A senior thesis in partial fulfillment of requirements for the Bachelor of Arts in Mathematics

PROBABILISTIC DISTANCE ANALYSIS OF THE WEIGHTED RAA CODE IN FINITE FIELDS

April 25, 2026

Nicole Xing

Advisors: Benjamin Fisch & Daniel Spielman

Yale University

nicole.xing@yale.edu, benjamin.fisch@yale.edu, daniel.spielman@yale.edu

Contents

1 Introduction

Repeat-accumulate (RA) codes are a class of error correcting codes first introduced in 1998 by Divsalar et al. in 1998 [1]. The RAA code over the binary field was proved to have good distance with cryptographically small failure probability in [2]. However, whether the same holds over finite fields was left as an open problem.

We are interested in using RAA codes for cryptography because they have a fast (generally linear) encoding procedure, good relative distance, and an efficient interactive oracle proof of proximity (IOPP) for multilinear evaluation. This makes them excellent candidates for applications in zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs), as described further in [2]. Since many cryptographic constructions operate over large finite fields, we are highly motivated to characterize the behavior of the RAA code beyond the binary case.

Previous work on RAA codes uses a direct combinatorial approach via the **input-output weight enumerator** (IOWE), which often leads to convoluted bounds that require scripts to confirm. Due to the widely hypothesized strength of the RAA code over large prime fields, we take a probabilistic approach and prove several properties of the code. In particular, our contributions are as follows:

- We show a probabilistic reduction to the binary case: for any fixed input weight and block length, the prime probability of failure is upper bounded by that of the binary.
- We prove several lemmas that characterize the probabilistic behavior of the code, including stochastic dominance by *i.i.d.* Binomial and Exponential random variables, exchangeability, and Markov behavior.
- We prove that a single round of accumulation (RA) over large prime fields is a strong amplifier of relative weight for all inputs above a minimum threshold, which can be paired with an outer code to create an efficient encoding scheme.

Lastly, for illustrative purposes, we provide progress towards a full two-stage bound of the RAA code over all input weights. Although each term of the bound can be shown to be asymptotically small, differences in the required parameters for each regime limit the complete bound from closing fully.

2 Related Work

Repeat-accumulate (RA) codes. Divsalar, Jin, and McEliece [1] introduced *repeat-accumulate* (RA) codes: the simplest turbo-like code, consisting of a rate $1/r$ repetition code followed by a single accumulator connected through a random interleaver. A fundamental limitation, however, is that the minimum distance of RA code ensembles grows at most sub-linearly with block length, making them asymptotically bad [1].

Repeat-accumulate-accumulate (RAA) and repeat multiple-accumulate (RMA) codes. Bazzi, Mahdian, and Spielman [3] derived worst-case upper bounds on the minimum distance of a broad class of turbo-like codes. Among their results, they showed that depth-three serially concatenated codes obtained by placing a repetition code before two accumulators through random permutations (RAA codes)can be asymptotically good. Kliewer et al. [4] proved that for certain parameters, the RAA code exhibits linear minimum distance growth with block length, with the growth rate approaching the Gilbert-Varshamov bound as M increases. In 2025, Brehm et al. [2] constructed Blaze, a multilinear polynomial commitment scheme built on RAA codes. In particular, they proved that the probability of distance failure for the RAA code over binary fields is cryptographically small.

Expand-accumulate (EA) codes. Boyle et al. [5] introduced *expand-accumulate* (EA) codes as a single-accumulation alternative to RA codes, designed for pseudorandom correlation generator (PCG) applications. The minimum distance of binary EA codes was analyzed in [5] via a random walk on the associated expander graph, showing that EA codes achieve constant relative distance with high probability. Our analysis of the Markov cancellations in the RAA code draw from this field of work.

Block et al. [6] generalized EA codes to arbitrary finite fields $\mathbb{F}_q$ and used them to construct field-agnostic SNARKs following the Brakedown framework. The distance analysis introduces a support-based union bound that avoids the $(q-1)$-factor blowup inherent in naive per-value union bounds over nonbinary alphabets. This decomposition technique is directly related to the proof strategy employed in this work, which utilizes weight randomization over the finite field to construct the same union bound.

3 Problem Setup

Notation. Let $x \in \mathbb{F}_q^n$ denote the input message, where the regime of interest has $n \in \{2^{10}, 2^{15}, \ldots\}$ be quite large, following the conventions of [2][7]. IFor a constant $r \in \mathbb{N}$, the **rate** of the code $1/r$ repeats the entries of x by r, which we represent with the matrix $R \in \{0,1\}^{rn \times n}$. $R_{i,j} = 1$ if and only if $\lfloor j/r \rfloor + 1 = i$. After repetition, we have the $N = rn$ long codeword m. In this work, we restrict the field order q to be larger than N; this is in line with many cryptographic constructions over prime fields of large order.

Each accumulation step of the repeat-multiple-accumulate code then permutes m and randomizes the nonzero entries of m. For any permutation matrix $\pi : [N] \to [N]$, its corresponding matrix Π follows $\Pi_{i,j} = 1$ if and only if $\pi(i) = j$. The weight-randomizing matrix V is a diagonal matrix in $\mathbb{F}_q^{N \times N}$, where each diagonal element is sampled uniformly randomly from F_q. After permutation and randomization, $V\Pi m$ is fed into an **accumulator**, represented by the matrix $A \in \{0,1\}^{N \times N}$ where $A_{i,j} = 1$ if and only if $i \geq j$. At each index i of $V\Pi m \in \mathbb{F}_q^N$, we can imagine the output of $AV\Pi m$ at the corresponding index as the prefix sum of $V\Pi m$ from 0 to i modulo q.

The permute-randomize-accumulate steps are repeated as needed: once for the RA code, twice for the RAA code, and so on. Overall, RAA generator matrix takes the form

$$G_{\text{RAA}} = A \cdot V_2 \cdot \Pi_2 \cdot A \cdot V_1 \cdot \Pi_1 \cdot R$$

where Π_1, Π_2 are the selected permutation matrices, and V_1, V_2 are the weight-randomizing matrices. Encoding proceeds through the following sequence of intermediate codewords:

$$x \xrightarrow{R} m \xrightarrow{\Pi_1, V_1} m_1 \xrightarrow{A} m_1^A \xrightarrow{\Pi_2, V_2} m_2 \xrightarrow{A} m_2^A.$$

We track the Hamming weight at each stage. Setting $\text{wt}(x) = w_0$, the successive weights are

$$\text{wt}(m) = r \cdot w_0 =: w_1, \qquad \text{wt}(m_1^A) =: L_1, \qquad \text{wt}(m_2^A) =: L_2.$$

We are interested in upper bounding the probability of distance failure for the RAA code; that is, for some fixed n, r, q, and cutoff relative weight $\delta \in (0, 1)$:

$$\mathbb{P}_{\Pi_1, V_1, \Pi_2, V_2}(\exists x \in F_q^n \mid \mathrm{wt}(\mathrm{RAA(x)}) < \delta N).$$

We use the uniform interleaver approach in this analysis, proving distance with respect to permutations Π and randomizations V that are sampled uniformly randomly.

3.1 RAA Urn Model

We can consider each round of the weighted RAA code with the following intuition. For an input x of fixed weight w_0, we draw without replacement from an urn with $N = r \cdot w_0$ balls inside. w_1 of the balls are white, each labeled with a random number from $\mathbb{F}_q$, and the remaining $N - w_1$ are black.

The first accumulation step can be understood as follows: after processing each successive draw from the urn, we keep track of the status of a "run." If no run is active and we draw a white ball, we immediately start a run, regardless of the value on it. If a run is active, black draws do not change the run status, and white draws deactivate a run with probability $1/(q-1)$. Consecutive black draws during a deactivated run are considered "gaps." After the urn has been drawn to completion, the total weight of the accumulator output, L_1, is the sum of all runs.[1]

The second permutation, weight randomization, and accumulation can be understood similarly, except the urn now contains L_1 white balls, which can take on any value from $\{1, \dots, N\}$ instead of $\{r, 2r, 3r, \dots, N\}$.

We will take a probabilistic view of the RAA code to derive simple bounds over the prime field. Since both accumulation steps sample without replacement and are themselves dependent on the outcomes of the permutations, there are many interactions to account for. We first derive a number of lemmas that characterize the behavior of the the random interleaver in a general and explicit way.

3.2 Nonzero Field Elements as Markov Run Interactions

First, we analyze the behavior of **runs**, or how white balls interact with each other.

[1] The length of a single run starts with the activation draw, but does not include the terminating draw. That is, if the activating and deactivating white balls are drawn at indices i and j, the length of the run is $j - i$, not $j - i + 1$.

Lemma 1 (Accumulator run status is Markov.)**.** *Define random variables R_i for $i = 1, \ldots, w_1$ as an indicator on the event that the prefix sum after processing the i-th white ball from the urn is nonzero. $R_1 = 1$ deterministically. Then rest of the (R_i) form a Markov chain with transition probabilities*

$$\mathbb{P}(R_i = 1 | R_{i-1} = 1) = (q-2)/(q-1)$$
$$\mathbb{P}(R_i = 0 | R_{i-1} = 1) = 1/(q-1)$$
$$\mathbb{P}(R_i = 1 | R_{i-1} = 0) = 1$$
$$\mathbb{P}(R_i = 0 | R_{i-1} = 0) = 0.$$

Proof. By definition, the above follows the Markov property. □

Lemma 2 (Cancellation count bound)**.** *Over $\mathbb{F}_q$, the cancellation set I satisfies $K = |I| \leq \lfloor w/2 \rfloor$ almost surely. Over $\mathbb{F}_2$, the cancellation set is the deterministic set $I_2 = \{2, 4, 6, \ldots, 2\lfloor w/2 \rfloor\}$, so $K_2 = \lfloor w/2 \rfloor$ exactly.*

Proof. The transition $R_i = 0$ forces $R_{i+1} = 1$ deterministically, since a zero prefix sum is immediately broken by any nonzero subsequent input. Then no two consecutive indices can both lie in I, and any antichain of $\{1, \ldots, w\}$ with no two adjacent elements has size at most $\lfloor w/2 \rfloor$. Hence $K \leq \lfloor w/2 \rfloor$.

Over $\mathbb{F}_2$, every nonzero field element is 1, so cancellations occur deterministically at every other white draw. Therefore, $R_i = 0$ if and only if i is even, giving $I_2 = \{2, 4, \ldots, 2\lfloor w/2 \rfloor\}$ and $K_2 = \lfloor w/2 \rfloor$. □

3.3 Zeros as Hypergeometric Gap Lengths

Next, we handle **gaps**: how black draws between white draws affect the accumulator output.

Lemma 3 (Black draw gaps are Negative Hypergeometric.)**.** *Define the random variables g_i as the length of the gap between the i-th and $(i+1)$-th white ball. There are $w_1 + 1$ such gaps, where the extra variable g_0 denotes the gap before the first white draw. Gaps are defined as consecutive black draws without replacement until encountering a white ball, so we have*

$$g_i \sim NHG\left(N - \left(\sum_{j<i} g_j\right) - i, w_1 - i, 1\right).$$

Proof. Again, this follows directly from definitions. □

For a given permutation Π applied to some vector v, we will refer to the indices of the resulting nonzero positions of Πv as the **support** of Πv. A similar notion of "BadSup" and "BadVal" is used in [6].

Let $s_1 < s_2 < \cdots < s_w$ denote the ordered support positions of the permuted vector. Define the **gap sequence**

$$g_0 := s_1 - 1, \qquad g_i := s_{i+1} - s_i - 1 \text{ for } 1 \le i \le w-1, \qquad g_w := N - s_w. \tag{1}$$

Each $g_i \ge 0$ and $\sum_{i=0}^{w} g_i = N - w$. The gaps are determined entirely by Π.

Lemma 4 (Exchangeability of the gap sequence)**.** *For a fixed input weight w, the gap sequence $(g_0, g_1, \ldots, g_w)$ is exchangeable: its distribution is invariant under all permutations of its coordinates. In particular, $(g_1, \ldots, g_w)$ is exchangeable conditioned on on $g_0 = c$ for any $c \in \{0, 1, \ldots, N-w\}$.*

Proof. Since Π is a uniformly random permutation of $[N]$, the support $\{s_1, \ldots, s_w\}$ is a uniformly random w-element subset of $[N]$. The mapping from ordered supports to gap sequences defined by (1) is a bijection between the $\binom{N}{w}$ ordered supports and the set $\mathcal{G} := \left\{(g_0, \ldots, g_w) \in \mathbb{N}^{w+1} : \sum_{i=0}^{w} g_i = N - w\right\}$. Since each ordered support is equally likely, the gap vector is uniformly distributed on $\mathcal{G}$.

The uniform distribution on $\mathcal{G}$ is invariant under all permutations of coordinates, since any permutation of a vector in $\mathcal{G}$ yields another vector in $\mathcal{G}$, and each such vector has the same probability $1/|\mathcal{G}| = 1/\binom{N}{w}$. Hence $(g_0, \ldots, g_w)$ is exchangeable.

For the conditional statement: fixing $g_0 = c$ restricts to the sub-simplex $\mathcal{G}_c := \{(g_1, \ldots, g_w) \in \mathbb{N}^w : \sum_{i=1}^{w} g_i = N - w - c\}$, on which the conditional distribution of $(g_1, \ldots, g_w)$ is again uniform by the same bijection argument applied to w positions in $\{c+1, \ldots, N\}$. Uniformity on $\mathcal{G}_c$ implies exchangeability of $(g_1, \ldots, g_w)$ given $g_0 = c$. □

Lemma 5 (Independence of cancellations and gaps)**.** *The cancellation set I is independent of the gap sequence $(g_0, g_1, \ldots, g_w)$.*

Proof. The gap sequence is determined entirely by the support positions $s_1 < \cdots < s_w$, which are the output of the permutation Π. The cancellation set I is determined entirely by the values $m_{s_1}, \ldots, m_{s_w} \in \mathbb{F}_q^*$, which are the outputs of the weight-randomization matrix V. By construction, Π and V are independent (they are separate random matrices in the code definition), so the gap sequence and the cancellation set are independent. $\square$

4 Probabilistic Reduction from Prime to Binary Worst-Case

From the previous section, we showed that nonzero/white draws in the prime case have a nondeterministic cancellation behavior. A natural next step is to wonder whether the binary accumulation step is a worst-case bound for any prime accumulation step. We prove this formally:

Theorem 1 (Single-stage worst-case binary reduction.)**.** *Let $L^{(q)}$ denote the output weight of the accumulator over $\mathbb{F}_q$ (for any prime $q \geq 2$) and $L^{(2)}$ the output weight over $\mathbb{F}_2$, both with the same input weight w and the same uniform random permutation distribution. Then*

$$\mathbb{P}(L^{(q)} \leq t) \leq \mathbb{P}(L^{(2)} \leq t) \qquad \textit{for all } t \geq 0. \tag{2}$$

Equivalently, $L_q \geq_{\mathrm{st}} L_2$.

Proof. We consider a single accumulation stage for fixed input weight $w \geq 1$ and block length N.

The accumulator output weight is

$$\begin{aligned} L &= \sum_{i=1}^{w} R_i(1+g_i) \\ &= (N - g_0) - \sum_{i=1}^{w} \mathbb{1}[R_i = 0]\,(1+g_i). \end{aligned} \tag{3}$$

Let $I := \{i \in \{2, \ldots, w\} : R_i = 0\}$ be the cancellation set with $K := |I|$.

Condition on $g_0 = c$. From 3, the event that the accumulator output is low, $\{L \leq t\}$, is equivalent to there being many cancellations, i.e.,

$$\left\{\sum_{i \in I}(1+g_i) \geq (N-c) - t\right\}. \tag{4}$$

Therefore, we can equivalently bound $\mathbb{P}\left(\sum_{i\in I}(1+g_i)\geq s \,\middle|\, |I|=K, g_0=c\right)$. By the law of total probability, this is a sum over fixed cancellation realizations J of size K:

$$\sum_{J\subseteq\{2,\ldots,w\}:|J|=K} \mathbb{P}\left(\sum_{i\in I}(1+g_i)\geq s \,\middle|\, I=J, g_0=c\right)\cdot\mathbb{P}\left(I=J \,\middle|\, |I|=K\right). \tag{5}$$

By Lemma 5, I is independent of $(g_1,\ldots,g_w)$, so for any $J=\{j_1<\cdots<j_K\}\subseteq\{2,\ldots,w\}$,

$$\mathbb{P}\left(\sum_{i\in I}(1+g_i)\geq s \,\middle|\, I=J, g_0=c\right)=\mathbb{P}\left(\sum_{k=1}^{K}(1+g_{j_k})\geq s \,\middle|\, g_0=c\right).$$

Furthermore, Lemma 4, tells us that $(g_1,\ldots,g_w)$ is exchangeable conditionally on $g_0=c$. Therefore, the joint distribution of $(g_{j_1},\ldots,g_{j_K})$ given $g_0=c$ is the same as that of $(g_1,\ldots,g_K)$ on the same condition:

$$\mathbb{P}\left(\sum_{k=1}^{K}(1+g_{j_k})\geq s \,\middle|\, g_0=c\right)=\mathbb{P}\left(\sum_{k=1}^{K}(1+g_k)\geq s \,\middle|\, g_0=c\right).$$

Since the above holds for *every* J with $|J|=K$, and I is independent of the gaps, it holds unconditionally on the specific realization of I. Then 5 is simply $\mathbb{P}\left(\sum_{k=1}^{K}(1+g_k)\geq s \,\middle|\, g_0=c\right)$.

Since each $(1+g_k)\geq 1>0$ almost surely and $K\leq\lfloor w/2\rfloor$ by Lemma 2, the partial sum $\sum_{k=1}^{K}(1+g_k)$ is non-decreasing in K almost surely. Therefore, for every fixed value of $(g_1,\ldots,g_w)$:

$$\sum_{k=1}^{K}(1+g_k)\leq\sum_{k=1}^{\lfloor w/2\rfloor}(1+g_k).$$

Taking probabilities:

$$\mathbb{P}\left(\sum_{k=1}^{K}(1+g_k)\geq s \,\middle|\, g_0=c\right)\leq\mathbb{P}\left(\sum_{k=1}^{\lfloor w/2\rfloor}(1+g_k)\geq s \,\middle|\, g_0=c\right).$$

But by the same argument[2], the right hand side of is precisely the binary distribution failure probability.

Setting $s = (N - c) - t$ and recalling (4):

$$\mathbb{P}(L^{(q)} \le t \mid g_0 = c) \le \mathbb{P}(L^{(2)} \le t \mid g_0 = c).$$

The gap $g_0 = s_1 - 1$ is determined entirely by the permutation Π. Note that g_0 has the *same distribution* under $\mathbb{F}_q$ and $\mathbb{F}_2$: both ensembles use the same uniform random permutation Π, and g_0 depends only on the support positions, not on the field values assigned by V.

Therefore, writing $\mu_c := \mathbb{P}(g_0 = c)$ for the common distribution, the law of total probability gives:

$$\mathbb{P}\big(L^{(q)} \le t\big) = \sum_{c=0}^{N-w} \mu_c \cdot \mathbb{P}\big(L^{(q)} \le t \mid g_0 = c\big) \le \sum_{c=0}^{N-w} \mu_c \cdot \mathbb{P}\big(L^{(2)} \le t \mid g_0 = c\big) = \mathbb{P}\big(L^{(2)} \le t\big),$$

establishing (2). $\square$

We remark that using 1 to reduce the two-stage prime failure probability to that of [2] is not straightforward, as

$$\begin{aligned}
\mathbb{P}\left(L_2^{(q)} \le \delta N\right) &= \sum_{w_2=1}^{N} \mathbb{P}\left(L_2^{(q)} \le \delta N \,\middle|\, L_1^{(q)} = w_2\right) \mathbb{P}\left(L_1^{(q)} = w_2\right) \\
&\le \sum_{w_2=1}^{N} \mathbb{P}\left(L_2^{(2)} \le \delta N \,\middle|\, L_1^{(2)} = w_2\right) \mathbb{P}\left(L_1^{(q)} = w_2\right)
\end{aligned}$$

cannot be trivially bounded by the binary LoTP expression. This motivates the following section, which provides more concrete bounds comparing a single stage of prime to binary RA.

[2] Over $\mathbb{F}_2$, Lemma 2 gives the deterministic cancellation set $I_2 = \{2, 4, \dots, 2\lfloor w/2 \rfloor\}$, so

$$C_2 = \sum_{k=1}^{\lfloor w/2 \rfloor} (1 + g_{2k}).$$

By Lemma 4, $(g_1, \dots, g_w)$ is exchangeable conditionally on $g_0 = c$, so the joint distribution of $(g_2, g_4, \dots, g_{2\lfloor w/2 \rfloor})$ given $g_0 = c$ equals the joint distribution of $(g_1, g_2, \dots, g_{\lfloor w/2 \rfloor})$ given $g_0 = c$. Therefore:

$$\mathbb{P}\Big(C_2 \ge s \,\Big|\, g_0 = c\Big) = \mathbb{P}\Big(\sum_{k=1}^{\lfloor w/2 \rfloor} (1 + g_{2k}) \ge s \,\Big|\, g_0 = c\Big) = \mathbb{P}\Big(\sum_{k=1}^{\lfloor w/2 \rfloor} (1 + g_k) \ge s \,\Big|\, g_0 = c\Big).$$

5 Probabilistic Single-round Amplification Bound

We obtain explicit bounds for a single RA stage over $\mathbb{F}_q$. Recall the cancellation formula from the proof of Theorem 1:

$$L_1 = (N - g_0) - \sum_{i \in I} (1 + g_i), \tag{6}$$

where $I \subseteq \{2, \ldots, w_1\}$ is the cancellation set and $K = |I|$. We will split by cases of fixed K to bound $\mathbb{P}(\exists x \in \mathbb{F}_q^n \mid L_1 < \delta N)$.

Proposition 1 (Zero-cancellation term)**.** *Fix $\phi, \delta \in (0,1)$ with $\phi < \delta$ and $r \geq 1$, and suppose $\frac{H(\phi)}{r} + \phi \log \delta < 0$. Then for all $w_0 \geq \phi N/r$ (so that $w_1 = rw_0 \geq \phi N$),*

$$\mathbb{P}(K = 0,\, L_1 < \delta N \mid w_0) \;\leq\; \delta^{w_1}.$$

In particular, the union bound over all input weights $w_0 \geq \phi N/r$,

$$\sum_{w_0 \,\geq\, \phi N/r} \binom{n}{w_0} \cdot \mathbb{P}(K = 0,\, L_1 < \delta N \mid w_0), \tag{7}$$

decays as $O(N) \cdot e^{-cN}$, where $c = |H(\phi)/r + \phi \log \delta| > 0$.

Proof. When $K = 0$ there are no cancellations, so (6) reduces to $L_1 = N - g_0$, and the event $\{L_1 < \delta N\}$ becomes $\{g_0 > (1-\delta)N\}$. Since the w_1 draws are sampled without replacement from $[N]$, the gap g_0 is the smallest order statistic, and the event $\{g_0 > (1-\delta)N\}$ is exactly the event that all w_1 draws land in $\{(1-\delta)N + 1, \ldots, N\}$, a set of size $\lfloor \delta N \rfloor$. Therefore

$$\mathbb{P}(K = 0,\, L_1 < \delta N \mid w_0) \;=\; \mathbb{P}(g_0 > (1-\delta)N) \;=\; \frac{\binom{\lfloor \delta N \rfloor}{w_1}}{\binom{N}{w_1}} \;=\; \prod_{j=0}^{w_1 - 1} \frac{\lfloor \delta N \rfloor - j}{N - j}. \tag{8}$$

Each factor satisfies $(\lfloor \delta N \rfloor - j)/(N - j) \leq \lfloor \delta N \rfloor / N \leq \delta$, so the product is at most δ^{w_1}.

For the union bound, the dominant summand in (7) is at $w_0 = \lceil \phi N/r \rceil$ (larger w_0 gives larger $w_1 = rw_0$ and hence smaller δ^{w_1}). Using $\binom{n}{w_0} \leq e^{nH(\phi)} = e^{NH(\phi)/r}$ and $w_1 \geq \phi N$, each summand is at most

$$e^{NH(\phi)/r} \cdot \delta^{\phi N} = e^{N(H(\phi)/r \,+\, \phi \log \delta)} = e^{-cN}.$$

Summing over at most $O(N)$ values of w_0 gives $O(N) \cdot e^{-cN}$. $\square$

We now handle the $K \geq 1$ correction term.

Proposition 2 (Cancellation term vanishes)**.** *Fix $\phi, \delta \in (0,1)$ with $\phi < \delta$ and $r \geq 1$, and suppose $\frac{H(\phi)}{r} + \phi \log \delta < 0$. Then for all $w_0 \geq \phi N/r$ (so that $w_1 = rw_0 \geq \phi N$),*

$$\mathbb{P}(K \geq 1,\ L_1 < \delta N \mid w_0) \ \leq\ \delta^{w_1} \cdot F(BN),$$

where $B := \phi^2(1-\delta)/((1-\phi)\delta)$ and $F(x) := \sum_{k=1}^{\infty} \frac{k+1}{(k!)^2} x^k$. In particular, the union bound over all input weights $w_0 \geq \phi N/r$,

$$\sum_{w_0 \,\geq\, \phi N/r} \binom{n}{w_0} \cdot \mathbb{P}(K \geq 1,\ L_1 < \delta N \mid w_0),$$

decays as $e^{-cN/2}$ for all $N \geq N_0$, where $c = \left|\frac{H(\phi)}{r} + \phi \log \delta\right| > 0$ and $N_0 = 16B/c^2$.

Proof. We bound $\mathbb{P}(K = k,\, L_1 < \delta N \mid w_0)$ for each fixed $k \geq 1$ and then sum over k.

Reduce to order statistics. Fix any realization $I = \{i_1, \ldots, i_k\} \subseteq \{2, \ldots, w_1\}$ with $|I| = k$. From (6),

$$\{L_1 < \delta N\} \ =\ \Big\{g_0 + \textstyle\sum_{j \in I} g_j > (1-\delta)N - k\Big\}.$$

The gaps $(g_0, g_{i_1}, \ldots, g_{i_k})$ are still random even after fixing I: by Lemma 5, the cancellation set I is independent of the gap sequence, so conditioning on $I = \{i_1, \ldots, i_k\}$ does not alter the distribution of the gaps. Therefore

$$\mathbb{P}\big(L_1 < \delta N \;\big|\; K = k,\ I = \{i_1, \ldots, i_k\}\big) \ =\ \mathbb{P}\big(g_0 + g_{i_1} + \cdots + g_{i_k} > (1-\delta)N - k\big).$$

By Lemma 4, the gaps are exchangeable, so the right-hand side takes the same value for every choice of k distinct indices from $\{2, \ldots, w_1\}$. In particular, it equals $\mathbb{P}(g_0 + g_1 + \cdots + g_k > (1-\delta)N - k)$, which does not depend on I. Since the conditional probability is constant over all realizations of I with $|I| = k$, removing the conditioning on I gives

$$\mathbb{P}(L_1 < \delta N \mid K = k,\, w_0) \;=\; \mathbb{P}\big(g_0 + g_1 + \cdots + g_k > (1-\delta)N - k\big).$$

The telescoping identity $g_0 + g_1 + \cdots + g_k = s_{k+1} - (k+1)$ (which follows directly from the definition (1) of the gap sequence) simplifies the right-hand side to $\mathbb{P}(s_{k+1} > (1-\delta)N + 1)$.

Since s_{k+1} is an integer, this equals $\mathbb{P}(s_{k+1} \geq \lfloor (1-\delta)N \rfloor + 2)$, which is at most $\mathbb{P}(s_{k+1} > \lfloor (1-\delta)N \rfloor)$. Setting $T := \lfloor (1-\delta)N \rfloor$, the event $\{s_{k+1} > T\}$ is equivalent to the event that at least $w_1 - k$ of the w_1 draws lie in $\{T+1, \ldots, N\}$. This set has exactly $N - T = N - \lfloor (1-\delta)N \rfloor = \lceil \delta N \rceil$ elements.

Since the w_1 draws are sampled without replacement from $[N]$, the count of draws landing in $\{T+1, \ldots, N\}$ is $\mathrm{Hyp}(N, \lceil \delta N \rceil, w_1)$, so

$$\mathbb{P}(L_1 < \delta N \mid K = k,\, w_0) \;\leq\; \mathbb{P}\big(\mathrm{Hyp}(N,\, \lceil \delta N \rceil,\, w_1) \geq w_1 - k\big). \tag{9}$$

Bound hypergeometric tail with dominant term. Write the right side of (9) as a sum of PMF values, setting $M := \lceil \delta N \rceil$ for brevity:

$$\mathbb{P}\big(\mathrm{Hyp}(N, M, w_1) \geq w_1 - k\big) = \sum_{j=w_1-k}^{w_1} \frac{\binom{M}{j}\binom{N-M}{w_1-j}}{\binom{N}{w_1}},$$

where the upper limit is w_1 because $w_1 = \phi N < \delta N \leq M$ (since $\phi < \delta$), so $\min(w_1, M) = w_1$. The sum therefore has exactly $k+1$ terms. The ratio of consecutive terms is

$$\frac{\binom{M}{j+1}\binom{N-M}{w_1-j-1}}{\binom{M}{j}\binom{N-M}{w_1-j}} = \frac{(M-j)(w_1-j)}{(j+1)(N-M-w_1+j+1)}.$$

At $j = w_1 - k$ and $w_1 = \phi N$, $M \approx \delta N$, this ratio is approximately $(\delta - \phi)k \,/\, (\phi(1-\delta)N) \to 0$ as $N \to \infty$. As j increases past $w_1 - k$, the factor $w_1 - j$ in the numerator decreases strictly while $j+1$ in the denominator increases strictly, so the ratio is decreasing in j. Thus, once the ratio is less than 1 at $j = w_1 - k$ — which holds for all sufficiently large N — it remains less than 1 for all subsequent terms. Since the sum has $k+1$

terms, each at most the $j = w_1 - k$ term:

$$\mathbb{P}\big(\mathrm{Hyp}(N, M, w_1) \geq w_1 - k\big) \;\leq\; (k+1) \cdot \frac{\binom{M}{w_1-k}\binom{N-M}{k}}{\binom{N}{w_1}}. \tag{10}$$

Product bound dominant term. Write the dominant-term ratio as With $M = \lceil \delta N \rceil$:

$$\frac{\binom{M}{w_1-k}\binom{N-M}{k}}{\binom{N}{w_1}} \;=\; \underbrace{\frac{\binom{M}{w_1-k}}{\binom{N}{w_1-k}}}_{\leq (\delta+1/N)^{w_1-k}} \cdot \underbrace{\frac{\binom{N}{w_1-k}}{\binom{N}{w_1}}}_{\leq (\phi/(1-\phi))^k} \cdot \underbrace{\binom{N-M}{k}}_{\leq ((1-\delta)N)^k/k!}.$$

The first factor satisfies

$$\binom{M}{w_1-k}\Big/\binom{N}{w_1-k} = \prod_{j=0}^{w_1-k-1}(M-j)/(N-j) \leq (M/N)^{w_1-k}.$$

Since $M = \lceil \delta N \rceil \leq \delta N + 1$, we have $M/N \leq \delta + 1/N$, so this factor is at most $(\delta + 1/N)^{w_1-k}$. For $w_1 \leq \phi N$ this is $\leq \delta^{w_1-k}(1 + 1/(\delta N))^{\phi N} \leq \delta^{w_1-k}e^{\phi/\delta}$, a constant-factor correction. For the second factor, write

$$\binom{N}{w_1-k}\Big/\binom{N}{w_1} = \prod_{j=0}^{k-1}(w_1-j)/(N-w_1+j+1).$$

Each factor is at most $\frac{w_1}{N-w_1} = \frac{\phi}{1-\phi}$ because $(w_1-j)(N-w_1) \leq w_1(N-w_1+j+1)$ reduces to $-jN \leq w_1$, which holds for all $j, N, w_1 \geq 0$. The third factor uses $N - M \leq (1-\delta)N$ and $\binom{m}{k} \leq m^k/k!$. Absorbing $e^{\phi/\delta}$ into the implicit constant and combining:

$$\mathbb{P}(L_1 < \delta N \mid K = k,\, w_0) \;\lesssim\; \frac{k+1}{k!} \cdot \delta^{w_1-k} \cdot \left(\frac{(1-\delta)\phi\, N}{1-\phi}\right)^k, \tag{11}$$

where the implicit constant $e^{\phi/\delta}$ is independent of N and k.

Bound on $\mathbb{P}(K = k)$. Each cancellation requires an independent field value to equal zero, an event of probability $1/(q-1)$. Since there are at most $\binom{w_1}{k}$ ways to choose k cancellation sites and the events at distinct sites are independent (Lemma 5),

$$\mathbb{P}(K = k) \;\le\; \binom{w_1}{k} \frac{1}{(q-1)^k} \;\le\; \frac{w_1^k}{k!\,(q-1)^k}.$$

With $q = N$ and $w_1 \le \phi N$, this gives $\mathbb{P}(K = k) \le (\phi N/(N-1))^k/k!$. For $N \ge 2$ we have $N/(N-1) \le 2$, so $\mathbb{P}(K = k) \le (2\phi)^k/k!$. In the estimates below we track the factor as $(\phi N/(N-1))^k/k!$ and note it converges to $\phi^k/k!$ as $N \to \infty$.

Joint bound and summation over k. Multiplying (11) by the Step 4 bound and using $w_1 \ge \phi N$:

$$\mathbb{P}(K = k,\ L_1 < \delta N \mid w_0) \;\le\; \frac{\phi^k}{k!} \cdot \frac{k+1}{k!} \cdot \delta^{w_1 - k} \cdot \left(\frac{(1-\delta)\phi\, N}{1-\phi}\right)^k + O(N^{k-1}) \cdot \delta^{w_1},$$

where the error from replacing $\phi N/(N-1)$ by ϕ is of lower order. Setting $B = \phi^2(1-\delta)/((1-\phi)\delta)$, one computes

$$\frac{\phi^k}{k!} \cdot \left(\frac{(1-\delta)\phi N}{1-\phi}\right)^k = \frac{(B\delta N)^k}{k!},$$

and factoring $\delta^{w_1-k} \cdot (B\delta N)^k = \delta^{w_1} \cdot (BN)^k$ gives

$$\mathbb{P}(K = k,\ L_1 < \delta N \mid w_0) \;\lesssim\; \frac{k+1}{(k!)^2} \cdot (BN)^k \cdot \delta^{w_1} \;=:\; a_k \cdot \delta^{w_1}.$$

Summing over $k \ge 1$ defines

$$F(x) := \sum_{k=1}^{\infty} \frac{k+1}{(k!)^2}\, x^k,$$

which converges for all x (consecutive ratio $(k+2)x/(k+1)^2 \to 0$), giving

$$\mathbb{P}(K \ge 1,\ L_1 < \delta N \mid w_0) \;\le\; \delta^{w_1} \cdot F(BN). \tag{12}$$

Union bound over messages. The number of inputs of weight w_0 is $\binom{n}{w_0}$. The dominant contribution in the sum over $w_0 \ge \phi N/r$ comes from $w_0 = \lceil \phi N/r \rceil$ (since larger w_0 means larger $w_1 = r w_0$, hence smaller

δ^{w_1}). With $n = N/r$ and $\binom{n}{w_0} \le \exp(nH(\phi)) = \exp(NH(\phi)/r)$:

$$\binom{n}{w_0} \cdot \mathbb{P}(K \ge 1,\, L_1 < \delta N \mid w_0) \;\le\; \exp\!\left(\frac{NH(\phi)}{r} + w_1 \log \delta\right) \cdot F(BN).$$

The exponent satisfies

$$\frac{NH(\phi)}{r} + \phi N \log \delta \;=\; -cN, \qquad c := \left|\frac{H(\phi)}{r} + \phi \log \delta\right| > 0,$$

where $c > 0$ by hypothesis. For the F factor, the series satisfies $F(x) = \sqrt{x}\, I_1(2\sqrt{x}) + I_0(2\sqrt{x}) - 1$ in terms of modified Bessel functions, so $F(x) \le e^{C\sqrt{x}}$ for large x, with $C = 2$. Thus $F(BN) \le e^{C\sqrt{BN}} = e^{2\sqrt{BN}}$, growing like $e^{C'\sqrt{N}}$. Since $-cN + C'\sqrt{N} \le -cN/2$ for all $N \ge (2C'/c)^2 = 16B/c^2 = N_0$,

$$\binom{n}{w_0} \cdot \mathbb{P}(K \ge 1,\, L_1 < \delta N \mid w_0) \;\le\; e^{-cN/2},$$

and the sum over polynomially many values of w_0 contributes at most $O(N) \cdot e^{-cN/2} \to 0$ as $N \to \infty$. □

Theorem 2 (Single-round amplification bound)**.** *Fix $\phi, \delta \in (0,1)$ with $\phi < \delta$ and $r \ge 1$, and suppose*

$$c \;:=\; \left|\frac{H(\phi)}{r} + \phi \log \delta\right| > 0.$$

Let $B := \phi^2(1-\delta)/((1-\phi)\delta)$ and $N_0 := 16B/c^2$. Then for all $w_0 \ge \phi N/r$ (so that $w_1 = rw_0 \ge \phi N$) and all $N \ge N_0$,

$$\mathbb{P}(L_1 < \delta N \mid w_0) \;\le\; \delta^{w_1}\left(1 + F(BN)\right), \tag{13}$$

where $F(x) := \sum_{k=1}^{\infty} \frac{k+1}{(k!)^2}\, x^k$ converges for all $x \ge 0$. In particular, summing over all input weights,

$$\sum_{w_0 \,\ge\, \phi N/r} \binom{n}{w_0} \cdot \mathbb{P}(L_1 < \delta N \mid w_0) \;\le\; O(N) \cdot e^{-cN/2} \;\to\; 0 \quad \textit{as } N \to \infty.$$

Proof. Decompose on the cancellation count $K = |I|$:

$$\mathbb{P}(L_1 < \delta N \mid w_0) = \mathbb{P}(K = 0,\, L_1 < \delta N \mid w_0) + \mathbb{P}(K \ge 1,\, L_1 < \delta N \mid w_0).$$

Add the two bounds obtained from Propositions 1 and 2 and use the union bound argument for each. □

5.1 Explicit Bounds for Single Round Repeat-Accumulate

The Theorem 2 bound $e^{-cN/2}$ only kicks in for $N \geq N_0$. Near $\phi_{\min}$, the exponent $c \approx 0$ which makes $N_0 \propto 1/c^2 \to \infty$. So even if we select a ϕ that is theoretically above threshold, it may require an astronomically large N before the bound is useful. Therefore, in practice we must use a slightly higher ϕ, which we denote as ϕ_{prac}: the smallest value such that $N_0(\phi) \leq N = rn$. The table below gives example $\phi_{\min}, \phi_{\text{prac}}$ for different parameters n, r, and δ.

While the union bound values are still weak for certain small n and ϕ_{prac}, we obtain extremely small failure probabilities for most choices of parameters. ϕ can also be increased above ϕ_{prac} to further improve the failure probability. Bold table rows indicate parameters where the union bounded failure probability is below 2^{-20}.

Table 1: Practical parameter choices for the first-stage amplification bound. Bold rows satisfy a 2^{-20} security threshold.

n	r	δ	$\phi_{\min}$	ϕ_{prac}	c	N_0	Union bound
2^{10}	4	1/3	0.0330	0.0471	4.28×10^{-3}	4 077	1.55×10^{-4}
	4	**1/2**	**0.1564**	**0.2017**	$\mathbf{1.41 \times 10^{-2}}$	**4 090**	$\mathbf{2.71 \times 10^{-13}}$
	6	1/3	0.0037	0.0058	4.30×10^{-4}	5 912	2.63×10^{-1}
	6	1/2	0.0416	0.0566	2.98×10^{-3}	6 124	1.05×10^{-4}
	8	1/3	0.0004	0.0007	5.00×10^{-5}	7 490	8.23×10^{-1}
	8	1/2	0.0106	0.0151	6.70×10^{-4}	8 126	6.26×10^{-2}
2^{15}	**4**	**1/3**	**0.0330**	**0.0352**	$\mathbf{5.80 \times 10^{-4}}$	**120 348**	$\mathbf{4.55 \times 10^{-18}}$
	4	**1/2**	**0.1564**	**0.1636**	$\mathbf{1.99 \times 10^{-3}}$	**128 841**	$\mathbf{6.14 \times 10^{-58}}$
	6	1/3	0.0037	0.0041	6.00×10^{-5}	128 168	5.03×10^{-4}
	6	**1/2**	**0.0416**	**0.0439**	$\mathbf{4.10 \times 10^{-4}}$	**194 853**	$\mathbf{3.54 \times 10^{-18}}$
	8	1/3	0.0004	0.0005	1.00×10^{-5}	58 819	9.73×10^{-2}
	8	1/2	0.0106	0.0113	9.00×10^{-5}	238 479	2.93×10^{-6}
2^{20}	**4**	**1/3**	**0.0330**	**0.0334**	$\mathbf{1.10 \times 10^{-4}}$	**3 063 763**	$\mathbf{2.27 \times 10^{-115}}$
	4	**1/2**	**0.1564**	**0.1577**	$\mathbf{3.50 \times 10^{-4}}$	**3 947 249**	**0** (Precision)
	6	**1/3**	**0.0037**	**0.0038**	$\mathbf{1.00 \times 10^{-5}}$	**3 248 067**	$\mathbf{1.32 \times 10^{-21}}$
	6	**1/2**	**0.0416**	**0.0420**	$\mathbf{7.00 \times 10^{-5}}$	**5 467 723**	$\mathbf{7.58 \times 10^{-108}}$
	8	**1/3**	**0.0004**	**0.0005**	$\mathbf{1.00 \times 10^{-5}}$	**58 819**	$\mathbf{2.22 \times 10^{-41}}$
	8	**1/2**	**0.0106**	**0.0108**	$\mathbf{3.00 \times 10^{-5}}$	**2 489 057**	$\mathbf{1.46 \times 10^{-73}}$

5.2 Amplifying Encoder Framework

While the above section fails to bound the failure probability for all input weights, it exhibits strong amplification. For modest parameters like $n = 2^{15}$ and $r = 4$, a *single* round of RA amplifies inputs with relative weight over 0.0352 to 0.33 with a 4.55×10^{-18} failure probability. For larger n, the failure probability drops massively. At weaker rates like $r = 8$, the $n = 2^{20}$ RA regime can amplify relative weights as low as 0.01 to 0.5 with a negligible failure probability of 1.46×10^{-73}.

Although the ϕ threshold is linear in n by construction, it represents a very low relative weight that drops as n increases. Therefore, we believe the prime RA code can be combined with an additional code to produce a viable encoding framework. The outer code simply needs to bring codewords up to the modest ϕ_{prac} threshold before feeding it to the RA code to essentially guarantee strong distance amplification.

6 Incomplete RAA Bound

Lastly, we provide progress towards a full two-stage bound on the RAA failure probability. Although this specific approach fails due to issues with the parameters, we believe it may be illustrative of the code behavior.

We wish to bound $\mathbb{P}(\exists x \in F_q^n \backslash \{0\} : L_2 < \delta N)$. To do so, we will condition on the input to the second accumulator, L_1. Because L_2 can be expressed as $\sum_{i=1}^{w_1} R_i(1 + g_i)$, we note that

$$\sum_{i=1}^{w_1}(1 + g_i) = w_1 + \sum_{i=1}^{w_1} g_i = w_1 + (N - w_1 - g_0) = N - g_0,$$

and split L_2 as follows:[3]

$$L_2 = \sum_{i=1}^{w_1} R_i(1 + g_i) = (N - g_0) - \sum_{i \geq 1 : R_i = 0} (1 + g_i).$$

Intuitively, the left expression of L_2 is the direct sum of all runs, whereas the right expression is all terms after

[3]We note that this approach has qualitative similarities to those taken by [6]. In those analyses, the accumulation step considered cases of poor nonzero "supports" and poor "values." This is akin to our decomposition, where $(N - g_0)$ represents a poor support via a long series of black draws before the first white, and poor values as the event of many cancellations when $R_i = 0$.

the first white draw $N - g_0$ minus the number of subsequent gaps of length $1 + g_i$ whenever $R_i = 0$.

Bounding $(N - g_0) - \sum_{i \geq 1 : R_i = 0}(1 + g_i) < \delta N$ is akin to bounding $g_0 + \sum_{i \geq 1 : R_i = 0}(1 + g_i) \geq (1 - \delta)N$. We will condition L_2 values by the input weight to the second accumulator L_1, and bound $\mathbb{P}_{\Pi_2, V_2, \Pi_1, V_1}(L_2 < \delta N)$ by

$$\mathbb{P}_{\Pi_2,V_2}(L_2 < \delta N | L_1 > N^\kappa, w_0 \text{ fixed})\mathbb{P}_{\Pi_1,V_1}(L_1 > N^\kappa | w_0 \text{ fixed}) +$$
$$\mathbb{P}_{\Pi_1,V_1}(L_2 < \delta N | L_1 \leq N^\kappa, w_0 \text{ fixed})\mathbb{P}_{\Pi_1,V_1}(L_1 \leq N^\kappa | w_0 \text{ fixed}).$$

For brevity, we will not write the "w_0 fixed" condition until taking union bounds over all w_0. Trivially bounding the L_1 probabilities by 1 and using the $(1 - \delta)N$ argument, the above is bounded by

$$\leq \mathbb{P}_{\Pi_2,V_2}\left(g_0 + \sum_{i \geq 1 : R_i = 0}(1 + g_i) > (1 - \delta)N | L_1 > N^\kappa\right) + \mathbb{P}_{\Pi_1,V_1}(L_2 < \delta N | L_1 \leq N^\kappa), \tag{14}$$

where we can split the left probability to get an upper bound of

$$\leq \mathbb{P}_{\Pi_2,V_2}\left(g_0 > \frac{1-\delta}{2}N | L_1 > N^\kappa\right) + \mathbb{P}_{\Pi_2,V_2}\left(\sum_{i \geq 1 : R_i = 0}(1 + g_i) > \frac{1-\delta}{2}N | L_1 > N^\kappa\right) + \mathbb{P}_{\Pi_1,V_1}(L_2 < \delta N | L_1 \leq N^\kappa) \tag{15}$$

Term 1 Bound. We will upper bound each term of (15), starting with the leftmost.

$$\begin{aligned}\mathbb{P}_{\Pi_2,V_2}\left(g_0 > \frac{1-\delta}{2}N | L_1 > N^\kappa\right) &= \sum_{\ell \geq N^\kappa} \mathbb{P}(L_1 = \ell | w_0)\mathbb{P}\left(g_0 > (1-\delta)N/2 | L_1 = \ell\right) \\ &= \sum_{\ell \geq N^\kappa} \mathbb{P}(L_1 = \ell | w_0)\frac{\binom{(1+\delta)N/2}{\ell}}{\binom{N}{\ell}}.\end{aligned}$$

$\frac{\binom{(1+\delta)N/2}{\ell}}{\binom{N}{\ell}}$ is maximized at $\ell = N^\kappa$, so we have [4]

$$\sum_{\ell \geq N^\kappa} \mathbb{P}(L_1 = \ell | w_0)\frac{\binom{(1+\delta)N/2}{\ell}}{\binom{N}{\ell}} \leq \frac{\binom{(1+\delta)N/2}{N^\kappa}}{\binom{N}{N^\kappa}} \sum_{\ell \geq N^\kappa} \mathbb{P}(L_1 = \ell | w_0) \leq \frac{\binom{(1+\delta)N/2}{N^\kappa}}{\binom{N}{N^\kappa}},$$

[4] For proof that this fraction decreases in ℓ, see the appendix.

because all $\mathbb{P}(L_1 = \ell | w_0)$ probabilities must sum to 1 for fixed w_0.[5] To bound this ratio, we write it as a product:

$$\frac{\binom{(1+\delta)N/2}{N^\kappa}}{\binom{N}{N^\kappa}} = \prod_{j=0}^{N^\kappa - 1} \frac{(1+\delta)N/2 - j}{N - j}.$$

For any $j \geq 0, a < 1$, the inequality $\frac{(1+\delta)N/2-j}{N-j} < (1+\delta)/2$ is satisfied. [6] When $j = 0$, the term $\frac{(1+\delta)N/2-j}{N-j}$ is simply $\frac{1+\delta}{2}$. We upper bound the above as

$$\frac{\binom{(1+\delta)N/2}{N^\kappa}}{\binom{N}{N^\kappa}} = \frac{1+\delta}{2} \prod_{j=1}^{N^\kappa - 1} \frac{(1+\delta)N/2 - j}{N - j} < \frac{1+\delta}{2} \left(\frac{1+\delta}{2}\right)^{N^\kappa - 1} = \left(\frac{1+\delta}{2}\right)^{N^\kappa}.$$

This equals $e^{N^\kappa \log \frac{1+\delta}{2}}$. Since $\frac{1+\delta}{2} < 1$, the exponent is negative, leaving to super-polynomial decay in n for any $\kappa > 0$. Therefore,

$$\mathbb{P}_{\Pi_2, V_2}\left(g_0 > \frac{1-\delta}{2} N | L_1 > N^\kappa\right) \leq \frac{\binom{(1+\delta)N/2}{N^\kappa}}{\binom{N}{N^\kappa}} \leq e^{N^\kappa \log \frac{1+\delta}{2}} = O(\exp(-C \cdot N^\kappa)). \tag{16}$$

Term 2 Bound. We move to bounding the second term of 15, $\mathbb{P}_{\Pi_2, V_2}\left(\sum_{i \geq 1 : R_i = 0} (1 + g_i) > \frac{1-\delta}{2} N | L_1 > N^\kappa\right)$. We will first need two lemmas on the random variables g_i and R_i.

Lemma 6. *Let $g_i \sim NHG(N - (\sum_{j<i} t_j) - i, w_1 - i, 1)$ for $i = 1, \ldots, L_1$. Then for fixed $L_1 = \ell$, each g_i is stochastically dominated by $X \sim Exp\left(\log \frac{N}{N-\ell}\right)$.*

We will write $\lambda = \log \frac{N}{N-\ell}$ for simplicity.

Proof.

$$\mathbb{P}(g_i \geq t) \leq \left(\frac{N-\ell}{N}\right)^t$$

[5] For the binomial coefficients to be well defined, we require $(1+\delta)N/2 > N^\kappa$.

[6] More specifically, this requires

$$\frac{1+\delta}{2} N - j < \frac{1+\delta}{2}(N - j) = \frac{1+\delta}{2} N - \frac{1+\delta}{2} j \implies -j < -\frac{1+\delta}{2} j \implies j\left(1 - \frac{1+\delta}{2}\right) > 0,$$

which holds for all $j \geq 1$. The $j = 0$ term is already handled explicitly.

. The CDF of X is given in closed form due to the exponential distribution, where

$$\mathbb{P}(X \geq t) := e^{-\lambda t} = e^{k \log((N-\ell)/N)} = \left(\frac{N-\ell}{N}\right)^t.$$

Therefore, $\mathbb{P}(g_i \geq t) \leq \mathbb{P}(X \geq t)$ for all t. $\square$

Lemma 7. *Let $K := |\{i \geq 1 : R_i = 0\}|$, the number of run-ending draws. Then for fixed $L_1 = \ell$, K is stochastically dominated by $B \sim Binom(\ell, \frac{1}{q-1})$.*[7]

Proof. The transitions R_i are controlled entirely by the nonzero values of the input. These values are determined by the weight-randomizing matrix V and are fully independent of the randomized support due to Π. Therefore, to describe the distribution of K, we need only consider the structure of the nonzero values of the input. Since cancellations in K must be non-consecutive due to the Markov structure and valid opportunities for cancellation have probability $1/(q-1)$, the result immediately follows. $\square$

We now proceed with the bound of $\mathbb{P}_{\Pi_2,V_2}\left(\sum_{i\geq 1:R_i=0}(1+g_i) > \frac{1-\delta}{2}N|L_1 > N^\kappa\right)$. It is known that if a random variable A is stochastically dominated by a random variable B, then for any nondecreasing function $f(x)$, $\mathbb{E}(f(A)) \leq \mathbb{E}(f(B))$. Applying this to g_i and X, $\mathbb{E}\left(e^{\theta g_i}|L_1\right) \leq \mathbb{E}\left(e^{\theta X}|L_1\right)$ for $\theta \geq 0$. Since $\mathbb{E}\left(e^{\theta X}|L_1\right)$ is precisely the MGF of $X \sim \text{Exp}(\lambda)$ for fixed L_1, we can evaluate:

$$\mathbb{E}\left(e^{\theta X}|L_1\right) = \int_0^\infty e^{\theta y}\lambda e^{-\lambda y}dy = \lambda\int_0^\infty e^{-(\lambda-\theta)y}dy,$$

which converges if $\theta < \lambda$. Then for $0 \leq \theta < \lambda$, the above equals

$$\lambda\left(0 - \frac{1}{-(\lambda-\theta)}\right) = \frac{\lambda}{\lambda-\theta} \implies \mathbb{E}\left(e^{\theta g_i}|L_1\right) \leq \frac{\lambda}{\lambda-\theta}.$$

By the negative association of the g_i, we can factor the MGFs:

$$\mathbb{E}\left(\prod_{i\geq 1:R_i=0} e^{\theta g_i}|L_1\right) \leq \prod_{i\geq 1:R_i=0}\left(\mathbb{E}\left(e^{\theta g_i}|L_1\right)\right) \leq \left(\frac{\lambda}{\lambda-\theta}\right)^{|\{i\geq 1:R_i=0\}|}.$$

[7]Intuition: because K counts transitions to 0, which cannot occur for adjacent i and $i+1$, we would expect that K is less than a random variable that always has a chance to transition to 0.

Then for a fixed $|\{i \geq 1 : R_i = 0\}| = K$,

$$\mathbb{E}\left(e^{\theta \sum_{i\geq 1:R_i=0}(1+g_i)}|K, L_1\right) \leq e^{\theta K}\left(\frac{\lambda}{\lambda-\theta}\right)^K = \left(\frac{e^\theta \lambda}{\lambda-\theta}\right)^K,$$

which we will abbreviate as $(m(\theta))^K$. Since for $0 \leq \theta < \lambda$, $m(\theta)^x$ increases in x, we can again exploit the stochastic domination of K:

$$\mathbb{E}(m(\theta)^K|K, L_1) \leq \mathbb{E}(m(\theta)^B|L_1).$$

And since we know for a random variable $\sim \text{Binom}(n, p)$ that $\mathbb{E}(x^B) := (1 - \frac{1}{p} + \frac{x}{p})^n$,

$$\mathbb{E}(m(\theta)^K|K, L_1) \leq \left(1 + \frac{m(\theta)-1}{q-1}\right)^{L_1}$$

for all K, which removes our dependence on it.

Now we want to convert from expectations to probabilities, which naturally motivates the usage of Markov's inequality. Since $f(x) = e^{\theta x}$ is increasing in x for $\theta > 0$, the following events are the same:

$$\mathbb{P}\left(\sum_{i\geq 1:R_i=0}(1+g_i) > \frac{1-\delta}{2}N|L_1\right) = \mathbb{P}\left(e^{\theta \sum_{i\geq 1:R_i=0}(1+g_i)} > e^{\theta \frac{1-\delta}{2}N}|L_1\right).$$

By Markov,

$$\begin{aligned}\mathbb{P}\left(e^{\theta \sum_{i\geq 1:R_i=0}(1+g_i)} > e^{\theta \frac{1-\delta}{2}N}|L_1\right) &\leq \mathbb{E}\left(e^{\theta \sum_{i\geq 1:R_i=0}(1+g_i)}\right)e^{-\theta\frac{1-\delta}{2}N} \\ &\leq \left(1 + \frac{m(\theta)-1}{q-1}\right)^{L_1} e^{-\theta\frac{1-\delta}{2}N}.\end{aligned}$$

Plugging in $m(\theta)$, the above is

$$\leq \inf_{\theta\in[0,\lambda)}\left(\exp\left(-\frac{1-\delta}{2}\theta N + L_1 \log\left(1 + \frac{\frac{e^\theta\lambda}{\lambda-\theta}-1}{q-1}\right)\right)\right)_{L_1}. \tag{17}$$

We will choose a universal θ^* and derive an upper bound that holds for all L_1. This amounts to upper bounding the exponent. Rewrite

$$\log\left(1+\frac{\frac{e^\theta\lambda}{\lambda-\theta}-1}{q-1}\right)=\log\left(\frac{q-2+\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}\right).$$

Since $\log(x)$ increases in x and $\log(x)\le x-1$ for $x>1$, $\frac{q-2+\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}<1+\frac{\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}$ gives

$$\log\left(\frac{q-2+\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}\right)\le\frac{\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}.$$

We wish to upper bound

$$-\frac{1-\delta}{2}\theta N+L_1\frac{\frac{e^\theta\lambda}{\lambda-\theta}}{q-1}=-\frac{1-\delta}{2}\theta N+L_1\frac{\frac{e^\theta}{1-\theta/\lambda}}{q-1}.$$

The righthand term decreases as λ increases. $\lambda=\log(N/(N-L_1))$, conditioned on $L_1\ge N^\kappa$, is smallest for $\lambda^*=\log(N/(N-N^\kappa))$. So for any L_1, plugging in λ^* for every λ term gives us a definite upper bound. Furthermore, L_1 is at most N, and we have

$$-\frac{1-\delta}{2}\theta N+L_1\frac{\frac{e^\theta}{1-\theta/\lambda}}{q-1}\le-\frac{1-\delta}{2}\theta N+\frac{N}{q-1}\cdot\frac{e^\theta}{1-\theta/\lambda^*}.$$

We now set $\theta^*=c\lambda^*$ for $c\in(0,1)$, which satisfies the required constraint $\theta^*<\lambda^*$. Since $q\ge n$ and $N=rn$, we have the exact upper bound

$$\frac{N}{q-1}\le\frac{N}{n-1}=\frac{rn}{n-1}\le 2r, \tag{18}$$

where the last inequality holds for all $n\ge 2$. Substituting $\theta^*=c\lambda^*$ and (18), the exponent of (17) is bounded above by

$$\mathcal{E}(c):=-\frac{(1-\delta)c\lambda^*}{2}N+2r\frac{e^{c\lambda^*}}{1-c}. \tag{19}$$

Lemma 8. *$\lambda^*N\ge N^\kappa$.*

Proof. Since $-\log(1-x)\ge x$ for all $x\in[0,1)$, applying this with $x=N^\kappa/N$ gives

$$\lambda^*=-\log\big(1-N^{\kappa-1}\big)\ge N^{\kappa-1},$$

and the desired statement follows. $\square$

Lemma 9. $e^{c\lambda^*} \leq \left(\frac{N}{N-N^\kappa}\right)^c \leq \frac{N}{N-N^\kappa}$ *for all* $c \in (0,1)$.

Proof. The first inequality is equality by definition of λ^*:

$$e^{c\lambda^*} = e^{c\log(N/(N-N^\kappa))} = \left(\frac{N}{N-N^\kappa}\right)^c.$$

We require $N^\kappa \leq N/2$, i.e., $N \geq 2^{1/(1-\kappa)}$, which is satisifed for all sufficiently large N. Then $\frac{N}{N-N^\kappa} > 1$, yielding the second inequality, and $c \leq 1$, so $\left(\frac{N}{N-N^\kappa}\right)^c \leq \frac{N}{N-N^\kappa}$.

$\square$

Applying Lemma 8 to the negative term and Lemma 9 to the positive term in (19):

$$\mathcal{E}(c) \leq -\frac{(1-\delta)c}{2}N^\kappa + \frac{2r}{1-c} \cdot \frac{N}{N-N^\kappa}$$

For $\mathcal{E}(c) < 0$, it suffices to require

$$\frac{(1-\delta)c}{2}N^\kappa > \frac{2r}{1-c} \cdot \frac{N}{N-N^\kappa}. \tag{20}$$

Since $N^\kappa \leq N/2$ implies $N/(N-N^\kappa) \leq 2$, the above requirement is implied by the cleaner sufficient condition

$$N^\kappa > \frac{8r}{(1-\delta)c(1-c)}. \tag{21}$$

The right-hand side is minimized over $c \in (0,1)$ at $c^* = 1/2$, where $c^*(1-c^*) = 1/4$, giving the explicit finite-N condition

$$N^\kappa > \frac{32r}{1-\delta}. \tag{22}$$

For all N satisfying (22), setting $c = 1/2$ and therefore $\theta^* = \lambda^*/2$, the exponent satisfies $\mathcal{E}(1/2) < 0$, and since this bound holds uniformly for all $L_1 \geq N^\kappa$, we conclude

$$\mathbb{P}_{\Pi_2,V_2}\left(\sum_{i\geq 1:R_i=0}(1+g_i) > \frac{1-\delta}{2}N \,\Big|\, L_1 > N^\kappa\right) \leq \sum_{\ell > N^\kappa} \mathbb{P}(L_1=\ell \mid w_0)\exp(\mathcal{E}(1/2))$$
$$\leq \exp(\mathcal{E}(1/2))$$
$$= \exp\left(-\frac{(1-\delta)\lambda^*}{4}N + 4r\cdot\frac{N}{N-N^\kappa}\right) = O\left(e^{-N^\kappa}\right). \quad (23)$$

Term 3 Bound. Lastly, it remains to bound the final term of (15): $\mathbb{P}_{\Pi_1,V_1}(L_2 < \delta N | L_1 \leq N^\kappa)$.

We use an existing bound on the probability of the first round failure from [2] for the binary RAA code:[8]

Proposition 3 (Blaze: First Stage Failure)**.** *Let $r, N \in \mathbb{N}$, r even, with $r \geq 4$, $\delta \in (0, 1/3)$, $\kappa = 1 - \frac{2}{r}(1+\varepsilon)$ for some $\varepsilon > 0$. If $r = 4$ assume further $\delta < 1/4$. Suppose there exists $\xi \in (0,1)$ such that the following holds*

$$N \geq max\left\{7,\, 2m+r+1,\, \frac{2}{1-\xi}m,\, \frac{(2/\xi)^{r/\varepsilon}}{r^{1/\varepsilon}}\right\}. \quad (24)$$

Then:

$$\mathbb{P}(\exists x : L_2 < \delta N) \leq O\left(N^{1+\kappa-r/2}\right).$$

This gives an upper bound on the expected number of low-weight messages, conditioned on the first stage of accumulation leading to low intermediate weight $\lfloor N^\kappa \rfloor$. By Markov's inequality, this implies an upper bound on the probability over all messages, since

$$\mathbb{P}(\exists x : L_2 < \delta N) \leq \mathbb{E}(\#x : L_2 < \delta N).$$

From Theorem 1, we know this is also an upper bound for term 3 in the prime case.[9]

[8] We abbreviate the explicit bound and additional conditions.
[9] In fact, it is quite weak, and could be significantly improved.

Combining All Terms. From (15), we want to bound the three part probability

$$\mathbb{P}_{\Pi_2,V_2}\left(g_0 > \frac{1-\delta}{2}N | L_1 > N^\kappa\right) + \mathbb{P}_{\Pi_2,V_2}\left(\sum_{i\geq 1: R_i=0}(1+g_i) > \frac{1-\delta}{2}N | L_1 > N^\kappa\right) + \mathbb{P}_{\Pi_1,V_1}(L_2 < \delta N | L_1 \leq N^\kappa)$$

over all input messages of length n. However, the union bound over the first two terms leads to conflicts in the parameters that prevent true overall decay. We have been unable to resolve them within the scope of this thesis, so we leave this incomplete proof for instructive purposes.

7 Conclusions

Overall, our approach is rather unconventional, as tight bounds are often obtained via directly analyzing the RAA IOWEs. This leads to issues like in the two-stage RAA bound, which fails due to explicit conditioning on the intermediate weight. However, we believe the probabilistic perspective in this work provides a useful lens with which to intuitively understand the accumulation of the code. It also leads to bounds which are often simple and, in contrast to other approaches, do not require memory-intensive scripts to see explicitly. Our case-splitting allows us to conduct detailed analyses into specific behaviors of the code, including bounds on the expected numbers of cancellations over the field[10], which have been questioned in previous work. Furthermore, our reduction-based theorem formally confirms the widely-conjectured intuition that, under the same conditions, the prime code ought to be bounded by a binary worst case cancellation behavior. Finally, our single-stage amplification analysis provides a useful tool for efficient encoding, provided an outer code which need only amplify weights to a very low threshold.

[10] See Appendix.

8 Acknowledgments

I would like to thank Professors Ben Fisch and Dan Spielman for advising this work and serving on my thesis committee. I am also grateful to Hadas Zeilberger for her generous supervision.

Many thanks to the Yale Department of Mathematics and the Department of Computer Science.

Lastly, thanks to my family and friends, without whom I would not be here.

Any errors in this work are my own.

References

[1] D. Divsalar, H. Jin, and R. J. McEliece. "Coding theorems for 'turbo-like' codes". In: *Proc. 36th Annual Allerton Conference on Communication, Control, and Computing*. Monticello, IL, USA, Sept. 1998, pp. 201–210.

[2] Paul Brehm, Dmitry Moshkovich, and Ron Rothblum. "Blaze: Fast SNARKs from Interleaved RAA Codes". In: *Advances in Cryptology - EUROCRYPT 2025*. Lecture Notes in Computer Science. to appear. Springer, 2025. URL: `https://eprint.iacr.org/2024/1609`.

[3] Louay Bazzi, Mohammad Mahdian, and Daniel A Spielman. "The minimum distance of turbo-like codes". In: *IEEE Transactions on Information Theory* 55.1 (2009), pp. 6–15. DOI: `10.1109/TIT.2008.2008114`.

[4] Jörg Kliewer, Kamil S. Zigangirov, and Daniel J. Costello Jr. "New results on the minimum distance of repeat multiple accumulate codes". In: *Proc. 45th Annual Allerton Conf. Commun., Control, Computing*. Monticello, IL, Sept. 2007, pp. 246–253.

[5] E. Boyle et al. "Correlated pseudorandomness from expand-accumulate codes". In: *Advances in Cryptology – CRYPTO 2022*. Vol. 13508. Lecture Notes in Computer Science. Cham: Springer, 2022, pp. 603–633. DOI: `10.1007/978-3-031-15979-4_21`.

[6] A. R. Block et al. "Field-agnostic SNARKs from expand-accumulate codes". In: *Advances in Cryptology – CRYPTO 2024*. Vol. 14929. Lecture Notes in Computer Science. Cham: Springer, 2024, pp. 276–307. DOI: `10.1007/978-3-031-68403-6_9`.

[7] 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. `https://eprint.iacr.org/2026/524`. 2026. URL: `https://eprint.iacr.org/2026/524`.

9 Appendix

Binomial Ratio Decreases with ℓ

Lemma 10. *For $a \in (0,1)$ and $0 \le \ell \le aN$, the ratio $\frac{\binom{aN}{\ell}}{\binom{N}{\ell}}$ is strictly decreasing in ℓ.*

Proof. Write the ratio as a product:

$$\frac{\binom{aN}{\ell}}{\binom{N}{\ell}} = \frac{(aN)!/(aN-\ell)!}{(N)!/(N-\ell)!} = \prod_{j=0}^{\ell-1} \frac{aN-j}{N-j}.$$

It suffices to show that the multiplicative factor by which the ratio changes when passing from ℓ to $\ell+1$ is strictly less than 1. This factor is

$$\frac{\binom{aN}{\ell+1}/\binom{N}{\ell+1}}{\binom{aN}{\ell}/\binom{N}{\ell}} = \frac{aN-\ell}{N-\ell}.$$

Since $a < 1$, we have $aN - \ell < N - \ell$ for all $\ell \ge 0$, and since $\ell \le aN < N$ the denominator is strictly positive. Therefore

$$\frac{aN-\ell}{N-\ell} < 1$$

for all $0 \le \ell \le aN - 1$, which establishes that the ratio is strictly decreasing in ℓ. □

Expected cancellations from spectral analysis. We also include an interesting technique to understand the expected number of cancellations in the prime RAA code. The simple transition matrix of the white draws allows us to analyze the chain's expected returns to zero (cancellation) in the following way:

Lemma 11 (Expected run cancellations)**.** *Let $w \ge 1$ be the number of white balls drawn, and recall that $R_1 = 1$ deterministically and $(R_i)_{i=2}^{w}$ is the run-status Markov chain with transitions*

$$\mathbb{P}(R_i = 1 \mid R_{i-1} = 1) = \tfrac{q-2}{q-1}, \qquad \mathbb{P}(R_i = 0 \mid R_{i-1} = 1) = \tfrac{1}{q-1},$$
$$\mathbb{P}(R_i = 1 \mid R_{i-1} = 0) = 1, \qquad \mathbb{P}(R_i = 0 \mid R_{i-1} = 0) = 0.$$

Let $K := |\{i \ge 1 : R_i = 0\}|$ be the number of cancellation steps. Then

$$\mathbb{E}[K] \;\le\; \frac{w}{q}. \tag{25}$$

In particular, since $w \leq N$ and $q \geq N$, we have $\mathbb{E}[K] \leq 1$.

Proof. Since $R_1 = 1$ always, state 0 can only be visited for $i \geq 2$, so

$$\mathbb{E}[K] = \sum_{i=2}^{w} \mathbb{P}(R_i = 0). \tag{26}$$

We evaluate $\mathbb{P}(R_i = 0)$ via the spectral decomposition of the transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ \dfrac{1}{q-1} & \dfrac{q-2}{q-1} \end{pmatrix}, \tag{27}$$

where rows and columns are indexed by states 0 and 1 respectively.

Eigenvalues and stationary distribution of P. The characteristic polynomial of P is

$$\lambda^2 - \frac{q-2}{q-1}\lambda - \frac{1}{q-1} = 0,$$

which after multiplying by $(q-1)$ becomes $(q-1)\lambda^2 - (q-2)\lambda - 1 = 0$. The discriminant equals $(q-2)^2 + 4(q-1) = q^2$, so the two eigenvalues are

$$\lambda_1 = \frac{(q-2)+q}{2(q-1)} = 1 \qquad \text{and} \qquad \lambda_2 = \frac{(q-2)-q}{2(q-1)} = \frac{-1}{q-1}.$$

The spectral gap is $1 - |\lambda_2| = \frac{q-2}{q-1}$, which is close to 1 for large q, reflecting rapid mixing.

The stationary distribution π satisfies $\pi P = \pi$ and $\pi_0 + \pi_1 = 1$. From the equation $\pi_0 \cdot 0 + \pi_1 \cdot \frac{1}{q-1} = \pi_0$ we get $\pi_1 = (q-1)\pi_0$, and normalisation gives

$$\pi_0 = \frac{1}{q}, \qquad \pi_1 = \frac{q-1}{q}.$$

Spectral decomposition. Since P has two distinct eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1/(q-1)$, we can write $P^k = c_1\Pi_1 + c_2(k)\Pi_2$ where Π_1 is the projection onto the stationary distribution and Π_2 captures the transient part. Explicitly, for the $(1 \to 0)$ entry (starting in state 1, ending in state 0):

$$P^k(1 \to 0) = \pi_0 + A \cdot \lambda_2^k$$

for some constant A to be determined. At $k = 0$, $P^0(1 \to 0) = 0$ (since the chain starts in state 1), giving $0 = \pi_0 + A$, hence $A = -\pi_0 = -1/q$. Therefore

$$\mathbb{P}(R_i = 0 \mid R_1 = 1) = P^{i-1}(1 \to 0) = \frac{1}{q}\left(1 - \left(\frac{-1}{q-1}\right)^{i-1}\right). \tag{28}$$

Bounding the sum. Since $|\lambda_2| = 1/(q-1) < 1$, we have $|(-1/(q-1))^{i-1}| \le 1/(q-1)^{i-1} \le 1$ for all $i \ge 2$. Thus $P^{i-1}(1 \to 0) \le \frac{1}{q}(1 + \frac{1}{(q-1)^{i-1}}) \le \frac{2}{q}$, demonstrating that the formula is a genuine probability.

To bound the sum (26), substitute (28):

$$\mathbb{E}[K] = \sum_{i=2}^{w} \frac{1}{q}\left(1 - \left(\tfrac{-1}{q-1}\right)^{i-1}\right) = \frac{w-1}{q} - \frac{1}{q}\sum_{k=1}^{w-1}\left(\frac{-1}{q-1}\right)^k.$$

The remaining sum is a finite geometric series with first term $-1/(q-1)$ and ratio $-1/(q-1)$. Its value is

$$\sum_{k=1}^{w-1}\left(\frac{-1}{q-1}\right)^k = \frac{-\frac{1}{q-1}\left(1 - \left(\frac{-1}{q-1}\right)^{w-1}\right)}{1 + \frac{1}{q-1}} = \frac{-1}{q}\left(1 - \left(\frac{-1}{q-1}\right)^{w-1}\right).$$

Since $|(-1/(q-1))^{w-1}| \le 1$, the factor $1 - (-1/(q-1))^{w-1}$ lies in $[0, 2]$, so this geometric series lies in $[-2/q,\, 0]$. It is non-positive, and therefore

$$\mathbb{E}[K] = \frac{w-1}{q} - \frac{1}{q} \cdot \left(\text{a value in } \left[-\frac{2}{q},\, 0\right]\right) \le \frac{w-1}{q} + \frac{2}{q^2} \le \frac{w-1}{q} + \frac{1}{q} = \frac{w}{q},$$

where the penultimate inequality uses $2/q^2 \le 1/q$ for $q \ge 2$. This shows (25).

Since $w \le N$ and $q \ge N$ by assumption, we have $\mathbb{E}[K] \le N/q \le 1$. That is, on average fewer than one cancellation occurs per accumulation pass, regardless of the input weight. $\square$

Interestingly, the bound $\mathbb{E}[K] \le w/q$ is tight to leading order: for large q, equation (28) gives $P^k(1 \to 0) \approx 1/q$

for each $k \geq 1$, so $\mathbb{E}[K] \approx (w-1)/q$. The original stochastic domination lemma (Lemma 7) bounds K by $\mathrm{Binom}(w, 1/(q-1))$, whose mean is $w/(q-1) > w/q$; the present spectral argument gives the tighter mean bound w/q.