By the definition of τ_n,

$$\eta_n^{\tau_n} \leqslant \frac{1}{k(q-1)}.$$

Consequently,

$$1 \leqslant 1 + (q-1)\eta_n^{\tau_n} \leqslant 1 + \frac{1}{k},$$

and hence

$$\ln\left(1 + (q-1)\eta_n^{\tau_n}\right) = o(1).$$

Thus,

$$\begin{aligned} \ln \theta_n &= \frac{k}{\tau_n} \left[\ln\left(1 + (q-1)\eta_n^{\tau_n}\right) - \ln q\right] \\ &= -\frac{k}{\tau_n} \ln q + o\left(\frac{k}{\tau_n}\right) \\ &= -\frac{1}{r} \frac{n}{\tau_n} \ln q + o\left(\frac{1}{r} \frac{n}{\tau_n}\right) \\ &= -\frac{q \ln q}{r(q-1)} \lambda_n (1 + o(1)). \end{aligned}$$

So, if $\lambda_n \to \infty$, the above equation implies

$$\ln \theta_n \to -\infty \iff \theta_n \to 0.$$

If $\lambda_n \to \gamma \in (0, \infty)$, then

$$\ln \theta_n \to -\frac{\gamma q \ln q}{r(q-1)} \iff \theta_n \to q^{-\frac{\gamma q}{r(q-1)}}.$$

Finally, if $\lambda_n \to 0$, then

$$\ln \theta_n \to 0 \iff \theta_n \to 1. \qquad \square$$

Lemma 4.4. *In addition to the quantities defined in the statement of Theorem 4.3, define*

$$\mathscr{V}_n := k\nu_n^n, \quad \lambda_n := \frac{np_n}{\ln n}.$$

Then:

1. *If* $\lambda_n \to \infty$,

$$\lim_{n\to\infty} \mathscr{V}_n = 0.$$

2. *If* $\lambda_n \to \gamma \in (0,\infty)$ *and* $\gamma(1-\Lambda_{q,\delta}) > 1$,

$$\lim_{n\to\infty} \mathscr{V}_n = 0,$$

3. *If* $\lambda_n \to 0$,

$$\mathscr{V}_n \to \infty.$$

Proof. Since $p_n = o(1)$,

$$\ln \nu_n = -(1-\Lambda_{q,\delta})p_n + O\left(p_n^2\right),$$

and hence

$$\ln \mathscr{V}_n = \ln n - (1-\Lambda_{q,\delta})np_n(1+o(1)) + O(1).$$

If $\lambda_n \to \infty$, then $np_n = \omega(\ln n)$, so

$$\ln \mathscr{V}_n \to -\infty \iff \lim_{n\to\infty} \mathscr{V}_n = 0.$$

If $\lambda_n \to \gamma \in (0,\infty)$, then $np_n = \gamma \ln n(1+o(1))$, so

$$\ln \mathscr{V}_n = [1-\gamma(1-\Lambda_{q,\delta}) + o(1)] \ln n.$$

Thus, $\lim_{n\to\infty} \mathscr{V}_n = 0$ whenever $\gamma(1-\Lambda_{q,\delta}) > 1$.

Finally, if $\lambda_n \to 0$, then $np_n = o(\ln n)$, giving

$$\ln \mathscr{V}_n = (1-o(1)) \ln n \to \infty \iff \mathscr{V}_n \to \infty.$$ □

Proof of Theorem 4.3. Let

$$\begin{aligned}
\lambda_n &:= \frac{np_n}{\ln n}, \\
\mu_{q,\delta} &:= \frac{q-1}{q}(1-\Lambda_{q,\delta}), \\
\mathcal{U}_1 &:= q^k\left(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_n\right)^n, \\
\mathcal{U}_2 &:= e(q-1)k\nu_n^n, \\
\mathcal{U}_3 &:= (1-\mu_{q,\delta})^n.
\end{aligned}$$

Then, from Theorem 1.4,

$$\Pr[\text{failure}] \leqslant \frac{\mathcal{U}_1 + \mathcal{U}_2 - \mathcal{U}_3}{q-2}.$$

Hence, for $\lim_{n\to\infty} \Pr[\text{failure}] = 0$, it suffices that

$$\lim_{n\to\infty} \mathcal{U}_1 + \mathcal{U}_2 - \mathcal{U}_3 = 0.$$

Furthermore, since $\mathcal{U}_3 = o(\mathcal{U}_1)$, it suffices that

$$\lim_{n\to\infty} \mathcal{U}_1 + \mathcal{U}_2 = \lim_{n\to\infty} q^k(1 - \mu_{q,\delta} + \mu_{q,\delta}\theta_n)^n + e(q-1)\lim_{n\to\infty} k\nu_n^n = 0.$$

Let

$$\mathcal{L}_1 := \lim_{n\to\infty} \mathcal{U}_1 = \lim_{n\to\infty} q^k(1 - \mu_{q,\delta} + \mu_{q,\delta}\theta_n)^n \quad \text{and} \quad \mathcal{L}_2 := \lim_{n\to\infty} \mathcal{U}_2 = \lim_{n\to\infty} k\nu_n^n.$$

If $p_n = \omega\left(\frac{\ln n}{n}\right)$, then $\lambda_n \to \infty$. In this regime, by Lemma 4.4, $\mathcal{L}_2 = 0$. Hence, the failure probability vanishes if $\mathcal{L}_1 = 0$. By Lemma 4.3 and Lemma 4.2, a sufficient condition for $\mathcal{L}_1 = 0$ is

$$q(1 - \mu_{q,\delta})^r < 1.$$

Unpacking the definition of $\mu_{q,\delta}$, this is equivalent to

$$\Lambda_{q,\delta} < B_{r,q}.$$

Thus,

$$\Lambda_{q,\delta} < B_{r,q} \implies \mathcal{L}_1 = 0 \implies \lim_{n\to\infty} \Pr[\text{failure}] = 0.$$

Before analyzing the next regime, we note that $\mathcal{U}_1, \mathcal{U}_2 \geqslant 0$. Hence, $\mathcal{L}_1, \mathcal{L}_2 \geqslant 0$. This implies that the sufficient condition $\mathcal{L}_1 + \mathcal{L}_2 = 0$ is necessarily equivalent to the condition

$$\mathcal{L}_1 = \mathcal{L}_2 = 0.$$

Now, if $p_n = \gamma\frac{\ln}{n}(1 + o(1))$ for some constant $\gamma \in (0, \infty)$, then $\lim_{n\to\infty} \lambda_n = \gamma$. In this regime, by Lemma 4.3, $\mathcal{L}_2 = 0$ whenever

$$\gamma(1 - \Lambda_{q,\delta}) > 1 \iff \Lambda_{q,\delta} < 1 - \frac{1}{\gamma}.$$

Moreover, by Lemma 4.3 and Lemma 4.2, a sufficient condition for $\mathcal{L}_1 = 0$ is

$$q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big)^r < 1 \iff \Lambda_{q,\delta} < \frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}},$$

where $\beta_{r,q} := q^{-\frac{\gamma q}{r(q-1)}}$. Thus,

$$\Lambda_{q,\delta} < \min\left\{\frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}}, 1 - \frac{1}{\gamma}\right\} \implies \mathcal{L}_1 = \mathcal{L}_2 = 0 \iff \mathcal{L}_1 + \mathcal{L}_2 = 0 \implies \lim_{n\to\infty} \Pr[\text{failure}] = 0.$$

Finally, if $p_n = o(\frac{\ln n}{n})$, then $\lim_{n\to\infty} \lambda_n = 0$. In this regime, by Lemma 4.4, $\mathcal{L}_2 = \infty$. Thus, our analysis does not yield a sufficient condition for vanishing failure probability. □

Remark 4.3 ($p_n = \frac{\mathfrak{c}}{n}$). The case of $p_n = \frac{\mathfrak{c}}{n}$ for a constant $\mathfrak{c}$ falls in the third regime: $p_n = o\left(\frac{\ln n}{n}\right)$. Hence, our analysis is inconclusive.

Remark 4.4 (Comparison of critical-regime threshold with fixed-p threshold). From Corollary 4.4, the fixed-p sufficient condition for vanishing failure probability can be written as

$$\Lambda_{q,\delta} < B_{r,q}, \qquad B_{r,q} := \frac{q^{1-\frac{1}{r}} - 1}{q-1}.$$

In the regime

$$p_n = \gamma \frac{\ln n}{n}, \qquad \gamma \in (0, \infty),$$

the preceding theorem gives the sufficient condition

$$\Lambda_{q,\delta} < \min\left\{ \frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}}, 1 - \frac{1}{\gamma} \right\}, \qquad \beta_{r,q} := q^{-\frac{\gamma q}{r(q-1)}}.$$

For every finite $\gamma > 0$, we have $0 < \beta_{r,q} < 1$. Moreover, since $B_{r,q} < 1$,

$$\frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}} < B_{r,q}.$$

Hence,

$$\min\left\{ \frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}}, 1 - \frac{1}{\gamma} \right\} < B_{r,q}.$$

So, the critical regime never yields a larger admissible $\Lambda_{q,\delta}$, and therefore never yields a larger admissible δ since $\Lambda_{q,\delta}$ is strictly increasing in δ.

Furthermore, there does not exist even a single admissible δ for some values of γ. A positive critical-regime δ requires both terms in the minimum to be positive. Thus,

$$\gamma > \max\left\{ 1, \frac{r(q-1)}{q \ln q} \ln \frac{1}{B_{r,q}} \right\}$$

is necessary for the critical-regime analysis to certify a positive δ.

Finally, as $\gamma \to \infty$,

$$\beta_{r,q} \to 0, \qquad 1 - \frac{1}{\gamma} \to 1.$$

Therefore,

$$\frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}} \to B_{r,q}.$$

Hence, the critical-regime threshold approaches the fixed-p threshold from below.

Conclusion: Increasing γ continuously recovers the fixed-p threshold, but no finite critical-scale choice $p_n = \gamma \frac{\ln n}{n}$ improves upon it. This signifies a trade-off between provable sufficient threshold and encoder sparsity.

Corollary 4.6 (Decay rate of the EA failure probability)**.** *Let $r, q > 2$ be fixed, let $n = rk \to \infty$, and let $\Lambda_{q,\delta} < 1$. Suppose that in the following regimes, the respective sufficient conditions for vanishing failure probability (from Corollary 4.4 and Theorem 4.3) are satisfied. Then:*

1. *If $p_n = p \in (0,1)$ is fixed, then*
$$\Pr[\textit{failure}] = e^{-\Theta_{r,q,\delta}(n)}.$$
2. *If $p_n = \omega\left(\frac{\ln n}{n}\right)$, then for every fixed $\xi > 0$,*
$$\Pr[\textit{failure}] = O_{r,q,\delta}\left(n^{-\xi}\right).$$
3. *If $p_n = \gamma\frac{\ln n}{n}(1 + o(1))$ for a constant γ, then*
$$\Pr[\textit{failure}] \leqslant n^{-\left[\gamma(1-\Lambda_{q,\delta})-1\right]+o(1)}.$$

Proof. Let
$$\begin{aligned}
\lambda_n &:= \frac{np_n}{\ln n},\\
\mu_{q,\delta} &:= \frac{q-1}{q}(1-\Lambda_{q,\delta}),\\
\mathcal{U}_n &:= q^k\left(1-\mu_{q,\delta}+\mu_{q,\delta}\theta_n\right)^n,\\
\mathcal{T}_n &:= k\nu_n^n.
\end{aligned}$$

From Theorem 1.4,
$$\Pr[\text{failure}] \leqslant \frac{\mathcal{U}_n + e(q-1)\mathcal{T}_n - (1-\mu_{q,\delta})^n}{q-2} \leqslant \frac{\mathcal{U}_n + e(q-1)\mathcal{T}_n}{q-2}.$$

Anyway, $\lim_{n\to\infty}(1-\mu_{q,\delta})^n = 0$. Hence, the scaling contention is between the other two terms. We analyze these two terms in each regime.

If $p_n = p$ is fixed, then, by Corollary 4.5, $\lim_{n\to\infty}\theta_n = 0$. Hence, by Lemma 4.2,
$$\lim_{n\to\infty}\mathcal{U}_n^{1/n} = q^{\frac{1}{r}}(1-\mu_{q,\delta}).$$

Equivalently,
$$\frac{\ln\mathcal{U}_n}{n} = \frac{\ln q}{r} + \ln(1-\mu_{q,\delta}) + o(1).$$

This implies
$$\ln\mathcal{U}_n = \left(\frac{\ln q}{r} + \ln(1-\mu_{q,\delta})\right)n + o(n),$$

or equivalently,
$$\mathcal{U}_n = e^{\left(\frac{\ln q}{r}+\ln(1-\mu_{q,\delta})\right)n+o(n)}.$$

Also, the sufficient condition for vanishing failure probability in this regime yields

$$\Lambda_{q,\delta} < B_{r,q} \iff q(1-\mu_{q,\delta})^r < 1 \iff q^{\frac{1}{r}}(1-\mu_{q,\delta}) < 1 \iff \frac{\ln q}{r} + \ln(1-\mu_{q,\delta}) < 0.$$

Hence, $\frac{\ln q}{r} + \ln(1-\mu_{q,\delta})$ is a negative constant. Thus,

$$\mathcal{U}_n = e^{-\Theta_{r,q,\delta}(n)}. \tag{4.9}$$

Furthermore, for fixed $p_n = p$, $\eta_{p,q} < 1$ is a constant which implies that ν_n is constant. Moreover, $\nu_n < 1$ since $0 < \mu_{q,\delta}, \eta_{p,q} < 1 \implies 0 < 1 - \mu_{q,\delta}(1-\eta_{p,q}) < 1$. Hence,

$$\mathcal{T}_n = k\nu_n^n = e^{n\ln\nu_n + \ln k} = e^{n\ln\nu_n + O_r(\ln n)}.$$

But $\ln \nu_n < 0$ since $\nu_n < 1$. Thus,

$$\mathcal{T}_n = e^{-\Theta_{r,q,\delta}(n)}.$$

This combined with Eq. (4.9) proves that the failure probability decays exponentially in the fixed-p regime.

A useful representation for ν_n in the $p_n = o(1)$ regime is

$$\nu_n = 1 - (1-\Lambda_{q,\delta})p_n. \tag{4.10}$$

If $p_n = \omega\left(\frac{\ln n}{n}\right)$, then $\lambda_n \to \infty$. Hence, by Lemma 4.4, $\theta_n \to 0$. Moreover, we have $\Lambda_{q,\delta} < B_{r,q}$. This is similar to the fixed-p case. So, we get that

$$\mathcal{U}_n = e^{-\Theta_{r,q,\delta}(n)}.$$

For $\mathcal{T}_n$, from Eq. (4.10),

$$\mathcal{T}_n = k\big(1-(1-\Lambda_{q,\delta})p_n\big)^n.$$

Taking logarithms, and using the result that $\ln(1-x) = -x + O\big(x^2\big)$ as $x \to 0$,

$$\begin{aligned}\ln \mathcal{T}_n &= \ln k + n\ln(1-(1-\Lambda_{q,\delta})p_n)\\ &= \ln k - (1-\Lambda_{q,\delta})np_n + O\Big(np_n^2\Big).\end{aligned}$$

Since $p_n = o(1)$, $np_n^2 = o(np_n)$. Hence,

$$\begin{aligned}\ln \mathcal{T}_n &= \ln n - (1-\Lambda_{q,\delta})np_n(1+o(1))\\ &= \ln n\big(1-(1-\Lambda_{q,\delta})\lambda_n(1+o(1))\big).\end{aligned}$$

Since $\lambda_n \to \infty$,

$$1-(1-\Lambda_{q,\delta})\lambda_n(1+o(1)) \to -\infty.$$

Hence, for every fixed $\xi > 0$ and sufficiently large n,

$$1-(1-\Lambda_{q,\delta})\lambda_n(1+o(1)) \leqslant -\xi.$$

Therefore, for all sufficiently large n,

$$\ln \mathcal{T}_n \leqslant -\xi \ln n.$$

Thus,

$$\mathcal{T}_n = O\Big(n^{-\xi}\Big).$$

Moreover, $\mathcal{U}_n$ is exponentially small. Therefore, for *every* $\xi > 0$,

$$\Pr[\text{failure}] = O\Big(n^{-\xi}\Big),$$

i.e., the failure probability decays faster than every inverse polynomial in this regime.

If $p_n = \gamma \frac{\ln n}{n}(1 + o(1))$ for a constant γ, then $\lim_{n\to\infty} \lambda_n = \gamma$. Hence, by Lemma 4.4,

$$\lim_{n\to\infty} \theta_n = \beta_{r,q} := q^{-\frac{\gamma q}{r(q-1)}}$$

Further, by Lemma 4.2,

$$\lim_{n\to\infty} \mathcal{U}_n^{1/n} = q^{\frac{1}{r}}\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big).$$

Equivalently,

$$\frac{\ln \mathcal{U}_n}{n} = \frac{\ln q}{r} + \ln\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big) + o(1).$$

This implies

$$\ln \mathcal{U}_n = \left(\frac{\ln q}{r} + \ln\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big)\right)n + o(n),$$

or equivalently

$$\mathcal{U}_n = e^{\left(\frac{\ln q}{r} + \ln\left(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\right)\right)n + o(n)}.$$

Also, the sufficient condition for vanishing failure probability in this regime yields

$$\begin{aligned}
\Lambda_{q,\delta} < \frac{B_{r,q} - \beta_{r,q}}{1 - \beta_{r,q}} &\iff q\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big)^r < 1 \\
&\iff q^{\frac{1}{r}}\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big) < 1 \\
&\iff \frac{\ln q}{r} + \ln\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big) < 0.
\end{aligned}$$

Hence, $\frac{\ln q}{r} + \ln\big(1 - \mu_{q,\delta} + \mu_{q,\delta}\beta_{r,q}\big)$ is a negative constant. Thus,

$$\mathcal{U}_n = e^{-\Theta_{r,q,\delta}(n)}.$$

For $\mathcal{T}_n$, from Eq. (4.10),

$$\mathcal{T}_n = k\big(1 - (1 - \Lambda_{q,\delta})p_n\big)^n.$$

Taking logarithms, and using the result that $\ln(1-x) = -x + O(x^2)$ as $x \to 0$,

$$\begin{aligned}\ln \mathcal{T}_n &= \ln k + n \ln(1 - (1 - \Lambda_{q,\delta}) p_n) \\ &= \ln k - (1 - \Lambda_{q,\delta}) n p_n + O\left(n p_n^2\right).\end{aligned}$$

Since $p_n = \Theta\left(\frac{\ln n}{n}\right) = o(1)$, $O\left(np_n^2\right) = o(np_n) = o(\ln n)$. Moreover, since $\lambda_n \to \gamma$, $np_n \to \gamma \ln n(1 + o(1))$. Hence,

$$\begin{aligned}\ln \mathcal{T}_n &= \ln k - (1 - \Lambda_{q,\delta}) n p_n + o(\ln n) \\ &= \ln n - \gamma(1 - \Lambda_{q,\delta}) \ln n + o(\ln n) \\ &= -[\gamma(1 - \Lambda_{q,\delta}) - 1 + o(1)] \ln n.\end{aligned}$$

Defining

$$\varsigma := \gamma(1 - \Lambda_{q,\delta}) - 1,$$

we obtain

$$\mathcal{T}_n = n^{-\varsigma + o(1)}.$$

Note that the constant $\gamma(1 - \Lambda_{q,\delta}) - 1$ is always positive. Indeed, the sufficient condition for vanishing failure probability in this regime yields

$$\Lambda_{q,\delta} < 1 - \frac{1}{\gamma} \iff \gamma(1 - \Lambda_{q,\delta}) - 1 > 0.$$

Thus, the decay of $\mathcal{T}_n$ is inverse polynomial up to subpolynomial factors. Moreover, $\mathcal{U}_n$ is exponentially small. Therefore, the failure probability decays inverse polynomially up to subpolynomial factors:

$$\Pr[\text{failure}] \leqslant n^{-\varsigma + o(1)}. \qquad \square$$

All together, this section proves Corollary 1.5.

5 More Parameter Comparisons and Generation

In this section, we give another sampling of parameters generated by our scripts using Theorems 1.1 to 1.4. We plot the field size q versus calculated distance δ, for target failure probability $\tau \leqslant 2^{-60}$. Our findings are presented in Figure 2.

5.1 Concrete Parameter Comparisons

Here, we compare our concrete distance calculations with several other codes in the literature. The codes we compare to have provable distance analyses over any (suitably large) finite field. In particular, we do not compare with the original Expand-Accumulate code [BCG+22], the Expand-Convolute code [RRT23] or the Block-Accumulate code [KPR+26] since these codes only prove their minimum distance guarantees with respect to $\mathbb{F}_2$.

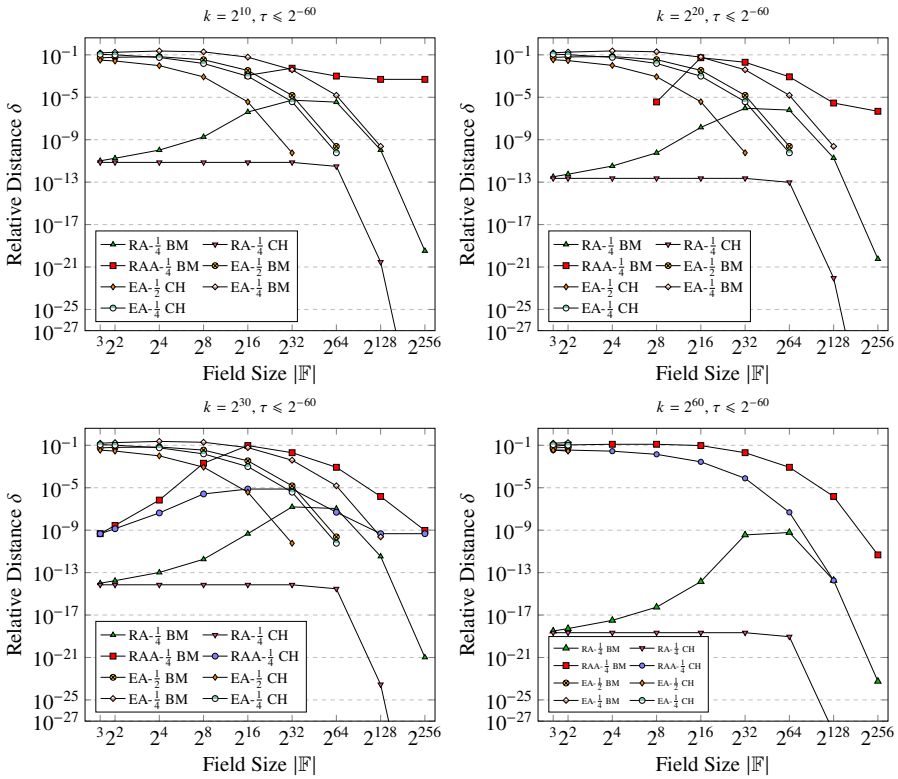

Figure 2: Another sampling of generated parameters for target failure probability $\leqslant 2^{-60}$. Omitted data points indicate that our binary search did not find an admissible bound within 50 iterations for the targeted parameters. For the EA code, the parameters for $p = 20\log(n)/n$ were selected.

Brakedown. To compare with Golovnev et al. [GLS+23], we use Haböck's improved analysis [Hab23]. We first remark that Haböck shows that for the Brakedown code, $\delta \leqslant \frac{1}{2}(1 - \frac{1}{\sqrt{2}}) \approx 0.146$ is the maximum possible attainable minimum distance, regardless of the rate ρ, as $q \to \infty$, and that some of our generated parameters are close to or exceed this bound; for example, $\delta = 1/8$ for the rate $\frac{1}{4}$ RAA code at $q = 2^8$ and $k = 2^{50}$, and $\delta = 0.2329553292$ for the rate $\frac{1}{4}$ EA code with Bernoulli probability $p = 20\log(n)/n$ and $q = 2^4$.

From [Hab23], parameters for $q = 2^{32}$ and $q = 2^{128}$ are given for the Brakedown code. For $q = 2^{128}$, the largest δ obtained is $\delta = 0.07$ for $\rho = 0.581$, and is valid for $\log(k) \in [15, 30]$; for the speed configuration, $\delta = 0.02$ and $\rho = 0.704$ are achieved. For our parameters under the same finite field size and $\rho = 1/4$, our RAA code reaches a maximum of $\delta = 0.0000305$ with failure probability 2^{-42} and $k = 2^{15}$, and becomes worse with increasing k; all of our other codes under our bounds are strictly worse.

For $q = 2^{32}$, [Hab23] shows that the same $\delta = 0.07$ and $\rho = 0.581$ hold for the Brakedown code for $\log(k) \in [13, 25]$; for the speed configuration, $\delta = 0.02$ and $\rho = 0.704$ are also achieved. For $k = 2^{15}$, our bound for the rate $\frac{1}{4}$ RAA code achieves $\delta = 0.02000526392$ except with probability 2^{-40}. We leave it as future work to compare the performance of Brakedown's speed configuration versus the RAA code in this case, though we believe Brakedown will still be faster due to the much higher rate.

Sketched Codes. Sketched codes leverage a base code $\mathbf{G} \in \mathbb{F}^{\alpha k \times n}$ along with a compression map $\mathbf{C} \in \mathbb{F}^{k \times \alpha k}$ for $\alpha < 1$, and encode a message $\mathbf{x}$ as $(\mathbf{x}, \mathbf{xCG})$. Both Lightning [QGZ26] and Bolt [GNR26] give constructions of sketched codes by setting $\mathbf{C}$ to be a sparse row-regular LDPC code [Gal62] and selecting various base codes $\mathbf{G}$ which have large minimum distance. Both results asymptotically yield $\delta = \Theta(1)$ and super-efficient linear-time encoding.

In Lightning, various rate-distance tradeoffs are presented. The parameters computed (via a union bound exhaustive computation) set $k = 2^{20}$ and target failure probability 2^{-100}; for their approach, the final bound they obtain is independent of the field size. For rate $\rho = 0.881$, they achieve $\delta = 0.005$, while for $\rho = 0.560$, they achieve $\delta = 0.03$. In Bolt, they target $q = 2^{32}$ and obtain $\delta = 0.01$ and $\delta = 0.07$ under different configurations, with all but $1/\text{poly}(n)$ probability.

To give some idea of how we compare, for $k = 2^{20}$, $r = 4$, and $q = 2^{32}$, our analysis yields $\delta = 0.0007797879$ except with 2^{-100} failure probability; this increases to $\delta = 0.01948211$ for 2^{-80} failure probability, and $\delta = 0.200164$ for 2^{-60} failure probability. Increasing the field size makes our achievable distance worse, and both the EA and RA codes also perform worse under our bounds.

Foldable Codes. Foldable codes, introduced by Zeilberger, Chen, and Fisch [ZCF24], are linear codes that are "folding friendly", meaning that they maintain distance and proximity guarantees when using them in the FRI IOPP [BBH+18]. For example, Reed-Solomon codes are foldable. Zeilberger, Chen, and Fisch introduce random foldable codes: given a base code $\mathbf{G}_0 \in \mathbb{F}_q^{k_0 \times n_0}$, the final generator matrix $\mathbf{G}_\ell \in \mathbb{F}_q^{k_\ell \times n_\ell}$ is recursively defined as

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{G}_{i-1} & \mathbf{G}_{i-1} \\ \mathbf{G}_{i-1} \cdot \mathbf{T}_{i-1} & \mathbf{G}_{i-1} \cdot -\mathbf{T}_{i-1} \end{bmatrix},$$

where $\mathbf{T}_{i-1}$ is a random full-rank diagonal matrix.

Notably, the analysis of Zeilberger, Chen, and Fisch shows that the base code $\mathbf{G}_0$ must have large minimum distance (in fact, [ZCF24] set $\mathbf{G}_0$ to be a (near) maximum distance separable code with large minimum distance), since the recursive definition of the generator matrix degrades the distance of $\mathbf{G}_0$ at every iteration. Moreover, their analysis assumes $q \geqslant 2^{10}$, which is a mild restriction.

For $q = 2^{31}$, rate $1/16$, and base code with distance $15/16$, BaseFold achieves $\delta = 0.5044$ at $k = 2^{20}$ except with probability 2^{-128}; another configuration for $k = 2^{20}$ and rate $1/16$ over $q = 2^{61}$ achieves $\delta = 0.484$. For rate $1/8$ and $k = 2^{20}$, $\delta = 0.557$ and $\delta = 0.728$ is achieved for $q = 2^{128}$ and $q = 2^{256}$, respectively.

Looking to our codes, at $k = 2^{20}$, rate $1/16$, and $q = 2^{32}$, the EA code with $p = 20\log(n)/n$ performs the best with $\delta = 0.2499840$ except with probability 2^{-100}. At $q = 2^{64}$, the RAA code performs the best with $\delta = 0.1122689$ except with probability 2^{-100}. For $q = 2^{128}$ and $q = 2^{256}$ at $k = 2^{20}$ and rate $1/8$, the RAA code performs the best, achieving (respectively) $\delta = 0.001321847$ and $\delta = 5.124819e-6$ except with probability 2^{-100}.

Encode-Repeat-Accumulate Codes. ERA codes, introduced by Baweja et al. [BFM+26], borrow from the idea of concatenated codes. Given some outer code $C_{\text{out}} \colon \mathbb{F}^k \to \mathbb{F}^{n/r}$, a message is first encoded with C_{out}, then encoded using the generalized RAA code. In [BFM+26, Table 1], a suitable base code of rate $\rho = 10/11$ is chosen, with varying input minimum distance (but always constant). For $k = 2^{20}$, $q = 2^{256}$, and input distance $\delta_{\text{out}} = 0.001$, when using the rate $1/4$ RAA code, they obtain $\delta = 0.2252$; with $\delta_{\text{out}} = 0.03$, they obtain $\delta = 0.2311$ with the same RAA code. Increasing the repetition factor of the RAA code greatly increases the minimum distance, which is expected due to what is known about RMA codes [KZC07]. All of these generated parameters exceed those generated by our bounds.

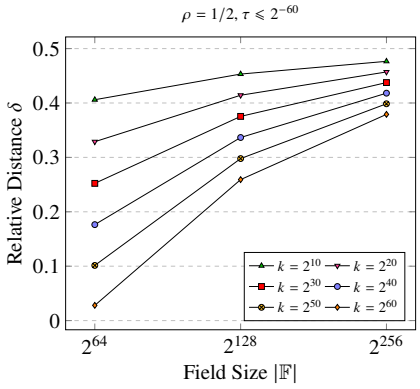

Figure 3: A sampling of generated parameters for the index 1, rank 2 (i.e., rate $1/2$) QA code for target failure probability $\leqslant 2^{-60}$. Omitted data points indicate our binary search did not find an admissible bound within 50 iterations for the targeted parameters.

Quasi-Abelian Codes. The Quasi-Abelian (QA) code, introduced by Bombar et al. [BCC+23], recently enjoyed a strong concrete distance analysis due to Li et al. [LLX+26]. Though not known at the time of writing if the code is asymptotically good, it exhibits excellent concrete performance. For message length k, rate ρ, and prime p, for target minimum distance δ, $c = 1/\rho$, and $\epsilon = 1 - H_p(\delta) - \frac{1+\log_p(k)}{c}$, [LLX+26] prove that the rank-1, index c QA code has minimum distance δ except with probability

$$\min\left\{\frac{c(c-1)N}{2p^2} + \frac{p^{-\lceil c-1/c\delta\rceil c\epsilon}}{(1-p^{-c\epsilon})(p-1)}, \frac{cN}{p} + \frac{p^{-\lceil 1/\delta\rceil c\epsilon}}{(1-p^{-c\epsilon})(p-1)}\right\}$$

This strong failure probability allows them to achieve very large concrete δ. For $k = 2^{20}$ and 128-bit fields, except with 2^{-127} probability, the code has distance $\delta = 0.4127$, and for the BN25519 field, the error probability is 2^{-250}.

We also implemented a binary search script to calculate concrete parameters for the QA code. As of this writing, this code has some of the strongest concrete minimum distance guarantees, and is very fast to encode: using $O(n \log n)$ additions and $O(n)$ multiplications, the code is essentially linear time (since field additions are much cheaper than multiplications).

In Figure 3, we give a sampling of QA code parameters we calculated using our scripts.

6 Conclusions and Open Problems

In this work, for any fixed set of parameters $(k, \rho, \mathbb{F}_q, \delta)$ such that $q \geqslant 3$, we give explicit concrete upper bounds on the failure probability that the $[k/\rho, k]_q$ RA, RAA, and EA code ensembles do not have distance at least δ. The RA code concretely matches its asymptotic performance: the distance is poor and degrades with increasing k. For the RAA and EA codes (under suitable parameters), we were able to obtain a maximum of $\delta = 1/8$ and $\delta = 0.2329553292$ at a failure probability of 2^{-60}. Along the way, we propose a fine-grained union bound for accumulate-type codes, which we believe to be of independent interest, that can be extended to other accumulate codes and more general turbo-like codes.

Our work leaves open many future directions. One direction is extending to other turbo-like codes. In Section 3.4, we briefly highlighted the modularity of our analysis and gave an overview of how to apply our

generalized union bound to the ERA code and the Expand-Convolute code. Many other codes exist within the framework of turbo-like codes, such as the fixed-weight EA code [Kha26a, LXL+26], block-accumulate codes [KPR+26], or the unscaled RAA code [Kha26b], and applying our analysis to these codes is a natural first-step. Another exciting direction is improving our bounds; as demonstrated, we obtain concretely large minimum distance when the field size is small. Though there are emerging applications which can utilize small fields [HGQ+26], improving our bounds for larger finite fields makes our work more applicable to zkSNARKs and PCGs. A natural direction to improve our bounds and prior work is to combine our approach with the union bound stratification other works use; we believe this is the most promising future direction to get concrete improvements to the minimum distance calculations. Finally, as an orthogonal direction, finding more cryptographic applications for codes over small fields is another exciting direction, as we have demonstrated that our codes have strong minimum distance guarantees over smaller fields.

References

[ABF26] Gal Arnon, Dan Boneh, and Giacomo Fenzi. Open problems in list decoding and correlated agreement. Cryptology ePrint Archive, Paper 2026/680, 2026. URL: https://eprint.iacr.org/2026/680.

[AHI+17] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. Ligero: lightweight sublinear arguments without a trusted setup. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 2087–2104. ACM, 2017. DOI: 10.1145/3133956.3134104. URL: https://doi.org/10.1145/3133956.3134104.

[AHK+26] Kasra Abbaszadeh, Hossein Hafezi, Jonathan Katz, and Sarah Meiklejohn. Single-server private outsourcing of zk-snarks. In *IEEE Symposium on Security and Privacy, SP 2026, San Francisco, CA, USA, May 18-21, 2026*, pages 3075–3090. IEEE, 2026. DOI: 10.1109/SP63933.2026.00194. URL: https://doi.org/10.1109/SP63933.2026.00194.

[AZ26] 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.

[BBH+18] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast Reed–Solomon interactive oracle proofs of proximity. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, 14:1–14:17. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. DOI: 10.4230/LIPIcs.ICALP.2018.14. URL: https://doi.org/10.4230/LIPIcs.ICALP.2018.14.

[BBH+19] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Scalable zero knowledge with no trusted setup. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part III*, volume 11694 of *Lecture Notes in Computer Science*, pages 701–732. Springer, 2019. DOI: 10.1007/978-3-030-26954-8_23. URL: https://doi.org/10.1007/978-3-030-26954-8%5C_23.

[BCC+23] Maxime Bombar, Geoffroy Couteau, Alain Couvreur, and Clément Ducros. Correlated pseudorandomness from the hardness of quasi-abelian decoding. In *Advances in Cryptology – CRYPTO 2023, Part IV*, pages 567–601. Springer, 2023. DOI: 10.1007/978-3-031-38551-3_18. URL: https://doi.org/10.1007/978-3-031-38551-3_18.

[BCF+25] Martijn Brehm, Binyi Chen, Ben Fisch, Nicolas Resch, Ron D. Rothblum, and Hadas Zeilberger. Blaze: fast SNARKs from interleaved RAA codes. In *Advances in Cryptology – EUROCRYPT 2025, Part IV*, pages 123–152. Springer, 2025. DOI: 10.1007/978-3-031-91134-7_5. URL: https://doi.org/10.1007/978-3-031-91134-7_5.

[BCG+17] Jonathan Bootle, Andrea Cerulli, Essam Ghadafi, Jens Groth, Mohammad Hajiabadi, and Sune K. Jakobsen. Linear-time zero-knowledge proofs for arithmetic circuit satisfiability. In *Advances in Cryptology – ASIACRYPT 2017, Part III*, pages 336–365. Springer, 2017. DOI: 10.1007/978-3-319-70700-6_12. URL: https://doi.org/10.1007/978-3-319-70700-6_12.

[BCG+18] Elette Boyle, Geoffroy Couteau, Niv Gilboa, and Yuval Ishai. Compressing vector OLE. In David Lie, Mohammad Mannan, Michael Backes, and XiaoFeng Wang, editors, *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*, pages 896–912. ACM, 2018. DOI: 10.1145/3243734.3243868. URL: https://doi.org/10.1145/3243734.3243868.

[BCG+20] Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, and Peter Scholl. Correlated pseudorandom functions from variable-density LPN. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1069–1080. IEEE, 2020. DOI: 10.1109/FOCS46700.2020.00103. URL: https://doi.org/10.1109/FOCS46700.2020.00103.

[BCG+22] 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, Part II*, pages 603–633. Springer, 2022. DOI: 10.1007/978-3-031-15979-4_21. URL: https://doi.org/10.1007/978-3-031-15979-4_21.

[BCG20] Jonathan Bootle, Alessandro Chiesa, and Jens Groth. Linear-time arguments with sublinear verification from tensor codes. In *Theory of Cryptography – TCC 2020, Part II*, pages 19–46. Springer, 2020. DOI: 10.1007/978-3-030-64378-2_2. URL: https://doi.org/10.1007/978-3-030-64378-2_2.

[BCK+22] Eli Ben-Sasson, Dan Carmon, Swastik Kopparty, and David Levit. Scalable and transparent proofs over all large fields, via elliptic curves - (ECFFT part II). In Eike Kiltz and Vinod Vaikuntanathan, editors, *Theory of Cryptography - 20th International Conference, TCC 2022, Chicago, IL, USA, November 7-10, 2022, Proceedings, Part I*, volume 13747 of *Lecture Notes in Computer Science*, pages 467–496. Springer, 2022. DOI: 10.1007/978-3-031-22318-1_17. URL: https://doi.org/10.1007/978-3-031-22318-1%5C_17.

[BCK+23] Eli Ben-Sasson, Dan Carmon, Swastik Kopparty, and David Levit. Elliptic curve fast fourier transform (ECFFT) part I: low-degree extension in time $O(n \log n)$ over all finite fields. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 700–737. SIAM, 2023. DOI: 10.1137/1.9781611977554.CH30. URL: https://doi.org/10.1137/1.9781611977554.ch30.

[BCL22] Jonathan Bootle, Alessandro Chiesa, and Siqi Liu. Zero-knowledge IOPs with linear-time prover and polylogarithmic-time verifier. In *Advances in Cryptology – EUROCRYPT 2022, Part II*, pages 275–304. Springer, 2022. DOI: 10.1007/978-3-031-07085-3_10. URL: https://doi.org/10.1007/978-3-031-07085-3_10.

[BCR+19] Eli Ben-Sasson, Alessandro Chiesa, Michael Riabzev, Nicholas Spooner, Madars Virza, and Nicholas P. Ward. Aurora: transparent succinct arguments for R1CS. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, volume 11476 of *Lecture Notes in Computer Science*, pages 103–128. Springer, 2019. DOI: 10.1007/978-3-030-17653-2_4. URL: https://doi.org/10.1007/978-3-030-17653-2%5C_4.

[BFK+24] 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, Part X*, pages 276–307. Springer, 2024. DOI: 10.1007/978-3-031-68403-6_9. URL: https://doi.org/10.1007/978-3-031-68403-6_9.

[BFM+26] 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. URL: https://eprint.iacr.org/2026/864. to appear in ASIACRYPT 2026.

[BGK+20] Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. DEEP-FRI: sampling outside the box improves soundness. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, 5:1–5:32. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. DOI: 10.4230/LIPIcs.ITCS.2020.5. URL: https://doi.org/10.4230/LIPIcs.ITCS.2020.5.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding: turbo-codes. 1. In *Proceedings of ICC '93 - IEEE International Conference on Communications*, volume 2, 1064–1070 vol.2, 1993. DOI: 10.1109/ICC.1993.397441.

[BM96a] Sergio Benedetto and Guido Montorsi. Design of parallel concatenated convolutional codes. *IEEE Trans. Commun.*, 44(5):591–600, 1996. DOI: 10.1109/26.494303. URL: https://doi.org/10.1109/26.494303.

[BM96b] Sergio Benedetto and Guido Montorsi. Unveiling turbo codes: some results on parallel concatenated coding schemes. *IEEE Trans. Inf. Theory*, 42(2):409–428, 1996. DOI: 10.1109/18.485713. URL: https://doi.org/10.1109/18.485713.

[BMS09] 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.

[BSG26] Mohammadtaghi Badakhshan, Susanta Samanta, and Guang Gong. Faster post-quantum zk-SNARK provers using the LCH polynomial basis. Cryptology ePrint Archive, Paper 2026/1784, 2026. URL: https://eprint.iacr.org/2026/1784.

[Cou20] Geoffroy Couteau. Learning parity with noise vesus linear tests, 2020. URL: https://geoffroycouteau.github.io/posts/LPN-linear-tests/.

[CRR21] Geoffroy Couteau, Peter Rindal, and Srinivasan Raghuraman. Silver: silent VOLE and oblivious transfer from hardness of decoding structured LDPC codes. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part III*, volume 12827 of *Lecture Notes in Computer Science*, pages 502–534. Springer, 2021. DOI: 10.1007/978-3-030-84252-9_17. URL: https://doi.org/10.1007/978-3-030-84252-9%5C_17.

[DI14] Erez Druk and Yuval Ishai. Linear-time encodable codes meeting the Gilbert–Varshamov bound and their cryptographic applications. In *Proceedings of the 5th Innovations in Theoretical Computer Science Conference*, pages 169–182, 2014. DOI: 10.1145/2554797.2554815.

[DJM98] 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.

[Gal62] Robert G. Gallager. Low-density parity-check codes. *IRE Trans. Inf. Theory*, 8(1):21–28, 1962. DOI: 10.1109/TIT.1962.1057683. URL: https://doi.org/10.1109/TIT.1962.1057683.

[Gil52] E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, 1952. DOI: 10.1002/j.1538-7305.1952.tb01393.x.

[GLS+23] Alexander Golovnev, Jonathan Lee, Srinath T. V. Setty, Justin Thaler, and Riad S. Wahby. Brakedown: linear-time and field-agnostic SNARKs for R1CS. In *Advances in Cryptology – CRYPTO 2023, Part II*, pages 193–226. Springer, 2023. DOI: 10.1007/978-3-031-38545-2_7. URL: https://doi.org/10.1007/978-3-031-38545-2_7.

[GM08] 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.

[GNR26] Kobi Gurkan, Andrija Novakovic, and Ron D. Rothblum. Bolt: faster SNARKs from sketched codes. In *Advances in Cryptology – CRYPTO 2026, Part IX*, pages 396–427. Springer, 2026. DOI: 10.1007/978-3-032-35424-2_14. URL: https://doi.org/10.1007/978-3-032-35424-2_14.

[GR11] Alexandre Graell i Amat and Eirik Rosnes. On the minimum distance properties of weighted nonbinary repeat multiple-accumulate codes. In *2011 Information Theory and Applications Workshop*, pages 135–139, 2011. DOI: 10.1109/ITA.2011.5743575.

[Hab23] Ulrich Haböck. Brakedown's expander code. Cryptology ePrint Archive, Paper 2023/769, 2023. URL: https://eprint.iacr.org/2023/769.

[HGQ+26] Kexi Huang, Yanpei Guo, Wenjie Qu, and Jiaheng Zhang. Linear-time, constant-depth blind polynomial commitments from generalized RAA codes, with an end-to-end blind SNARK implementation. Cryptology ePrint Archive, Paper 2026/487, 2026. URL: https://eprint.iacr.org/2026/487.

[JLX+26] Yuhao Jia, Zhe Li, Chaoping Xing, Yizhou Yao, and Chen Yuan. Quasar: a field-agnostic polynomial commitment scheme with polylogarithmic verification from quasi-abelian codes. Cryptology ePrint Archive, Paper 2026/1839, 2026. URL: https://eprint.iacr.org/2026/1839.

[KCL11] Yongsang Kim, Kyungwhoon Cheun, and Hyuntack Lim. Performance of weighted nonbinary repeat-accumulate codes over GF(q) with q-ary orthogonal modulation. *IEEE Transactions on Communications*, 59(5):1208–1212, May 2011. DOI: 10.1109/TCOMM.2011.022811.090194.

[Kha26a] Majid Khabbazian. Linear distance for fixed-row-weight expand–accumulate codes over arbitrary fields. Cryptology ePrint Archive, Paper 2026/1753, 2026. URL: https://eprint.iacr.org/2026/1753.

[Kha26b] Majid Khabbazian. Sharp minimum-distance lower tails for RAA codes. Cryptology ePrint Archive, Paper 2026/1075, 2026. URL: https://eprint.iacr.org/2026/1075.

[Kha26c] Majid Khabbazian. Sharp minimum-distance tails for BAA codes. Cryptology ePrint Archive, Paper 2026/2007, 2026. URL: https://eprint.iacr.org/2026/2007.

[KPR+26] Vladimir Kolesnikov, Stanislav Peceny, Rahul Rachuri, Srinivasan Raghuraman, Peter Rindal, and Harshal Shah. Block-accumulate codes: accelerated linear codes for PCGs and ZK. In *Advances in Cryptology – CRYPTO 2026, Part VIII*, pages 397–429. Springer, 2026. DOI: 10.1007/978-3-032-35418-1_13. URL: https://doi.org/10.1007/978-3-032-35418-1_13.

[KZC07] Jörg Kliewer, Kamil S. Zigangirov, and Daniel J. Costello Jr. New results on the minimum distance of repeat multiple accumulate codes. In *45th Annual Allerton Conference on Communication, Control, and Computing*, pages 1097–1102, 2007.

[KZG10] Aniket Kate, Gregory M. Zaverucha, and Ian Goldberg. Constant-size commitments to polynomials and their applications. In Masayuki Abe, editor, *Advances in Cryptology - ASIACRYPT 2010 - 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings*, volume 6477 of *Lecture Notes in Computer Science*, pages 177–194. Springer, 2010. DOI: 10.1007/978-3-642-17373-8_11. URL: https://doi.org/10.1007/978-3-642-17373-8%5C_11.

[KZK+08] Jörg Kliewer, Kamil S. Zigangirov, Christian Koller, and Daniel J. Costello Jr. Coding theorems for repeat multiple accumulate codes, 2008. arXiv: 0810.3422 [cs.IT]. URL: https://arxiv.org/abs/0810.3422.

[LLX+26] Zhe Li, Hongqing Liu, Chaoping Xing, Yizhou Yao, and Chen Yuan. More efficient SNARKs via quasi-abelian codes: faster, smaller, and field-agnostic. Cryptology ePrint Archive, Paper 2026/939, 2026. URL: https://eprint.iacr.org/2026/939.

[LXL+26] Chongrong Li, Runtian Xu, Yun Li, Yu Yu, and Yuncong Hu. EA codes approaching singleton bound (with application to field-agnostic SNARKs). Cryptology ePrint Archive, Paper 2026/1782, 2026. URL: https://eprint.iacr.org/2026/1782.

[NST25] Vineet Nair, Ashish Sharma, and Bhargav Thankey. BrakingBase - a linear prover, polylogarithmic verifier, field agnostic polynomial commitment scheme. In *Advances in Cryptology – ASIACRYPT 2025, Part V*, pages 427–455. Springer, 2025. DOI: 10.1007/978-981-95-5116-3_14. URL: https://doi.org/10.1007/978-981-95-5116-3_14.

[Pfi03] Henry D. Pfister. *On the Capacity of Finite State Channels and the Analysis of Convolutional Accumulate-m Codes*. PhD thesis, University of California, San Diego, La Jolla, CA, USA, March 2003. URL: https://pfister.ee.duke.edu/thesis/thesis.pdf.

[PR26] Stanislav Peceny and Peter Rindal. Chosen-block BAA codes: fast, near the gilbert–varshamov bound, and field-agnostic. Cryptology ePrint Archive, Paper 2026/1903, 2026. URL: https://eprint.iacr.org/2026/1903.

[PS03] Henry D. Pfister and Paul H. Siegel. The serial concatenation of rate-1 codes through uniform random interleavers. *IEEE Transactions on Information Theory*, 49(6):1425–1438, 2003. DOI: 10.1109/TIT.2003.811907.

[QGZ26] Wenjie Qu, Yanpei Guo, and Jiaheng Zhang. Lightning, field-agnostic super-efficient polynomial commitment scheme. Cryptology ePrint Archive, Paper 2026/258, 2026. URL: https://eprint.iacr.org/2026/258.

[RF09] Chiara Ravazzi and Fabio Fagnani. Spectra and minimum distances of repeat multiple-accumulate codes. *IEEE Transactions on Information Theory*, 55(11):4905–4924, 2009. DOI: 10.1109/TIT.2009.2030459.

[RG11] Eirik Rosnes and Alexandre Graell i Amat. On the analysis of weighted nonbinary repeat multiple-accumulate codes, 2011. DOI: 10.48550/arXiv.1101.5966. arXiv: 1101.5966 [cs.IT]. URL: https://arxiv.org/abs/1101.5966.

[Rin26] Peter Rindal. Expander-based codes at the gilbert–varshamov bound. Cryptology ePrint Archive, Paper 2026/2315, 2026. URL: https://eprint.iacr.org/2026/2315.

[RRT23] Srinivasan Raghuraman, Peter Rindal, and Titouan Tanguy. Expand-convolute codes for pseudorandom correlation generators from LPN. In *Advances in Cryptology – CRYPTO 2023, Part IV*, pages 602–632. Springer, 2023. DOI: 10.1007/978-3-031-38551-3_19. URL: https://doi.org/10.1007/978-3-031-38551-3_19.

[RS60] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, 1960. DOI: 10.1137/0108018. URL: https://doi.org/10.1137/0108018.

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42(6):1723–1731, 1996. DOI: 10.1109/18.556668.

[Var57] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akadamii Nauk*, 117:739–741, 1957.

[XZS22] Tiancheng Xie, Yupeng Zhang, and Dawn Song. Orion: zero knowledge proof with linear prover time. In *Advances in Cryptology – CRYPTO 2022, Part IV*, pages 299–328. Springer, 2022. DOI: 10.1007/978-3-031-15985-5_11. URL: https://doi.org/10.1007/978-3-031-15985-5_11.

[Yan04] Kyeongcheol Yang. Weighted nonbinary repeat-accumulate codes. *IEEE Transactions on Information Theory*, 50(3):527–531, 2004. DOI: 10.1109/TIT.2004.825255.

[ZCF23] Hadas Zeilberger, Binyi Chen, and Ben Fisch. BaseFold: efficient field-agnostic polynomial commitment schemes from foldable codes. Cryptology ePrint Archive, Paper 2023/1705, 2023. URL: https://eprint.iacr.org/2023/1705.

[ZCF24] Hadas Zeilberger, Binyi Chen, and Ben Fisch. BaseFold: efficient field-agnostic polynomial commitment schemes from foldable codes. In *Advances in Cryptology – CRYPTO 2024, Part X*, pages 138–169. Springer, 2024. DOI: 10.1007/978-3-031-68403-6_5. URL: https://doi.org/10.1007/978-3-031-68403-6_5.

[Zei26] Hadas Zeilberger. Khatam: proximity gaps for multilinear evaluation for all linear codes. In *Advances in Cryptology – CRYPTO 2026, Part IX*, pages 592–611. Springer, 2026. DOI: 10.1007/978-3-032-35424-2_20. URL: https://doi.org/10.1007/978-3-032-35424-2_20.

[ZXZ+20] Jiaheng Zhang, Tiancheng Xie, Yupeng Zhang, and Dawn Song. Transparent polynomial delegation and its applications to zero knowledge proof. In *2020 IEEE Symposium on Security and Privacy*, pages 859–876. IEEE, 2020. DOI: 10.1109/SP40000.2020.00052. URL: https://doi.org/10.1109/SP40000.2020.00052.