Distance of RAA Codes over Large Finite Fields (with Applications in zkSNARKs and PCGs)

Pariya Akhiani and Yupeng Zhang

University of Illinois Urbana Champaign
{akhiani2,zhangyp}@illinois.edu

Abstract. Error-correcting codes play a central role in modern cryptography, enabling efficient constructions of primitives such as zero knowledge proofs and secure multiparty computations. Among them, repeat-accumulate-accumulate (RAA) codes have recently attracted significant attention due to their linear-time encoding and good distance properties. However, prior work only established provable distance guarantees over the binary field or over finite fields whose size is smaller than the message length. The underlying techniques do not extend to larger fields. This restriction is significant, as many cryptographic constructions based on error-correcting codes operate over large finite fields.

In this paper, we prove that RAA codes achieve constant relative distance $0 < \delta \leq \frac{1}{2}$ with high probability over large finite fields. Moreover, we resolve an open conjecture by showing for the first time that the distance of RAA codes improves over large fields. We provide both theoretical analysis and empirical evidence demonstrating that, compared to the binary case, large fields yield strictly better distance guarantees with much smaller failure probabilities.

Our empirical evaluation shows that for message length $n = 2^{20}$ and repetition factor $r = 4$, the RAA code over a 31-bit prime field achieves relative distance $1/2$, except with failure probability 2^{-20}. This is better than the best provable distance of $\delta = 0.2$ with failure probability 2^{-10} over the binary field. Leveraging our improved distance bounds, we obtain a 2.6× reduction in the proof size of a polynomial commitment scheme based on RAA codes compared to using the binary-field bound, and a 5.4× reduction compared to a prior construction based on expand-accumulate codes.

1 Introduction

Error-correcting codes play an important role in many constructions of cryptographic primitives. For example, they are used to build pseudorandom correlation generators (PCGs), which are powerful tools to generating correlated randomness in oblivious transfers (OT) and vector oblivious linear evaluations (VOLE) with sublinear communication cost. There is also an important category of zero-knowledge succinct non-interactive arguments of knowledge (zkSNARKs) based on error-correcting codes, with advantages of efficient prover time, transparent setup and plausibly post-quantum security. In these schemes, efficiency is

largely determined by the encoding time of the underlying error-correcting code, whereas security relies on the distance of the code. Consequently, developing error-correcting codes with efficient encoding and good distance has become a central research focus in these areas.

The *Repeat–Accumulate–Accumulate* (RAA) code has recently regained attention due to its linear-time encoding. In [BCF+25], RAA code is proven to achieve good distance both asymptotically and concretely, and is used to construct an efficient zkSNARK scheme called Blaze. However, the proof is only for RAA codes over the binary field, and does not generalize to larger finite fields, which are of great interest in cryptographic applications. As a consequence, many constructions utilizing RAA codes over finite fields such as [BFRW25, BMMS25] only have heuristic security[1]. In fact, Blaze [BCF+25] has to be constructed over a binary extension field due to this limitation, unlike most zkSNARKs based on error-correcting codes that are over large finite fields.

It is widely conjectured that the distance of this family of codes improves over larger fields [BCG+22]. However, no work in the literature has formally proven this conjecture. This is because analyzing the distance over a large finite field is technically challenging, as techniques such as Markov's inequality and the naive union bound, which are effective in the binary setting, no longer work. Almost all existing papers in the literature on this family of error-correcting codes based on the accumulator matrix focus on the binary field. The only exception is [BFK+24], where the authors provided a formal proof that the expand-accumulate (EA) code has good distance over any field with high probability. Unfortunately, the provable distance bound is even weaker than the binary case due to limitations of the proof techniques, contradicting the conjecture that distance should improve as the field size increases.

Our Contributions. In this paper, we solve this open problem by formally proving that the RAA code over a large finite field achieves good distance with high probability. More importantly, we show – both theoretically and through concrete parameter instantiations – that our provable distance is strictly better than that of the binary-field counterpart, a result that has not been achieved in any prior work. Specifically, our contributions are summarized as follows:

– We show that an RAA code with message length n, repetition factor r and codeword length $N = rn$ over a finite field of size q has constant relative distance $\delta \leq \frac{1}{2}$, except with a small failure probability that is inverse polynomial in N. The dominating terms of the failure probability are $\tilde{O}(N^{2-r}) + O(\frac{1}{q-1})$, where $\tilde{O}(\cdot)$ hides polylogarithmic factors. This failure probability is strictly smaller than that in the binary case, which is $O(N^{1+\gamma-\frac{r}{2}})$ for a constant $0 < \gamma < 1$, as shown in [BCF+25]. Moreover, our provable distance bound is also strictly better than the binary-field bound, which only guarantees $\delta < \frac{1}{3}$. Our key technical contribution is a fine-grained analysis of the accumulator

[1] The focus of these papers is on the Interactive Oracle Proof of Proximity (IOPP), which works with any error-correcting code, but RAA code is used to achieve linear prover time.

matrix over large fields. This technique can be applied to strengthen the distance analysis of other codes in this family over finite fields, such as the EA code.

- As a by-product, our proof strategy can also be used to improve the analysis of the RAA code over the binary field. In particular, we reduce the dominating term of the failure probability to $\tilde{O}(N^{1-\frac{r}{2}})$, which may be of independent interest. We note that this improved failure probability remains worse than the failure probability achieved over large finite fields.
- Finally, we implement Sage scripts to compute the provable distance and corresponding failure probability under various parameter settings. Our evaluation shows that for $n = 2^{20}$, $r = 4$, the RAA code over a 31-bit prime field achieves a relative distance of $\delta = \frac{1}{2}$, except with failure probability 2^{-20}. This significantly improves upon the binary setting, which guarantees only $\delta = 0.2$ with failure probability 2^{-10}.

  To demonstrate the practical implications of our improved bounds, we evaluate the proof size of a polynomial commitment scheme (PCS) constructed from error-correcting codes, a fundamental building block of many zkSNARK systems. Leveraging our tighter distance bound, the resulting proof size for a polynomial of size $m = 2^{20}$ is 0.48 MB. This is 2.6× smaller compared to using the provable distance of the RAA code in [BCF+25] over a binary extension field, and is 5.4× smaller compared to the PCS in [BFK+24], which is based on the EA code over a finite field.

Applications in Cryptography. Our results have important applications in cryptography. As mentioned above, zkSNARKs based on error-correcting codes have drawn great attention recently. The constant relative distance is critical to prove their security through the proximity gap, while the encoding time dominates the prover efficiency. Most schemes are over large finite fields, due to the use of randomness and the Schwartz-Zippel Lemma [Sch80, Zip79] in the constructions. The only other candidate error-correcting code with linear-time encoding and constant relative distance is the (generalized) Spielman code [Spi95], but its concrete distance is very poor in practice ($\delta = 0.07$ for $n = 2^{30}$ [GLS+23]), leading to large proof size of tens of megabytes. Our result provides another option to build zkSNARKs with linear prover time over any finite field using the RAA code, while both the prover time and the proof size are better than those using the Spielman code.

Another important application is constructing PCGs based on error-correcting codes with sublinear communication. The large distance of the code is crucial to establish their security based on variants of Learning Parity with Noise (LPN) assumptions, where almost all known attacks fit in the framework of finding low-weight codewords [BCG+22]. Though the RAA code has not been directly used in PCG constructions, other codes in the same family such as the EA code [BCG+22] and the Expand-Convolute code [RRT23] have been used, and the RAA code can serve as a drop-in replacement to improve the computation time to linear over large finite fields. Overall, our results will enable new constructions with better performance in both areas.

	Field	Asymptotic Failure Probability	δ	Concrete Failure Probability
[BCF+25]	$\mathbb{F}_2$	$O(N^{1+\gamma-\frac{r}{2}})$	0.2	2^{-7}
[AHKM25]	$\mathbb{F}_q$	-	0.05	2^{-21}
Our result	$\mathbb{F}_q$	$\tilde{O}(N^{2-r})$	0.5	2^{-11}

Table 1: Comparison with existing works on the distance of RAA code. "-" means not explicitly derived in the paper. The concrete δ and failure probability are for $n = 2^{15}, r = 4, q > 2^{31}$.

Moreover, cryptographic applications operate in parameter regimes that differ substantially from those traditionally studied in coding theory. In particular, for security reasons, the finite field size q is typically asymptotically larger than any polynomial in n, a regime that has not been carefully analyzed in the coding theory literature. For example, in [RA11], the authors claim that the RAA code is asymptotically good over finite fields. However, their proof assumes that $n >> q$ as $n \to \infty$, and thus the result follows in a straightforward manner from the analysis without the size of the field. Our work is the first to show that the RAA code achieves good distance over the large finite fields required by cryptographic applications.

1.1 Related Works

Binary field. The family of repeat–multiple–accumulate (RMA) codes was proposed and analyzed in several early works [DP97, DJM98, PS03, BMS08, KZC07, KZKC08, FR08]. It was shown that RA codes with a single accumulator matrix are not asymptotically good, whereas codes with more than one accumulator matrices, such as RAA codes, achieve constant relative distance with high probability as $n \to \infty$. These works primarily focus on asymptotic behavior, and the concrete parameters obtained in the proofs are not practical. More recently, RAA codes were carefully analyzed in [BCF+25] to obtain improved concrete parameters and were used to construct a SNARK called Blaze. All of these results are over the binary field, and the techniques do not generalize easily to larger finite fields. Moreover, even over the binary field, the parameters remain suboptimal. In particular, the failure probability is $O(N^{1+\gamma-\frac{r}{2}})$, which exceeds $O(N^{-1})$ when $r = 4$. As a result, the code is only effective for large n, and additional techniques – such as puncturing and explicitly checking all messages of weight 1 and 2 – must be introduced to boost the success probability.

A variant called the EA code was introduced in [BCG+22], which replaces the first repeat-accumulate stage of the RAA code with an expander matrix. The work proves that EA codes achieve good distance over the binary field and uses them to construct a PCG. However, the encoding time of the EA code increases to $O(n \log n)$.

Finite fields. The good-distance property of RMA codes was extended to larger finite fields in [Yan04, KCL11, RA11, AR11] under the name of weighted non-binary RMA codes. However, some parts of their proofs assume that q is fixed and that n can grow much larger than q as $n \to \infty$, which does not match the parameter regime relevant to cryptographic applications. In [BFK+24], the authors provided the first formal analysis of the EA code over large finite fields with $q \gg n$ and proved that the EA code achieves good distance. This technique was later generalized to RAA codes over large finite fields in [AHKM25], where the authors presented numerical estimates of the failure probability but did not derive asymptotic bounds. Across all of these works, the provable failure probabilities are strictly worse than those obtained over the binary field, which contradicts the natural conjecture that the distance should improve as the field size increases.

A comparison with related works with concrete numbers is in Table 1. For [BCF+25], we wrote a script to compute the upper bound of the failure probability using the equations in the paper, without additional testing and puncturing. For [AHKM25], we use the numbers reported in the paper. As shown in the table, our analysis results in much better distance with smaller failure probability. See Section 5 for more experimental results.

Concurrent works. In a concurrent work [HGQZ26], Huang et al. also proved that RAA code has good distance over a finite field. However, their proof assumes that q is a constant while $n \to \infty$. Specifically, they are bounding Equation (10) with the q-terms from the naive union bound, and the repetition factor r grows with $\log q$ in the proof (see [HGQZ26, Lemma 3]). Therefore, the code does not have constant rate if q grows with n. Our proof does not rely on such an assumption. Concretely, our bound on the failure probability is always either better than or the same as their bound, as we are taking the minimum of Equation (9) and Equation (10).

Subsequent works. In a subsequent work [Kha26], Majid Khabbazian further addressed the $O(\frac{1}{q-1})$ term in our bounds. They showed that if $q - 1 \geq (eN)^2$ and $r \geq 9$, the upper bound of the failure probability of $\delta \leq \frac{1}{2}$ is dominated by $\tilde{O}(N^{1-r})$, without the q terms. In [BFMM26], Baweja et al. proposed Encode-Repeat-Accumulate (ERA) codes, where the distance can go beyond $\frac{1}{2}$ for large r. However, the construction requires another code with constant relative distance (which can be concretely poor) before the repetition and accumulations.

1.2 Technical Overview

The encoding of an RAA code is defined by a repetition code replicating each element of the message r times, followed by two accumulators computing the prefix-sum of their input vectors. The input vector is permuted before each accumulator, and the sampling of the permutation matrices defines a specific RAA code and is the random space of the probability whether the distance of the code is good or not. When defined over a large finite field, each permutation matrix $\mathbf{P}$ can be viewed as the product of a binary matrix $\mathbf{\Pi}$ permuting the positions

of elements and a diagonal matrix $\mathbf{V}$ randomizing the values of elements. See Section 2.2 for the formal definition of the RAA code.

Fine-grained analysis of IOWE. Our starting point is the input-output-weight-enumerator (IOWE) of the accumulator matrix, which counts the number of distinct input vectors of weight w that lead to output vectors of weight h. As shown in [AR11, BFK+24], the formula over a finite field of size q is

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

where N is the size of the vector as well as the length of the final codeword. The IOWE has been used to analyze the distance of the RAA code in many prior works, as the probability of an input vector of weight w outputting a vector of h after accumulation can be computed as $\frac{A_{w,h}^{N,q}}{\binom{N}{w}(q-1)^w}$, over the random choice of the permutation matrix before accumulation. With this probability, one could upper-bound the probability of the existence of a "bad" codeword of weight $w_3 < \delta N$ by the union bound summing over all weights of nonzero messages $w_0 \in [1, n]$, intermediate weights $w_2 \in [1, N]$ and output weights $w_3 \in [0, \delta N)$ ($w_1 = rw_0$ is determined by w_0). Unfortunately, applying the union bound naively does not work in the finite field case. This is because the total number of possible messages of weight w_0 is $\binom{n}{w_0}(q-1)^{w_0}$. When multiplied by the probability above derived from the IOWE, the upper bound becomes exponential in $(q-1)$ for some w_0, w_1, w_2, w_3 in their ranges, and is thus meaningless. For example, when $i = w - 1$ in the IOWE, the exponent of the q terms in $\binom{n}{w_0}(q-1)^{w_0} \cdot \frac{A_{w,h}^{N,q}}{\binom{N}{w}(q-1)^w}$ is $\approx (q-1)^{w_0 - 1}$, exponentially large for $w_0 > 1$.

In this paper, we propose a fine-grained analysis showing that for all messages of weight w_0, the probability that there exists at least one with intermediate weight equal to w_2 after the first accumulator matrix is upper-bounded by

$$\binom{n}{w_0} \sum_{i=w_1-2w_0-1}^{w_1-1} \frac{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}} + O\left(\frac{1}{q-1}\right).$$

Similarly, the probability that there exists at least one output codeword with weight w_3 after two accumulators is upper-bounded by

$$\binom{n}{w_0} \sum_{i+j > w_2 + (r-2)w_0 - 4} \frac{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}} \cdot \frac{\binom{w_3-1}{\lceil \frac{w_2-j}{2} \rceil - 1} \binom{N-w_3}{\lfloor \frac{w_2-j}{2} \rfloor} \binom{w_3 - \lceil \frac{w_2-i}{2} \rceil}{j}}{\binom{N}{w_2}} + O\left(\frac{1}{q-1}\right)$$

The analysis consists of union bounds in two steps, one on the random permutation of positions defined by $\mathbf{\Pi}$ and the other on the random values defined by $\mathbf{V}$. See Section 3.1 and Section 3.2 for the formal statements and proofs.

Comparing to the approaches in prior works, the upper bounds in [AR11, BFK+24, AHKM25] essentially drop the terms related to $q-1$, resulting in the same summand with the combinatorial terms above but with i from 0 to w_1-1 and j from 0 to w_2-1. Because of this, the bound is strictly larger than that of the binary case, as the latter probability consists of a single term of the summand with $i=j=0$. It makes the provable bound worse than that of the binary case, as we are bounding the failure probability that there exists a "bad" codeword. We show that our fine-grained analysis is critical to address this issue, making the bound better than that of the binary case.

Amplification of stage 1. We then divide the analysis into two stages, one for the repetition and the first accumulator (i.e., the RA code), and the other for the second accumulator. We show that the first stage of the RA code can amplify the weight of the message with high probability. That is, for all w_0 up to $O(\frac{n}{\log^2 n})$, $w_2 = O(w_0 \log N)$ after the first accumulator, except for a small probability that is $\tilde{O}(N^{2-r})+O(\frac{1}{q-1})$, where r is the repetition factor and is thus the inverse of the rate of the code. The failure probability is significantly smaller than that of the binary case, which is $O(N^{1+\gamma-\frac{r}{2}})$ for a constant $0<\gamma<1$ as shown in [BCF+25]. Note that although the analysis in [BCF+25] also consists of two stages, our technique is completely different as we deal with the case of large finite fields as explained above. See Section 3.1 for the details.

Good distance with amplification in stage 1. After that, we further show that with the amplification of the first stage, the weight of the codeword w_3 after the second accumulator must be $\geq \delta N$, except for a small probability that is $O(N^{-\eta_1})+O(\frac{1}{q-1})$. This is inverse polynomial in N and can be made arbitrarily small depending on the constant of the amplification of w_2. See Section 3.3 for the formal statement and the proof. Moreover, we show in Appendix A.2 that our δ can be much larger than the provable distance of the binary case using this approach, while the failure probability is significantly smaller than that of the binary case.

High-weight messages. Combining the two stages, so far we can show that the codewords of all messages of weight up to $O(\frac{n}{\log^2 n})$ are "good" with high probability, which we call low-weight messages. We note that the low-weight case must exist in our proof strategy, as r is a constant and $N=rn$, thus it is impossible to amplify the weight for messages with $w_0=\omega(\frac{n}{\log n})$ in stage 1. We are only able to prove the amplification for $w_0=O(\frac{n}{\log^2 n})$. Nevertheless, we complete the proof of high-weight messages following a similar approach in [RA11], which builds on top of the proofs in [AR09] and [KZKC08]. Our fine-grained analysis is crucial in this case as well to show that for concretely good parameters, such as $r=4$ and $\delta=1/2$, the failure probability that there exists a low-weight codeword for any high-weight message is negligible in N. See Section 4 for the details.

2 Preliminaries

We use $\mathbb{F}_q$ to denote a finite field of size q. λ denotes the security parameter, and $\mathsf{negl}(\lambda)$ denotes a negligible function in λ. For positive integer n, we define $[n] := \{0, \ldots, n-1\}$. We use boldface lowercase letters for vectors and boldface uppercase letters for matrices. Vectors are row vectors by default. For a vector $\mathbf{v}$, $\mathsf{wt}(\mathbf{v})$ denotes the Hamming weight (i.e., number of nonzero entries) of $\mathbf{v}$. For two vectors $\mathbf{u}, \mathbf{v}$ of the same dimension, $\Delta(\mathbf{u}, \mathbf{v}) = \mathsf{wt}(\mathbf{u} - \mathbf{v})$ denotes their Hamming distance. Let the support of a vector $\mathbf{x}$ of size N be the set of indices of nonzero entries: $\mathrm{supp}(\mathbf{x}) := \{\iota \in [N] : \mathbf{x}_\iota \neq 0\}$. Additionally, it will be helpful to define $\mathrm{supp}(N, w) := \{\mathrm{supp}(\mathbf{x}) | \mathbf{x} \in \mathbb{F}_q^N, \mathsf{wt}(\mathbf{x}) = w\}$ as the set of the supports of all vectors of length N and weight w. The size of $\mathrm{supp}(N, w)$ is $\binom{N}{w}$.

2.1 Error-Correcting Codes

A linear error-correcting code of codeword length N and message length $n \leq N$ over a field $\mathbb{F}_q$ is an n-dimensional linear subspace $C \subseteq \mathbb{F}_q^N$. Such a code is defined by a rank-n generator matrix $\mathbf{G} \in \mathbb{F}_q^{n \times N}$; the encoding of a message $\mathbf{x} \in \mathbb{F}_q^n$ is given by $\mathbf{x} \cdot \mathbf{G} \in \mathbb{F}_q^N$. The *rate* of the code is $R = n/N$. The *distance* of the code C is the minimum Hamming distance between any two codewords in C. When C is a linear code, the minimum distance is equal to the minimum weight of any nonzero codeword $\min_{\mathbf{u} \in C, \mathbf{u} \neq \mathbf{0}} \mathsf{wt}(\mathbf{u})$. We let $\delta := \min_{\mathbf{u} \in C, \mathbf{u} \neq \mathbf{0}} \mathsf{wt}(\mathbf{u})/N$ denote the relative distance of C. We say that a code (or formally a family of codes $\{C_n\}$ for different n) is asymptotically good if R and δ remain constants as $n \to \infty$.

2.2 Repeat-Accumulate-Accumulate Code

The encoding matrix of the RAA code of message length n and codeword length N consists of three types of matrices defined over $\mathbb{F}_q$:

- Repetition matrix $\mathbf{R} \in \mathbb{F}_q^{n \times N}$: the matrix repeats each entry of the message vector $r = \frac{N}{n}$ times. The rate of the RAA code is thus $R = \frac{1}{r}$.
- Permutation matrix $\mathbf{P} \in \mathbb{F}_q^{N \times N}$: the matrix randomly permutes the entries of the input vector. When defined over a large finite field, the matrix also multiplies each entry with a random nonzero value over the finite field. In our analysis, it is easier to view this permutation as the product of two $N \times N$ matrices $\mathbf{P} = \mathbf{\Pi} \cdot \mathbf{V}$, where $\mathbf{\Pi}$ is a permutation matrix with binary values for permutation $\pi : [N] \to [N]$ such that $\mathbf{\Pi}[i, j] = 1$ if and only if $\pi(i) = j$; $\mathbf{V}$ is a randomization matrix with nonzero values over the diagonal sampled uniformly from $\mathbb{F}_q$. We sometimes use $\mathbf{P}$ for short in our analysis.
- Accumulator matrix $\mathbf{A} \in \mathbb{F}_q^{N \times N}$: the matrix computes the prefix-sum of the input vector, hence the name "accumulator". $\mathbf{A}[i, j] = 1$ if and only if $i \leq j$.

With these matrices, the generator matrix of the RAA code is defined as

$$\mathbf{G} = \mathbf{R} \cdot \mathbf{\Pi}_1 \cdot \mathbf{V}_1 \cdot \mathbf{A} \cdot \mathbf{\Pi}_2 \cdot \mathbf{V}_2 \cdot \mathbf{A},$$

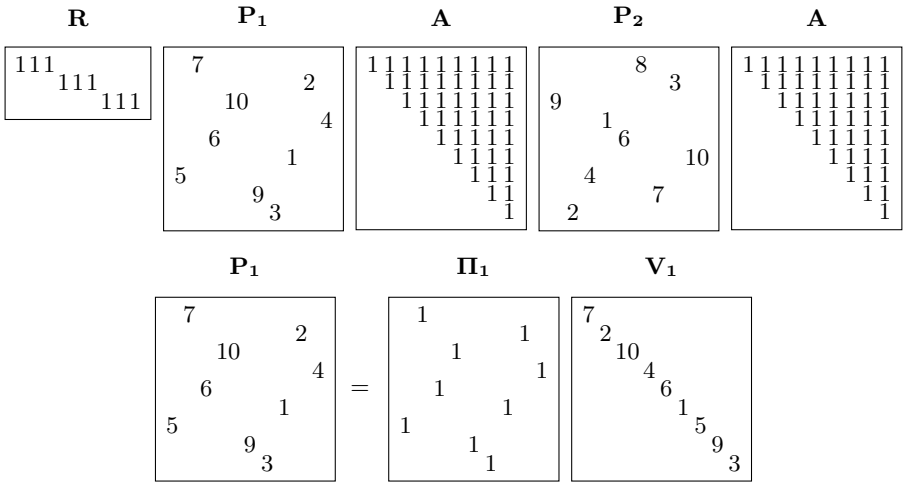

Fig. 1: A pictorial representation of the generator matrix for an RAA code over $\mathbb{F}_{11}$ with rate $r = 1/3$ and codeword length $N = 9$.

where $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ are two permutations of length N sampled uniformly at random, and $\mathbf{V}_1$ and $\mathbf{V}_2$ are $N \times N$ diagonal matrices with diagonal entries sampled uniformly from $\{1, 2, \ldots, q-1\}$. An example of the generator matrix is shown in Figure 1.

Though RAA codes with $r = 2$ can be asymptotically good [BMS08], $r \geq 3$ is required for good concrete parameters [KZC07, KZKC08]. In this paper, we consider the case of $r \geq 4$ for the convenience of our proof, and we believe our results can be generalized to the case of $r = 3$ with careful analysis.

In our analysis in the paper, we use w_0 to denote the weight of a specific message $\mathbf{x}$, $w_1 = rw_0$ to denote the weight of the vector after the repetition matrix $\mathbf{R}$, w_2 as the weight after the first accumulator matrix $\mathbf{A}$, and w_3 as the weight after the second accumulator matrix, which is the weight of the output codeword. The notations are summarized in Table 2.

2.3 IOWE of Accumulator Matrix A

The *input-output-weight-enumerator (IOWE)* counts the exact number of input vectors of weight w that lead to output vectors of weight h for a fixed matrix. Counting vectors in this way is an important technique used to analyze the distance of the family of RAA codes in the literature. The IOWE of the accumulator matrix $\mathbf{A}$ over a finite field $\mathbb{F}_q$ is given by the following theorem:

Symbol	Description
r	repetition factor
n	message length
N	codeword length, $N = rn$
$\mathbf{x}$	input vector $\in \mathbb{F}_q^n$
$\mathbf{R}$	repetition matrix $\in \mathbb{F}_q^{n \times N}$
$\mathbf{A}$	accumulator matrix, $\in \mathbb{F}_q^{N \times N}$
w_0	input weight, $\mathsf{wt}(\mathbf{x})$
w_1	weight after repetition, $\mathsf{wt}(\mathbf{xR}) = rw_0$
w_2	weight after the first accumulation, $\mathsf{wt}(\mathbf{xRP}_1\mathbf{A})$
w_3	output weight, $\mathsf{wt}(\mathbf{xRP}_1\mathbf{AP}_2\mathbf{A})$

Table 2: Notation.

Theorem 1 ([AR11, BFK+24]). *The IOWE of the accumulator matrix $\mathbf{A} \in \mathbb{F}_q^{N \times N}$ with input weight w and output h is:*

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

We restate the proof in [BFK+24] below with only minor notational changes and highlight the critical properties of the IOWE we use in our analysis.

Proof. Let $\mathbf{x}$ denote the input to $\mathbf{A}$, and $\mathbf{y} = \mathbf{x} \cdot \mathbf{A}$ be the corresponding output. Equivalently, $\mathbf{y}_a = \sum_{\iota=1}^{a} \mathbf{x}_\iota$. Following the proof in [BFK+24], we can define a *run* in $\mathbf{x}$ to be a sequence of consecutive indices $a, a+1, \ldots, b-1, b$ such that $y_{a-1} = 0$, $y_a, y_{a+1}, \ldots, y_{b-1} \neq 0$, and $y_b = 0$. We say that $\mathbf{x}_a$ *starts* the run and $\mathbf{x}_b$ *ends* it. We define a *non-run* in $\mathbf{x}$ to be a sequence of consecutive indices that are between two runs. We adopt the convention that $\mathbf{x}$ always starts with a non-run which can be empty. As an example, let $\mathbf{x} = (5, 2, 0, 3, 1, 0, 0, 0, 7, 4, \ldots)$ over $\mathbb{F}_{11}$. Then,

$$(\underbrace{\quad}_{\text{1st non-run}} \underbrace{5, 2, 0, 3, 1,}_{\text{1st run}} \underbrace{0, 0, 0}_{\text{2nd non-run}} , \underbrace{7, 4,}_{\text{2nd run}} \ldots).$$

Notice that runs always start and end with non-zero elements and non-runs only contain zeroes. However, the elements within the run can be zero or non-zero.

Let i denote the number of nonzero elements of $\mathbf{x}$ that do not start or end a run. If $\mathsf{wt}(\mathbf{x}) = w$, then the number of runs is $\phi = \lceil \frac{w-i}{2} \rceil$, as there must be $w - i$ elements that start or end a run. If there is an odd number of such elements, then that means that the last run does not end, which explains the need for the ceiling. The number of non-runs is $\phi' = \lfloor \frac{w-i}{2} \rfloor + 1$, where the 1 comes from our convention of always starting with a non-run.

In order to have output weight $\mathsf{wt}(\mathbf{y}) = h$, we must have $h - \phi$ entries within a run. This is from subtracting the run starts; the run ends correspond to a zero output entry. There are $\binom{h-1}{\phi-1}$ ways to place $h - \phi$ indistinguishable entries in

ϕ runs. The remaining $(N-w)-(h-\phi-k)$ zeros must be placed in the ϕ' non-runs; There are $\binom{N-h}{\phi'}$ ways of arranging them. So,

$$\binom{h-1}{\phi-1}\binom{N-h}{\phi'}$$

fixes the support (the location of nonzero entries) of the output $\mathbf{y}$. To emphasize, all entries in $\mathbf{y}$ corresponding to a run start or an internal run entry are nonzero, while all entries that end a run or are within a nonrun are zeros.

Within a run, recall that there can be zero entries and nonzero entries. There are $\binom{h-\phi}{i}$ ways to place the i internal nonzero elements in the $h-\phi$ internal run spots in $\mathbf{x}$. So, for a fixed i, there are at most

$$\binom{h-1}{\phi-1}\binom{N-h}{\phi'}\binom{h-\phi}{i}$$

possible arrangements of nonzero w values in $\mathbf{x}$ that will map to an output of weight h. We can think of this as fixing the support for $\mathbf{x}$.

Each of the ϕ runs must start with a nonzero value, giving $(q-1)^\phi$ possibilities. Each of the i internal run nonzero elements can take $(q-2)$ values, as they additionally cannot take the value that ends the run. Once the run start and the internal run values are selected, the value of the run end is fixed. This gives the quantity

$$\binom{h-1}{\phi-1}\binom{N-h}{\phi'}\binom{h-\phi}{i}(q-1)^\phi(q-2)^i$$

Summing over all possible internal run nonzeros $i=0$ to $w-1$ gives the result.

Properties of IOWE. A simple fact about the output weight is that $h \geq \lceil \frac{w}{2} \rceil$, i.e., the output weight cannot be less than half of the input weight after the accumulator matrix (see [BFK+24, Lemma 2.1] for a proof). Moreover, the third binomial coefficient in Theorem 1 implies that $i \leq h - \lceil \frac{w-i}{2} \rceil$, or equivalently $i \leq 2h - w + 1$. That is, h can be as small as $\lceil \frac{w}{2} \rceil$, but in the range of the summation $0 \leq i \leq w-1$, when $2h-w+1 < w-1$, the terms in the summand of Equation (1) for $2h-w+1 < i \leq w-1$ are all 0, following the convention that $\binom{a}{b} = 0$ if $b > a$.

An important fact that we will use later is that for fixed w, h, i, there are

$$\binom{h-1}{\phi-1}\binom{N-h}{\phi'}$$

possible supports for $\mathbf{y}$, corresponding to the choices of run starts and the placement of run ends and non-run entries. All run starts and entries within a run in $\mathbf{x}$ will create a nonzero entry in $\mathbf{y}$, and all run ends and nonrun entries will correspond to a zero in $\mathbf{y}$. Now, for a fixed support for $\mathbf{y}$, there are $\binom{h-\phi}{i}$ possible supports for $\mathbf{x}$ that could map to $\mathbf{y}$ under $\mathbf{A}$. This is because the support of $\mathbf{y}$

fixes the positions and lengths of the runs and nonruns, but not the position of the internal run nonzero entries. Recall that there can be zeros inside of a run. The i nonzero elements that do not start or end a run can be placed in any of the $h - \phi$ internal run positions.

Consequently, the total number of support pairs $(s_{\mathbf{x}}, s_{\mathbf{y}})$ such that a vector of support $s_{\mathbf{x}}$ can map to a vector of support $s_{\mathbf{y}}$ under $\mathbf{A}$ is

$$\binom{h-1}{\phi-1}\binom{N-h}{\phi'}\binom{h-\phi}{i}$$

This fact will be crucial for our analysis later.

As a special case over the binary field $\mathbb{F}_2$, the IOWE becomes

$$A_{w,h}^{N,2} = \binom{h-1}{\lceil \frac{w}{2} \rceil - 1}\binom{N-h}{\lfloor \frac{w}{2} \rfloor}, \tag{2}$$

which is exactly the same as the first term of the sum in Equation (1) with $i = 0$ and $q = 2$.

3 Low-Weight Messages

In this section, we prove that for a randomly sampled RAA code over a large finite field, the probability that there exists a *low-weight* message with low codeword weight is very small. Combined with the high-weight case in Section 4, we prove that the RAA code is asymptotically good. Moreover, we show for the first time that the distance over a large finite field is much better than the RAA code over the binary field, with a much smaller failure probability. Formally speaking, we will prove the following theorem:

Theorem 2. *For a randomly sampled RAA code over $\mathbb{F}_q$ as defined in Section 2.2, for nonzero messages $\mathbf{x}$ with weight $\mathsf{wt}(x) \leq \frac{n}{c' \log^2 n}$, for any constant relative distance $0 < \delta \leq \frac{1}{2}$,*

$$\Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x} \cdot \mathbf{G}) \leq \delta N\right] \leq \tilde{O}(N^{-\eta}) + O(\frac{1}{q-1}), \tag{3}$$

where $\mathbf{G} = \mathbf{R} \cdot \mathbf{P}_1 \cdot \mathbf{A} \cdot \mathbf{P}_2 \cdot \mathbf{A}$ is the encoding matrix of the RAA code and the probability is over the random choices of $\mathbf{P}_1$ and $\mathbf{P}_2$.

That is, the output weights of all low-weight messages are larger than δN, except with a small probability that is inverse polynomial in N, plus a term that is inverse in q. The threshold $\frac{n}{c' \log^2 n}$ of low-weight messages will become clear in our proving techniques, and constants $c', \eta > 1$ will be specified later.

We will divide the analysis of Theorem 2 into two stages. The first stage is the RA code part, which corresponds to multiplying a message $\mathbf{x}$ by $\mathbf{RP}_1\mathbf{A}$. The second stage is further multiplying the result of stage 1 by $\mathbf{P}_2\mathbf{A}$. We will first prove that stage 1 amplifies the weight of message from w_0 to $w_2 > cw_0 \log N$ for all low weight messages. Then, we will prove the distance property after stage 2 conditioning on the amplification from stage 1.

3.1 Probability of Amplification in Stage 1

In this section, we will prove that with high probability, the weights of all low-weight messages are amplified by the first stage of the RAA code. Conditioning on the first stage amplification will allow us to prove a tighter failure probability bound for the second stage.

Lemma 1. *For a uniformly sampled* $\mathbf{P}_1$*,*

$$\Pr_{\mathbf{P}_1}[\exists \mathbf{x} \in \mathbb{F}_q^n \text{ s.t. } 0 < \mathsf{wt}(\mathbf{x}) < n/c' \log^2 n \wedge \mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) \leq cw_0 \log N] \leq \tilde{O}(N^{-\eta_2}) + O\left(\frac{1}{q-1}\right).$$

Proof. Recall that for a fixed i, there are

$$S(i, w_1, w_2) := \binom{w_2 - 1}{\lceil \frac{w_1 - i}{2} \rceil - 1} \binom{N - w_2}{\lfloor \frac{w_1 - i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1 - i}{2} \rceil}{i}$$

possible supports that will map an input of weight w_1 to an output of weight w_2 under accumulation. In total, there are $\binom{N}{w_1}$ possible supports of size w_1. Since the permutation $\mathbf{P}_1$ is uniform, we have

$$\Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) = w_2 \text{ with } i \text{ internal run nonzeros} | \mathrm{supp}(\mathbf{xR}) = s_1] \leq \frac{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor}\binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}}$$

The reason why this is not an exact equality is that vectors with the same support may map to a different weight depending on the values of the nonzeros. To remove the conditional, notice that for any input vector $\mathbf{x}$, there is exactly one possible support $\mathbf{xR}$. So, our condition $\{\mathrm{supp}(\mathbf{xR}) = s_1\}$ is equivalent to a fixed support of $\mathbf{x}$, $\{\mathrm{supp}(\mathbf{x}) = s_0\}$, where $|\mathrm{supp}(\mathbf{x})| = w_0 = w_1/r$. Taking a union bound over all supports of $\mathbf{x}$ of weight w_0 and reindexing $w_1 = rw_0$, we have the following:

$$\Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) = w_2 \text{ with } i \text{ internal run nonzeros} \mid \mathsf{wt}(\mathbf{x}) = w_0] \leq \binom{n}{w_0} \frac{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2 - \lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{N}{rw_0}}$$

We can get a different upper bound by considering the values as well as the support:

$$\Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) = w_2 \text{ with } i \text{ internal run nonzeros} | \mathrm{supp}(\mathbf{xR}) = w_1 \text{ with fixed values }] = \frac{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor}\binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}} \frac{(q-1)^{\lceil \frac{w_1-i}{2} \rceil}(q-2)^i}{(q-1)^{w_1}}$$

Since $\mathbf{xR}$ is fixed by $\mathbf{x}$, we can once again formulate the condition in terms of the support and values of $\mathbf{x}$ and take a union bound for all inputs of $\mathsf{wt}(\mathbf{x}) = w_0$.

$$\Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) = w_2 \text{ with } i \text{ internal run nonzeros} | \mathsf{wt}(\mathbf{x}) = w_0]$$
$$\leq \binom{n}{w_0}(q-1)^{w_0} \frac{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2-\lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{N}{rw_0}} \frac{(q-1)^{\lceil \frac{rw_0-i}{2} \rceil}(q-2)^i}{(q-1)^{rw_0}}$$

Now, we have two upper bounds for the same probability. Notice that for some values of i, the first bound is smaller, but for other values of i, the second bound will be smaller depending on the value of q. The idea is to take a union bound over i using the minimum of the two upper bounds. In particular, when $i \leq (r-2)w_0 - 1$, the second bound is $O\left((q-1)^{-1}\right)$. Invoking this refined union bound over i, we get

$$\Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) = w_2 | \mathsf{wt}(\mathbf{x}) = w_0]$$
$$\leq \sum_{i=(r-2)w_0-1}^{rw_0-1} \binom{n}{w_0} \frac{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2-\lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{N}{rw_0}} + O\left(\frac{1}{q-1}\right)$$

Since we wanted to bound the probability of insufficient amplification, $\mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) \leq cw_0 \log N$, we will need to take union bounds over the insufficient w_2 and the low-weight w_0.

$$\Pr_{\mathbf{P}_1}[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{xRP}_1\mathbf{A}) \leq cw_0 \log N \,|\, w_0 \leq n/c' \log^2 n]$$
$$\leq \sum_{w_0=1}^{n/c' \log^2 n} \sum_{w_2=1}^{cw_0 \log N}$$
$$\left[\sum_{i=(r-2)w_0-1}^{rw_0-1} \binom{n}{w_0} \frac{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2-\lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{N}{rw_0}} + O\left(\frac{1}{q-1}\right)\right]$$
$$= \sum_{w_0=1}^{n/c' \log^2 n} \sum_{w_2=1}^{cw_0 \log N} \left[\sum_{i=(r-2)w_0-1}^{rw_0-1} \binom{n}{w_0} \frac{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2-\lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{N}{rw_0}}\right]$$
$$+ O\left(\frac{1}{q-1}\right). \quad (4)$$

Treatment of the term $O(\frac{1}{q-1})$. As shown in the proof above, the term $O(\frac{1}{q-1})$ hides factors that depend on N. Technically, the term could still be large even if $q >> N$. However, as shown in [BFK+24], these factors related to N are also inverse polynomials in N for some parameter regimes. Moreover, this term can be made smaller for larger r, e.g., $O(\frac{1}{(q-1)^2})$ for $r = 6$ with a similar analysis.

Therefore, we view the term as extremely small and focus on the bracketed combinatorial terms in the rest of this paper. In a subsequent work [Kha26], Majid Khabbazian shows that if $q-1 \geq (eN)^2$, the failure probability is completely dominated by the combinatorial terms.

Asymptotic behavior of the combinatorial terms. To complete the proof of Lemma 1, we will prove two auxiliary lemmas for $S(i, w_1, w_2) = \binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}$.

Lemma 2. *For all $w_2 \leq N/2$, $S(i, w_1, w_2)$ is increasing with w_2 on the interval $i \in [(r-2)w_0 - 1, rw_0 - 1]$.*

Proof. We will consider the ratio between consecutive terms over w_2 and show that the ratio is strictly greater than 1 under our conditions.

$$\begin{aligned}
\frac{S(i, w_1, w_2+1)}{S(i, w_1, w_2)} &= \frac{\binom{w_2}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_2-1}{\lfloor \frac{w_1-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil + 1}{i}}{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_2}{\lfloor \frac{w_1-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}} \\
&= \frac{w_2}{w_2 - \lceil \frac{w_1-i}{2} \rceil + 1} \frac{N - w_2 - \lfloor \frac{w_1-i}{2} \rfloor}{N - w_2} \frac{w_2 - \lceil \frac{w_1-i}{2} \rceil + 1}{w_2 - \lceil \frac{w_1-i}{2} \rceil - i + 1} \\
&= \frac{w_2}{w_2 - \lceil \frac{w_1-i}{2} \rceil - i + 1} \frac{N - w_2 - \lfloor \frac{w_1-i}{2} \rfloor}{N - w_2}
\end{aligned}$$

To show that this is increasing with w_2, we need to have the following:

$$\begin{aligned}
&\frac{w_2}{w_2 - \lceil \frac{w_1-i}{2} \rceil - i + 1} \frac{N - w_2 - \lfloor \frac{w_1-i}{2} \rfloor}{N - w_2} > 1 \\
&\Leftrightarrow w_2 \left(N - w_2 - \lfloor \frac{w_1 - i}{2} \rfloor \right) > (N - w_2) \left(w_2 - \lceil \frac{w_1 - i}{2} \rceil - i + 1 \right) \\
&\Leftrightarrow w_2 (N - w_2) - w_2 \left(\lfloor \frac{w_1 - i}{2} \rfloor \right) > w_2(N - w_2) - (N - w_2) \left(\lceil \frac{w_1 - i}{2} \rceil + i - 1 \right) \\
&\Leftrightarrow w_2 \left(\lfloor \frac{rw_0 - i}{2} \rfloor \right) < (N - w_2) \left(\lceil \frac{rw_0 - i}{2} \rceil + i - 1 \right)
\end{aligned}$$

Since we are considering $i \in [(r-2)w_0 - 1, rw_0 - 1]$, a sufficient condition can be derived by maximizing the LHS over i and minimizing the RHS:

$$\begin{aligned}
w_2 w_0 &< (N - w_2)((r-2)w_0 - 1) \\
w_2 &< N \frac{(r-2)w_0 - 1}{(r-1)w_0 - 1}
\end{aligned}$$

Thus, since $r \geq 4$, we can say that $S(i, w_1, w_2)$ is increasing with w_2 for $w_2 \leq N/2$. □

Lemma 3. *Let $w_1 \geq 4w_0$. Then, the summand $S(i, w_1, cw_0 \log N)$ is decreasing with i on the interval $i \in [(r-2)w_0 - 1, rw_0 - 1]$.*

Proof. We will split this into cases corresponding to the parity of $rw_0 - i$.
Case 1 (odd $rw_0 - i$). Let $rw_0 - i = 2m + 1$, where m is an integer in the interval $[0, w_0]$. We will consider the ratio between consecutive terms over i and show that it is strictly less than 1 as done in the corresponding lemma for δN.

$$\begin{aligned}\frac{S(i+1, w_1, cw_0 \log N)}{S(i, w_1, cw_0 \log N)} &= \frac{\binom{cw_0 \log N - 1}{m-1}\binom{N - cw_0 \log N}{m}\binom{cw_0 \log N - m}{i+1}}{\binom{cw_0 \log N - 1}{m}\binom{N - cw_0 \log N}{m}\binom{cw_0 \log N - m - 1}{i}} \\ &= \frac{m}{cw_0 \log N - m}\frac{cw_0 \log N - m}{i+1} = \frac{m}{rw_0 - 2m}\end{aligned}$$

In order to have $\frac{m}{rw_0 - 2m} < 1$, we need $rw_0 > 3m$, or $r > 3$.

Case 2 (even $rw_0 - i$). Let $rw_0 - i = 2m$, where m is now in the interval $[1, w_0]$. Since we are considering low input weights, we can use the following bound:

$$\begin{aligned}w_0 &\leq \frac{n}{c' \log^2 n} = \frac{N/r}{c' \log^2 (N/r)} = \frac{N/r}{c'(\log N - \log r)^2)} \\ &= \frac{N}{rc'(\log^2 N - 2 \log N \log r - \log^2 r)} \leq \frac{N}{rc'(\log^2 N - 3 \log N \log r)}\end{aligned}$$

We will again compute the ratio between consecutive terms over i:

$$\begin{aligned}&\frac{S(i+1, w_1, cw_0 \log N)}{S(i, w_1, cw_0 \log N)} = \frac{\binom{cw_0 \log N - 1}{m-1}\binom{N - cw_0 \log N}{m-1}\binom{cw_0 \log N - m}{i+1}}{\binom{cw_0 \log N - 1}{m-1}\binom{N - cw_0 \log N}{m}\binom{cw_0 \log N - m}{i}} \\ &= \frac{m}{N - cw_0 \log N - m + 1}\frac{cw_0 \log N - m - i}{i+1} \\ &\leq \frac{cw_0 \log N}{N - cw_0 \log N - w_0}\frac{w_0}{(r-2)w_0} \\ &\leq \frac{cw_0 \log N}{N(1 - \frac{c}{rc'(\log N - 3 \log r)} - \frac{1}{rc'(\log^2 N - 3 \log N \log r)})}\frac{1}{r-2} \\ &\leq \frac{cw_0 \log N}{N/2}\frac{1}{r-2} \\ &\leq \frac{2c}{c'r(r-2)(\log N - 3 \log r)} \\ &\leq \frac{c}{c'r(\log N - 3 \log r)}\end{aligned}$$

The third to last inequality follows from c and r being constants. The last line follows from $r \geq 4$. Since c is a constant, we can conclude that the term is strictly less than 1 for large enough N. Thus $S(i, w_1, cw_0 \log N)$ is decreasing with i on the interval $i \in [(r-2)w_0 - 1, rw_0 - 1]$. □

With these two lemmas, focusing on the combinatorial terms of Equation (4):

$$\sum_{w_0=1}^{n/c'\log^2 n}\sum_{w_2=1}^{cw_0\log N}\binom{n}{w_0}\sum_{i=(r-2)w_0-1}^{rw_0-1}\frac{\binom{w_2-1}{\lceil\frac{rw_0-i}{2}\rceil-1}\binom{N-w_2}{\lfloor\frac{rw_0-i}{2}\rfloor}\binom{w_2-\lceil\frac{rw_0-i}{2}\rceil}{i}}{\binom{N}{rw_0}}, \tag{5}$$

by Lemma 2, as $w_2 \le N/2$, the numerator of the summand $S(i, w_1, w_2)$ increases with w_2, and thus we bound Equation (5) by the maximum of w_2 at $cw_0 \log N$:

$$\sum_{w_0=1}^{n/c'\log^2 n}(cw_0\log N)\binom{n}{w_0}\sum_{i=(r-2)w_0-1}^{rw_0-1}\frac{\binom{cw_0\log N-1}{\lceil\frac{rw_0-i}{2}\rceil-1}\binom{N-cw_0\log N}{\lfloor\frac{rw_0-i}{2}\rfloor}\binom{cw_0\log N-\lceil\frac{rw_0-i}{2}\rceil}{i}}{\binom{N}{rw_0}}$$

By Lemma 3, the numerator decreases with i as well. Therefore, we can bound the equation above as

$$\le \sum_{w_0=1}^{n/c'\log^2 n}(2w_0+1)(cw_0\log N)\binom{n}{w_0}\frac{\binom{cw_0\log N-1}{w_0}\binom{N-cw_0\log N}{w_0}\binom{cw_0\log N-w_0}{(r-2)w_0-1}}{\binom{N}{rw_0}}$$

Consider the individual summand for a fixed w_0. First, we will apply the following properties of the binomial coefficient: $\binom{n-1}{k} = \frac{n-k}{n}\binom{n}{k}$, $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$.

$$\begin{aligned}
&(2w_0+1)(cw_0\log N)\binom{n}{w_0}\frac{\binom{cw_0\log N-1}{w_0}\binom{N-cw_0\log N}{w_0}\binom{cw_0\log N-w_0}{(r-2)w_0-1}}{\binom{N}{rw_0}}\\
&= (2w_0+1)(cw_0\log N-w_0)\frac{(r-2)w_0}{cw_0\log N-(r-1)w_0+1}\\
&\qquad\cdot\binom{n}{w_0}\frac{\binom{cw_0\log N}{w_0}\binom{N-cw_0\log N}{w_0}\binom{cw_0\log N-w_0}{(r-2)w_0}}{\binom{N}{rw_0}}\\
&\le \frac{3w_0^2(r-2)}{1-(r-1)/c\log N}\left(\frac{en}{w_0}\right)^{w_0}\left(\frac{w_0}{n}\right)^{rw_0}\left(\frac{cew_0\log N}{w_0}\right)^{w_0}\\
&\qquad\cdot\left(\frac{e(N-cw_0\log N)}{w_0}\right)^{w_0}\left(\frac{e(cw_0\log N-w_0)}{(r-2)w_0}\right)^{(r-2)w_0}\\
&\le \frac{3w_0^2(r-2)}{1-(r-1)/c\log N}\left(\frac{eN}{rw_0}\right)^{w_0}\left(\frac{rw_0}{N}\right)^{rw_0}(ce\log N)^{w_0}\\
&\qquad\cdot\left(\frac{eN}{w_0}\right)^{w_0}\left(\frac{ec\log N}{r-2}\right)^{(r-2)w_0}\\
&\le \frac{3w_0^2(r-2)}{1-(r-1)/c\log N}\left[\frac{e^{r+1}w_0^{r-2}r^{r-1}}{(r-2)^{r-2}}\frac{(c\log N)^{r-1}}{N^{r-2}}\right]^{w_0}\\
&= \frac{3w_0^2(r-2)}{1-(r-1)/c\log N}\left(\frac{e^{r+1}r^{r-1}}{(r-2)^{r-2}}\right)^{w_0}\left(\frac{w_0^{r-2}(c\log N)^{r-1}}{N^{r-2}}\right)^{w_0}
\end{aligned} \tag{6}$$

Since e, c, r are constants, the equation above is equal to

$$O\left(w_0^2(g\log N)^{w_0}(\frac{w_0 c\log N}{N})^{(r-2)w_0}\right), \tag{7}$$

where $g = \frac{e^{r+1}r^{r-1}}{(r-2)^{r-2}}$ is a constant.

We further differentiate the following two cases:

Case 1: $1 \le w_0 \le O(\log N)$.

In this case, the numerator is $\mathsf{polylog}(N)^{w_0}$ while the denominator is $\mathsf{poly}(N)^{w_0}$. Therefore, the upper bound is $O\left(\frac{(\log N)^{(r-1)}}{N^{(r-2)}}\right)$ after summing over all w_0, and the dominating term is $w_0 = 1$. This is consistent with the intuition that the nonzero messages with the least weight of 1 are the most challenging to amplify.

Case 2: $w_0 = \omega(\log N)$.

In this case, for any $w_0 \le n/c'\log^2 n$, as $N = rn$, Equation (7) is $\le O\left(w_0^2(\frac{c^{r-2}g}{(rc')^{r-2}})^{w_0}\frac{1}{\log N}^{(r-3)w_0}\right)$. As $r \ge 4$, c, r, c', g are all constants, it is $O\left((\frac{c''}{(\log N)^{(r-3)}})^{w_0}\right)$ for $c'' = \frac{c^{r-2}g}{(rc')^{r-2}}$, which is negligible for $w_0 = \omega(\log N)$. As Equation (7) is summed over $w_0 = \omega(\log N)$ to $\frac{n}{c'\log^2 n}$, the probability is still negligible after the summation.

Combining the two cases completes the proof of Lemma 1, where $\eta_2 = r - 2$ due to the first case dominated by $w_0 = 1$. □

Moreover, Equation (7) explains the reason of the threshold $\frac{n}{c'\log^2 n}$ of low-weight messages. The term $\frac{w_0 c\log N}{N}$ needs to be at most $\frac{1}{\log N}$ to cancel other terms in the numerator and make the bound smaller than 1.

In Appendix A.1, we prove a bound for the amplification of the first stage of the RAA code over the binary field where the failure probability is $\tilde{O}(N^{1-\frac{r}{2}})$. As explained in the introduction, this is a better result for the binary RAA than the bound in [BCF+25, Theorem 9.1], where the failure probability is $O(N^{1+\gamma-\frac{r}{2}})$ for a constant $0 < \gamma < 1$. In either case, we clearly show that the failure probability over a large finite field ($\tilde{O}(N^{2-r})$) is strictly smaller than the RAA code over the binary field. For example, when $r = 4$, the failure probability over a large finite field is $\tilde{O}(N^{-2})$, which is already good enough for large N, while the failure probability over the binary field is only $\tilde{O}(N^{-1})$.

3.2 RAA Failure Probability

We will now derive a similar bound as Lemma 1 for both stages of the RAA code. This time, let i denote the number of internal run elements in $\mathbf{x}' := \mathbf{xR\Pi}_1$ and j denote the number of internal run elements in $\mathbf{x}'' := \mathbf{xR\Pi}_1\mathbf{A}$.

Lemma 4. *For uniformly sampled* $\mathbf{P}_1$ *and* $\mathbf{P}_2$*,*

$$\Pr_{\mathbf{P_1},\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3] \leq \sum_{i+j>w_2+(r-2)w_0-4} \binom{n}{w_0} \frac{\binom{w_2-1}{\phi-1}\binom{N-w_3}{\phi'}\binom{w_2-\phi}{i}}{\binom{N}{w_1}} \cdot \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} + O\left(\frac{1}{q-1}\right) \tag{8}$$

Proof. Let us fix i, j, w_0, w_2, w_3. We will use the law of total probability to derive an upper bound for the probability of having a final output weight w_3 with j internal run entries given an intermediate weight of w_2 with i internal run entries and an input weight of w_0. We will exploit properties of both the repetition matrix $\mathbf{R}$ and the accumulator $\mathbf{A}$.

For a fixed support $s_{\mathbf{x}''}$ of weight w_2, similar to the first accumulator in Section 3.1,

$$\Pr_{\mathbf{\Pi}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries } |\mathrm{supp}(\mathbf{x}'') = s_{\mathbf{x}''}] \leq \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}}$$

where $\tau = \left\lceil \frac{w_2-j}{2} \right\rceil, \tau' = \left\lfloor \frac{w_2-j}{2} \right\rfloor + 1$ denote the number of runs and non-runs respectively. This is just by the definition of the binomial coefficient terms and the fact that $\mathbf{\Pi}_2$ is a uniformly drawn permutation.

For a fixed support $s_{\mathbf{x}'}$ of weight w_1, let $\#(s_{\mathbf{x}'})$ denote the number of possible supports $s_{\mathbf{x}''}$ of weight w_2 with i internal run entries that can come from $s_{\mathbf{x}'}$ after accumulation. Then,

$$\Pr_{\mathbf{\Pi}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries } |\mathrm{supp}(\mathbf{x}') = s_{\mathbf{x}'}] \leq \#(s_{\mathbf{x}'}) \cdot \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}}.$$

For a fixed input support $s_{\mathbf{x}}$, there is exactly one corresponding support $s_{\mathbf{xR}}$ after repetition. This support is permuted to $s_{\mathbf{x}'}$ with probability $\frac{1}{\binom{N}{w_1}}$. So, we can say that

$$\Pr_{\mathbf{\Pi_1},\mathbf{\Pi}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries } |\mathrm{supp}(\mathbf{x}) = s_{\mathbf{x}}] \leq \frac{\#(s_{\mathbf{x}'})}{\binom{N}{w_1}} \frac{\binom{w_3-1}{r-1}\binom{N-w_3}{r'}\binom{w_3-r}{j}}{\binom{N}{w_2}}.$$

If we union bound over all supports $s_{\mathbf{x}}$ of weight w_0, we have

$$\Pr_{\mathbf{\Pi_1},\mathbf{\Pi}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries }] \leq \binom{n}{w_0} \frac{\sum_{\mathbf{x}' \in \mathrm{supp}(N,w_0)} \#(s_{\mathbf{x}'})}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}}.$$

Now we consider $\sum_{\mathbf{x}' \in \mathrm{supp}(N,w_0)} \#(s_{\mathbf{x}'})$ with the argument discussed in Section 2. For a fixed i, we know that there are $\binom{w_2-1}{\phi-1}\binom{N-w_2}{\phi'}$ supports for $\mathbf{x}'\mathbf{A}$, where ϕ, ϕ' denote the number of runs and non-runs respectively. There are $\binom{w_2-1}{\phi-1}\binom{N-w_3}{\phi'}\binom{w_2-\phi}{i}$ support pairs $(s_{\mathbf{x}'}, s_{\mathbf{x}''})$ such that a vector of support $s_{\mathbf{x}'}$ can map to a vector of support $s_{\mathbf{x}''}$ under $\mathbf{A}$. Using this fact, we have our first upper bound:

$$\begin{aligned}
&\Pr_{\boldsymbol{\Pi}_1,\boldsymbol{\Pi}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries }] \\
&\quad \leq \binom{n}{w_0} \frac{\binom{w_2-1}{\phi-1}\binom{N-w_2}{\phi'}\binom{w_2-\phi}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}}. \qquad (9)
\end{aligned}$$

Next, if we consider the values, we can derive a different upper bound. For notation purposes, we emphasize that we denote the number of runs and nonruns of $\mathbf{x}''$ with τ and τ' respectively and the number of runs and nonruns of $\mathbf{x}'$ with ϕ and ϕ'.

$$\begin{aligned}
&\Pr_{\mathbf{P}_1,\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries }] \\
&\quad = \Pr_{\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3 \wedge \mathbf{x}''\mathbf{A} \text{ has } j \text{ internal run entries } | \mathsf{wt}(\mathbf{x}'') = w_2 \\
&\qquad \wedge \mathbf{x}'' \text{ has } i \text{ internal run entries }] \cdot \Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{x}'') = w_2 \wedge \mathbf{x}'' \text{ has } i \text{ internal run entries }] \\
&\quad \leq \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} \frac{(q-1)^\tau (q-1)^j}{(q-1)^{w_2}} \\
&\qquad \cdot \Pr_{\mathbf{P}_1}[\mathsf{wt}(\mathbf{x}'') = w_2 \wedge \mathbf{x}'' \text{ has } i \text{ internal run entries } | \mathsf{wt}(\mathbf{x}') = w_1] \\
&\quad \leq \binom{n}{w_0}(q-1)^{w_0} \frac{\binom{w_3-1}{\phi-1}\binom{N-w_3}{\phi'}\binom{w_3-\phi}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} \\
&\qquad \cdot \frac{(q-1)^\phi (q-1)^i}{(q-1)^{w_1}} \frac{(q-1)^\tau (q-1)^j}{(q-1)^{w_2}} \qquad (10)
\end{aligned}$$

where the final line follows from a union bound over all messages of weight w_0.

We will now use a similar idea as in the first stage to union bound over i and j; use the minimum of both upper bounds. However, since we are considering both RAA stages simultaneously, we have different bounds over i and j to get our $O(1/(q-1))$ term. In particular, our condition on i and j is that $w_0 + \phi + i + \tau + j - w_1 - w_2 \leq -1$. Note that $\phi + i = \lceil (w_1 - i)/2 \rceil + i = w_1 - \lfloor (w_1 - i)/2 \rfloor$ and $\tau + j = \lceil (w_2 - j)/2 \rceil + j = w_2 - \lfloor (w_2 - j)/2 \rfloor$, so this condition is equivalent to

$$w_0 - \left\lfloor \frac{w_1 - i}{2} \right\rfloor - \left\lfloor \frac{w_2 - j}{2} \right\rfloor \leq -1.$$

A slightly looser requirement is

$$w_0 - \left\lfloor \frac{w_1 + w_2 - i - j}{2} \right\rfloor + 1 \leq -1.$$

If we have that $w_1 + w_2 - i - j \geq 2w_0 + 4$, or $i + j \leq w_2 + (r-2)w_0 - 4$, we have that the corresponding summand is $O(1/(q-1))$. Finally, we apply the union bound over i and j, resulting in:

$$\Pr_{\mathbf{P_1},\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3] \leq \sum_{i+j>w_2+(r-2)w_0-4} \binom{n}{w_0} \frac{\binom{w_2-1}{\phi-1}\binom{N-w_3}{\phi'}\binom{w_2-\phi}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} + O\left(\frac{1}{q-1}\right)$$

□

3.3 Probability of Good Distance in Stage 2

Now, we will consider the full RAA code for low-weight messages conditioning on the first stage succeeding. We will use Lemma 4, derived in the previous section. In particular, since $w_0 \geq 1$, we can use the following ranges for i and j:

$$\Pr_{\mathbf{P_1},\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3] \leq \sum_{i=(r-2)w_0-1}^{rw_0-1} \sum_{j=w_2-2w_0-1}^{w_2-1} \binom{n}{w_0} \frac{\binom{w_2-1}{\phi-1}\binom{N-w_3}{\phi'}\binom{w_2-\phi}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} + O\left(\frac{1}{q-1}\right)$$

Lemma 5. *For randomly sampled* $\mathbf{P}_2$ *and possible messages* $\mathbf{x}$ *with weight* $0 < w_0 \leq \frac{n}{c' \log^2 n}$*, suppose the weight* $\mathsf{wt}(\mathbf{x}\mathbf{R}\mathbf{P}_1\mathbf{A}) = w_2 > cw_0 \log N$*. Then, for any constant relative distance* $0 < \delta < \frac{1}{2}$*,*

$$\Pr_{\mathbf{P}_2}[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x}\mathbf{R}\mathbf{P}_1\mathbf{A}\mathbf{P}_2\mathbf{A}) \leq \delta N] \leq O(N^{-\eta_1}) + O\left(\frac{1}{q-1}\right).$$

The full proof is provided in Appendix B.1.

Finally, by combining Lemma 5 in Section 3.2 and Lemma 1 in Section 3.1 through the law of total probability, we are able to prove Theorem 2 as:

$$\begin{aligned}
&\Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x}\cdot\mathbf{G}) \leq \delta N\right] \\
=&\Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x}\cdot\mathbf{G}) \leq \delta N | \text{ Stage 1 succeeds }\right] \cdot \Pr[\text{ Stage 1 succeeds }] \\
&+ \Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x}\cdot\mathbf{G}) \leq \delta N | \text{ Stage 1 fails }\right] \cdot \Pr[\text{ Stage 1 fails }] \\
\leq&\Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x}\cdot\mathbf{G}) \leq \delta N | \text{ Stage 1 succeeds }\right] + \Pr[\text{ Stage 1 fails }] \\
=&\tilde{O}(N^{-\eta_1}) + \tilde{O}(N^{-\eta_2}) + O(\frac{1}{q-1}) = \tilde{O}(N^{-\eta}) + O(\frac{1}{q-1}),
\end{aligned}$$

for $\eta = \min(\eta_1, \eta_2)$.

In Appendix A.2, we show that the upper bound of the probability in Lemma 5 over a large finite field $\mathbb{F}_q$ is strictly smaller than the provable bound over the binary field $\mathbb{F}_2$.

4 High-Weight Messages

In this section, we prove that for a randomly sampled RAA code, the probability that there exists a *high-weight* message with low codeword weight is negligible, which completes the full distance analysis of the RAA code.

Theorem 3. *For a randomly sampled RAA code over $\mathbb{F}_q$ as defined in Section 2.2, for nonzero messages $\mathbf{x}$ with weight $\mathsf{wt}(x) > \frac{n}{c' \log^2 n}$, for any constant relative distance $0 < \delta \leq \frac{1}{2}$,*

$$\Pr\left[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{x} \cdot \mathbf{G}) \leq \delta N\right] \leq \exp(-O(N/\log N)) + O(\frac{1}{q-1}),$$

where $\mathbf{G} = \mathbf{R} \cdot \mathbf{P}_1 \cdot \mathbf{A} \cdot \mathbf{P}_2 \cdot \mathbf{A}$ is the encoding matrix of the RAA code and the probability is over the random choices of $\mathbf{P}_1$ and $\mathbf{P}_2$.

To prove this theorem, we will use Section 3.2 again, but this time union bounding over the high weight messages. Focusing on the combinatorial terms, the probability we are trying to bound is

$$\sum_{w_0=\frac{n}{c' \log^2 n}+1}^{n} \sum_{w_2=\left\lfloor \frac{rw_0}{2} \right\rfloor}^{2\delta N} \sum_{w_3=\left\lfloor \frac{w_2}{2} \right\rfloor}^{\delta N} \sum_{i=w_1-2w_0-1}^{w_1-1} \sum_{j=w_2-2w_0-1}^{w_2-1} \binom{n}{w_0} \frac{\binom{N-w_2}{\lceil \frac{w_1-i}{2} \rceil}\binom{w_2-1}{\lfloor \frac{w_1-i}{2} \rfloor}\binom{w_2-\lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}} \cdot \frac{\binom{N-w_3}{\lceil \frac{w_2-j}{2} \rceil}\binom{w_3-1}{\lfloor \frac{w_2-j}{2} \rfloor}\binom{w_3-\lceil \frac{w_2-j}{2} \rceil}{j}}{\binom{N}{w_2}}. \quad (11)$$

Note that the summand is exactly the same as the bound in [AHKM25, Appendix A]. However, the ranges of w_0, w_2, w_3 are different due to our fine-grained analysis, which are crucial to prove that the bound is negligible for high-weight messages in our case.

For this section, it will be helpful to work with a different formulation of the IOWE. In this new formulation, we consider summing over the number of runs instead of the number of internal run elements.

Theorem 4. *The IOWE of the accumulator matrix $\mathbf{A} \in \mathbb{F}_q^{N \times N}$ with input weight w and output h is:*

$$A_{w,h}^{N,q} = \sum_{k=\max(1,w-h)}^{\lfloor w/2 \rfloor} \left[\binom{h-1}{k-1}\binom{N-h}{k}\binom{h-k}{w-2k}(q-1)^k(q-2)^{w-2k} + \binom{h-1}{k-1}\binom{N-h}{k-1}\binom{h-k}{w-2k+1}(q-1)^k(q-2)^{w-2k+1} \right]. \quad (12)$$

Proof. This can be seen by re-indexing our original IOWE in Theorem 1 with the change of variables $k = \lceil \frac{w-i}{2} \rceil$. Under this transformation, the sum is over the number of runs k in the input vector. This will result in two terms in each summand corresponding to the cases of there being k nonruns and $k+1$ nonruns. □

As we primarily care about the asymptotic behavior and analyze the exponent of the failure probability in this section, we merge the two cases above via a very loose upper bound to simplify the computations. We take the ratio of the two terms in the summand.

$$\frac{\binom{h-1}{k-1}\binom{N-h}{k-1}\binom{h-k}{w-2k+1}}{\binom{h-1}{k-1}\binom{N-h}{k}\binom{h-k}{w-2k}} = \frac{k}{N-h-k+1} \cdot \frac{h+k-w}{w-2k+1}$$

Since $k, h+k-w \leq N$, we know that the numerator term is at most N^2 times the denominator term. Therefore, the equation below is an upper bound of Equation (11):

$$\sum_{w_0=\frac{n}{c'\log^2 n}+1}^{n} \sum_{w_2=\lfloor \frac{rw_0}{2} \rfloor}^{2\delta N} \sum_{w_3=\lfloor \frac{w_2}{2} \rfloor}^{\delta N} \sum_{i=0}^{w_0}\sum_{j=0}^{w_0} \binom{n}{w_0} N^2 \cdot \frac{\binom{N-w_2}{i}\binom{w_2-1}{i}\binom{w_2-i}{w_1-2i}}{\binom{N}{w_1}} \cdot \frac{\binom{N-w_3}{j}\binom{w_3-1}{j}\binom{w_3-j}{w_2-2j}}{\binom{N}{w_2}}, \tag{13}$$

which will be the equation we bound in the rest of this section.

Following the approach in [AR09], we write w_0, w_2, w_3, i, j in terms of N as:

$$w_0 = \alpha N^a \qquad w_2 = \beta N^b \qquad w_3 = \rho N^c \qquad i = \gamma_1 N^{e_1} \qquad j = \gamma_2 N^{e_2}.$$

With these, $w_1 = rw_0 = r\alpha N^a$. Moreover, as $w_0 > \frac{n}{c'\log^2 n}$ and the output weight of the accumulator matrix is no less than half of its input weight, we have $1 - \frac{2\log\log N}{\log N} \leq a \leq b \leq c \leq 1$. To prove that Equation (13) is negligible, we follow the approach in [AR09] and differentiate between two cases: (1) any of w_0, w_2, w_3 is sublinear in N (i.e., any of a, b, c is < 1); (2) all of them are linear in N (i.e., $a = b = c = 1$). We summarize the differences from [AR09] below:

- There are multiple sub-cases for Case (1) and the proofs of some sub-cases were omitted in [AR09]. We provide full proofs for all sub-cases.
- We observe that our fine-grained analysis is crucial to obtain good distance due to some sub-cases. In particular, we are able to show that the failure probability is negligible for any $\delta \leq \frac{1}{2}$ when the ranges of i, j are $[0, w_0]$ in Equation (13). In [AR09], i, j can be as large as $\lfloor \frac{w_1}{2} \rfloor$ and $\lfloor \frac{w_2}{2} \rfloor$, respectively, and the provable bound is smaller than 1 only for small δ (e.g., $\delta < 0.2$).
- The same summand in Equation (13) was used for all messages with nonzero weights to show that the probability converges to 0 when $N \to \infty$ in [AR09]. However, this does not yield good concrete numbers as the probability could be inverse polynomial in N with poor parameters. We show that by restricting $w_0 > \frac{n}{c'\log^2 n}$ for high-weight messages, the failure probability is negligible.
- For Case (2), the failure probability is shown to be negligible in [AR09] only for large r and small δ. We are able to show this for $r = 4$ and $\delta = \frac{1}{2}$ with our fine-grained analysis.

4.1 Sublinear Case

For clarity, define the following

$$S(i,j,w_0,w_2,w_3) := \binom{n}{w_0} \frac{\binom{N-w_2}{i}\binom{w_2-1}{i-1}\binom{w_2-i}{w_1-2i}}{\binom{N}{w_1}} \frac{\binom{N-w_3}{j}\binom{w_3-1}{j-1}\binom{w_3-j}{w_2-2j}}{\binom{N}{w_2}}$$

Lemma 6. *Assume at least one of w_0, w_2, w_3 is sublinear in N, then the summand term $S(i,j,w_0,w_2,w_3) = \exp(-O(N/\log N))$.*

The proof is deferred to Appendix C.1.

4.2 Linear Case

We are left with the case where $a = b = c = 1$, i.e.,w_0, w_1, w_2, w_3 are all linear in N. We rely on another variant of Stirling's inequality shown in [BCF+25, Lemma 9.4]:

$$0.61664 \cdot 2^{nH(\frac{k}{n})} \le \binom{n}{k} \le 0.67352 \cdot \frac{2^{n\cdot H(\frac{k}{n})}}{\sqrt{n}},$$

where $H(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function. The goal is to write $\binom{n}{k}$ as $\mathrm{poly}(n)\cdot 2^{n\cdot H(\frac{k}{n})}$ and the constants are not used explicitly in the derivation. By applying the bound above to our summand $N^2 \cdot S(i,j,w_0,w_2,w_3)$, we have

$$N^2 \cdot S(i,j,w_0,w_2,w_3) \le \mathrm{poly}(N) \cdot 2^{f(\alpha,\beta,\rho,\gamma_1,\gamma_2)N},$$

where

$$\begin{aligned} &f(\alpha,\beta,\rho,\gamma_1,\gamma_2) = \\ &H(r\alpha)(\frac{1}{r}-1) - H(\beta) + (1-\beta)H(\frac{\gamma_1}{1-\beta}) + \beta H(\frac{\gamma_1}{\beta}) + (\beta-\gamma_1)H(\frac{r\alpha-2\gamma_1}{\beta-\gamma_1}) \\ &\quad + (1-\rho)H(\frac{\gamma_2}{1-\rho}) + \rho H(\frac{\gamma_2}{\rho}) + (\rho-\gamma_2)H(\frac{\beta-2\gamma_2}{\rho-\gamma_2}), \end{aligned} \tag{14}$$

with the boundary conditions that $r\alpha/2 \le \beta \le 2\delta$, $\gamma_1, \gamma_2 \le \alpha$ and $\rho \le \delta$. By carefully bounding this function and showing it is always negative on the region, we can prove that each summand is $\exp(-O(N))$.

Lemma 7. *For $r = 4$ and $\delta = 1/2$, when w_0, w_2, w_3 are all linear in N, the summand $S(i,j,w_0,w_2,w_3) = \exp(-O(N))$.*

In Appendix C.2, we provide a proof for this result for the $r = 4$ and $\delta = 1/2$ case. Cases for larger r and smaller δ follow similarly with better constants.

Combining the two cases, $S(i,j,w_0,w_2,w_3) \le \exp(-O(N/\log N))$ for high-weight messages. For the entire probability in Equation (13), the ranges of w_0, w_2, w_3, i, j in the summation are all polynomial in N, thus the total probability remains $\exp(-O(N/\log N))$, which completes the proof of Theorem 3.

Message length n	Rate $1/r$	Scheme	Distance δ	Failure probability
2^{15}	1/4	Blaze	0.2	2^{-7}
		Ours	0.5	2^{-11}
	1/6	Blaze	0.2	2^{-17}
		Ours	0.5	2^{-31}
	1/8	Blaze	0.2	2^{-26}
		Ours	0.5	2^{-52}
2^{20}	1/4	Blaze	0.2	2^{-10}
		Ours	0.5	2^{-20}
	1/6	Blaze	0.2	2^{-25}
		Ours	0.5	2^{-49}
	1/8	Blaze	0.2	2^{-39}
		Ours	0.5	2^{-80}
2^{25}	1/4	Blaze	0.2	2^{-13}
		Ours	0.5	2^{-28}
	1/6	Blaze	0.2	2^{-33}
		Ours	0.5	2^{-67}
	1/8	Blaze	0.2	2^{-53}
		Ours	0.5	2^{-106}

Table 3: Distance and failure probability of our analysis and Blaze [BCF+25].

5 Evaluations

5.1 Distance of RAA Codes

Settings. In this section, we present the empirical evaluations of the relative distance and the failure probability of RAA codes. We implemented Sage scripts to calculate our upper bounds with different parameters, as well as the upper bounds in [BCF+25] for the binary case. The scripts were developed based on the implementations of [BFK+24] and [BCF+25].

For our bounds, we focus on the low-weight case in Section 3, as the failure probability is negligible for high-weight messages. The tables report numbers from a script using the upper bounds in Equation (6)+Equation (20), without the $O(\frac{1}{q-1})$ terms. For Blaze, we drop the negligible term as well (Stage 2 in their case) and implement [BCF+25, Equation (9)] as the upper bound with $\gamma = 0.36$ as in their implementation.

Comparison with Blaze. The empirical results are shown in Table 3 for different values of n and r. For the smallest instance of $n = 2^{15}$ and $r = 4$, the distance of RAA codes over the binary field is only 0.2, with a failure probability of 2^{-7}. With our analysis, the distance increases to 0.5 with a failure probability of 2^{-11}. The small instance of $n = 2^{15}$ already appears in some applications. For example, for polynomial commitments based on error-correcting codes [GLS+23], the message length is only square-root in the size of the polynomial, see Section 5.2.

Message length n	Rate $1/r$	Failure probability		
		$\delta = 0.3$	$\delta = 0.4$	$\delta = 0.5$
2^{15}	$1/4$	2^{-19}	2^{-17}	2^{-11}
	$1/6$	2^{-34}	2^{-33}	2^{-31}
2^{20}	$1/4$	2^{-22}	2^{-21}	2^{-20}
	$1/6$	2^{-53}	2^{-51}	2^{-49}
2^{25}	$1/4$	2^{-28}	2^{-30}	2^{-32}
	$1/6$	2^{-67}	2^{-69}	2^{-72}

Table 4: δ and failure probability of our analysis with different parameters.

When $r = 6$, our failure probability drops significantly to 2^{-31}, which is already considered acceptable in the literature, as this failure probability is statistical in the sampling phase of the RAA code and is not based on any cryptographic assumptions. By contrast, the failure probability of Blaze is 2^{-17}. Moreover, the failure probabilities decay dramatically when n becomes larger, which is expected as they are inverse polynomial in n. We note that the failure probability of Blaze in the table is without the additional testing on all messages of weight 1 or 2, thus the probability is higher than those reported in [BCF+25]. The additional testing can be applied to reduce the failure probability of both cases[2].

Compared with other related works, the analysis in [AHKM25] only achieves the relative distance $\delta = 0.05$ with the failure probability of 2^{-21} for $n = 2^{15}$ and $r = 4$ over a large finite field. The analysis in [BFK+24] on the EA code achieves the relative distance of $\delta = 0.1$ with the failure probability of 2^{-100} for $n = 2^{10}$ and rate $\frac{1}{2}$. However, this distance is based on a conjecture [BFK+24, Conjecture 4.6], and the encoding time of EA code is $O(n \log n)$.

Overall, our experiments demonstrate that our provable distance is significantly better than those in prior works. In particular, we show that the distance of RAA codes over a large finite field is better than that of the binary field, with a much smaller failure probability.

Additional results. In Table 4, we provide additional results for the distance and failure probability of our analysis. This is achieved by tweaking the constant c in Equation (20) and Equation (6). Recall that the failure probability of Stage 2 in Equation (20) increases as c decreases, while the probability of Stage 1 in Equation (6) decreases. For a target provable distance δ, we adjust c to balance the failure probabilities of the two stages. As shown in the table, for the same parameters of n and r, we can achieve smaller provable distances with smaller failure probabilities.

[2] For RAA codes over $\mathbb{F}_q$, one does not need to try all messages with different values. It suffices to test the positions of nonzeros and see if the weight remains high after accumulation, using the IOWE of the accumulator matrix.

Scheme	Distance δ	Proof Size	
		$m = 2^{20}$	$m = 2^{30}$
RAA over $\mathbb{F}_{2^\ell}$ [BCF+25]	0.2	1.26 MB	40 MB
RAA over $\mathbb{F}_q$ [AHKM25]	0.05	5.16 MB	165 MB
EA over $\mathbb{F}_q$ [BFK+24]	0.1	2.56 MB	82 MB
Ours	0.5	0.48 MB	15 MB

Table 5: Proof size of PCS with different codes and distances.

5.2 Proof Size of PCS based on RAA Code

To further demonstrate the advantage of our RAA code analysis, we evaluate the performance of a polynomial commitment scheme (PCS) using the RAA code. The construction of the PCS in [GLS+23] is presented in Protocol 1 of Appendix D. The RAA code over a large finite field can be used as a drop-in replacement in existing implementations such as that of [BFK+24] based on the EA code. The superior encoding time of the RAA code, and thus the good prover time of the PCS, has already been demonstrated in [BCF+25]. Therefore, we focus on the proof size of the PCS. We rely on the following theorem to calculate the proof size:

Theorem 5 ([GLS+23]). *For a polynomial of size m, Protocol 1 is a PCS with soundness error $\frac{m}{q} + (1 - \frac{\delta}{3})^t + (1 - \frac{2\delta}{3})^t$, where δ is the relative distance of the code, and t is the number of opened columns in Protocol 1.*

With this theorem, to achieve the soundness error of $2^{-\lambda}$, it suffices to set $t = \lambda \cdot \frac{1}{-\log_2(1-\delta/3)}$ (assuming that q is large enough). Therefore, larger δ leads to smaller t, and thus smaller proof size of the PCS. In our evaluation, we set $\lambda = 128$, and set the dimensions of the matrix C representing the coefficients of the polynomial as $\sqrt{m} \times \sqrt{m}$. The proof consists of t columns and 2 messages computed as linear combinations of the rows, i.e., $(t+2)\sqrt{m}$ in total. We note that it is possible to further optimize the proof size by changing the dimensions of the coefficient matrix [GLS+23], but to clearly show the relationship between the distance and the proof size, we do not apply this optimization.

Table 5 shows the proof size of the PCS using different codes and distances. "Scheme" refers to the distance achieved by the analysis in the corresponding paper. As the analysis of [BCF+25] is for the binary field, the PCS has to be constructed over the binary extension field $\mathbb{F}_{2^\ell}$, over which the RAA code obviously inherits the distance of the base field. Our analysis applies to both $\mathbb{F}_q$ and $\mathbb{F}_{2^\ell}$ as we do not utilize any other property of the field than its size. As shown in the table, with our better distance, the proof size is improved by 2.6× compared to the PCS based on the provable distance of RAA codes in [BCF+25]. It is 5.4× smaller than the proof size of the PCS in [BFK+24] using the EA code, and 11× smaller than that using the distance achieved in [AHKM25]. As PCS is a major building block of many zkSNARK schemes, our improvement directly leads to smaller proof size of zkSNARKs using RAA codes.

Acknowledgments

This work is supported by the National Science Foundation (NSF) under Grant No. 2401481. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these institutes.

References

AHKM25. Kasra Abbaszadeh, Hossein Hafezi, Jonathan Katz, and Sarah Meiklejohn. Single-server private outsourcing of zk-snarks. *Cryptology ePrint Archive*, 2025.

AR09. Alexandre Graell I Amat and Eirik Rosnes. Good concatenated code ensembles for the binary erasure channel. *IEEE Journal on Selected Areas in Communications*, 27(6):928–943, 2009.

AR11. 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 1–5. IEEE, 2011.

BCF+25. Martijn Brehm, Binyi Chen, Ben Fisch, Nicolas Resch, Ron D Rothblum, and Hadas Zeilberger. Blaze: Fast snarks from interleaved raa codes. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 123–152. Springer, 2025.

BCG+22. Elette Boyle, Geoffroy Couteau, Niv Gilboa, Yuval Ishai, Lisa Kohl, Nicolas Resch, and Peter Scholl. Correlated pseudorandomness from expand-accumulate codes. In *Annual International Cryptology Conference*, pages 603–633. Springer, 2022.

BFK+24. Alexander R Block, Zhiyong Fang, Jonathan Katz, Justin Thaler, Hendrik Waldner, and Yupeng Zhang. Field-agnostic snarks from expand-accumulate codes. In *Annual International Cryptology Conference*, pages 276–307. Springer, 2024.

BFMM26. 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.

BFRW25. Benedikt Bünz, Giacomo Fenzi, Ron D Rothblum, and William Wang. Tensorswitch: Nearly optimal polynomial commitments from tensor codes. *Cryptology ePrint Archive*, 2025.

BMMS25. Anubhav Baweja, Pratyush Mishra, Tushar Mopuri, and Matan Shtepel. Query-optimal iopps for linear-time encodable codes. *Cryptology ePrint Archive*, 2025.

BMS08. Louay Bazzi, Mohammad Mahdian, and Daniel A Spielman. The minimum distance of turbo-like codes. *IEEE Transactions on Information Theory*, 55(1):6–15, 2008.

CBBZ22. Binyi Chen, Benedikt Bünz, Dan Boneh, and Zhenfei Zhang. Hyperplonk: Plonk with linear-time prover and high-degree custom gates. *Cryptology ePrint Archive*, 2022.

DJM98. Dariush Divsalar, Hui Jin, and Robert J McEliece. Coding theorems for" turbo-like" codes. In *Proceedings of the annual Allerton Conference on Communication control and Computing*, volume 36, pages 201–210. University Of Illinois, 1998.

DP97. D Divsalar and F Pollara. Serial and hybrid concatenated codes with applications. In *Proc. Int. Symp. on Turbo Codes and Related Topics*, pages 80–87, 1997.

FR08. F. Fagnani and C. Ravazzi. Spectra and minimum distances of repeat multiple accumulate codes. *IEEE Transactions on Information Theory*, 2008.

GLS+23. Alexander Golovnev, Jonathan Lee, Srinath Setty, Justin Thaler, and Riad S Wahby. Brakedown: Linear-time and field-agnostic snarks for R1CS.

In *Annual International Cryptology Conference*, pages 193–226. Springer, 2023.

GWC19. Ariel Gabizon, Zachary J Williamson, and Oana Ciobotaru. Plonk: Permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. *Cryptology ePrint Archive*, 2019.

HGQZ26. 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.

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, 2011.

Kha26. Majid Khabbazian. Asymptotically optimal distance-tail bounds for large-field RAA codes. Cryptology ePrint Archive, Paper 2026/1075, 2026.

KZC07. J. Kliewer, K. S. Zigangirov, and Jr. D. J. Costello. New results on the minimum distance of repeat multiple accumulate codes. In *Proceedings of the 45th Annual Allerton Conference on Communication, Control, and Computing*, 2007.

KZKC08. J. Kliewer, K. S. Zigangirov, C. Koller, and Jr. D. J. Costello. Coding theorems for repeat multiple accumulate codes. *IEEE Transactions on Information Theory*, 2008.

Mer87. Ralph C Merkle. A digital signature based on a conventional encryption function. In *Conference on the theory and application of cryptographic techniques*, 1987.

PS03. H. D. Pfister and P. H. Siegel. The serial concatenation of rate-1 codes through uniform random interleavers. *IEEE Transactions on Information Theory*, 49(6):1425–1438, 2003.

RA11. Eirik Rosnes and Alexandre Graell i Amat. On the analysis of weighted nonbinary repeat multiple-accumulate codes. *arXiv preprint arXiv:1101.5966*, 2011.

RRT23. Srinivasan Raghuraman, Peter Rindal, and Titouan Tanguy. Expand-convolute codes for pseudorandom correlation generators from lpn. *Cryptology ePrint Archive*, 2023.

Sch80. Jacob T Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM (JACM)*, 27(4):701–717, 1980.

Set20. Srinath Setty. Spartan: Efficient and general-purpose zkSNARKs without trusted setup. In *Advances in Cryptology – CRYPTO 2020*, pages 704–737, 2020.

Spi95. Daniel A Spielman. Linear-time encodable and decodable error-correcting codes. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 388–397, 1995.

Tha22. Justin Thaler. Proofs, arguments, and zero-knowledge. *Foundations and Trends® in Privacy and Security*, 4(2–4):117–660, 2022.

XZZ+19. Tiacheng Xie, Jiaheng Zhang, Yupeng Zhang, Charalampos Papamanthou, and Dawn Song. Libra: Succinct zero-knowledge proofs with optimal prover computation. In *Annual International Cryptology Conference*, pages 733–764. Springer, 2019.

Yan04. Kyeongcheol Yang. Weighted nonbinary repeat-accumulate codes. *IEEE Transactions on Information Theory*, 50(3):527–531, 2004.

Zip79. Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation*, 1979.

A RAA over the Binary Field

A.1 Amplification of RA over the Binary Field

In this section, we want to prove that $w_2 > cw_0 \log n$ for some constant c with high probability over $\mathbb{F}_2$ for low-weight messages (i.e. when $w_0 = O(n/\log^2 n)$). In other words, we want to show the following probability is small:

$$\begin{aligned}
&\Pr_{\mathbf{\Pi}}[\exists \mathbf{x} \neq 0 \text{ s.t. } \mathsf{wt}(\mathbf{xR\Pi A}) \leq cw_0 \log n | \mathsf{wt}(\mathbf{x}) \leq c'n/\log^2 n] \\
&\quad = \sum_{w_0=1}^{c'n/\log^2 n} \binom{n}{w_0} \Pr_{\mathbf{\Pi}}[\mathsf{wt}(\mathbf{xR\Pi A}) \leq cw_0 \log n | \mathsf{wt}(\mathbf{x}) = w_0] \\
&\quad \leq \sum_{w_0=1}^{c'n/\log^2 n} \binom{n}{w_0} \sum_{w_2=0}^{cw_0 \log n} \Pr_{\mathbf{\Pi}}[\mathsf{wt}(\mathbf{xR\Pi A}) = w_2 | \mathsf{wt}(\mathbf{x}) = w_0] \\
&\quad = \sum_{w_0=1}^{c'n/\log^2 n} \binom{n}{w_0} \sum_{w_2=0}^{cw_0 \log n} \frac{\binom{w_2-1}{\lceil \frac{rw_0}{2} \rceil - 1}\binom{rn-w_2}{\lfloor \frac{rw_0}{2} \rfloor}}{\binom{rn}{rw_0}}
\end{aligned}$$

where the third line follows from the union bound and the last line follows from uniformity of $\mathbf{\Pi}$.

For simplicity, we will assume that $r = 2k$ for some integer k. We will approximate the innermost sum using Stirling's Approximation $(\frac{n}{k})^k \leq \binom{n}{k} \leq (\frac{en}{k})^k$.

$$\begin{aligned}
&\leq \sum_{w_2=kw_0}^{cw_0 \log n} \left(\frac{e(w_2-1)}{kw_0-1}\right)^{kw_0-1} \left(\frac{e(rn-w_2)}{kw_0}\right)^{kw_0} \left(\frac{rw_0}{rn}\right)^{rw_0} \\
&\leq \sum_{w_2=kw_0}^{cw_0 \log n} \left(\frac{e^2(w_2-1)(rn-w_2)}{kw_0(kw_0-1)}\right)^{kw_0} \left(\frac{w_0}{n}\right)^{2kw_0} \\
&= \left(\frac{e^2 w_0}{n^2 k(kw_0-1)}\right)^{kw_0} \sum_{w_2=kw_0}^{cw_0 \log n} ((rn-w_2)(w_2-1))^{kw_0}
\end{aligned}$$

Now we will compute the maximum of this sum given that $w_0 \in (c, n/\log^2 n)$. Notice that the summand is a quadratic of the form $f(w_2) = -w_2^2 + (rn+1)w_2 - rn$, so the maximum is at $w_2 = \frac{rn+1}{2}$. But recall that $w_2 \in [kw_0, cw_0 \log n]$. If we assume that $w_0 \leq \frac{r}{2c}\frac{n}{logn}$, then we can say that the maximum is at $w_2 = cw_0 \log n$. Realistically this will hold for most values of c' since $c' = O(1)$ and

we just need $c' \leq \frac{r}{2c} \log n$. So, we continue with this assumption:

$$\begin{aligned}
&\leq \left(\frac{e^2 w_0}{n^2 k(kw_0 - 1)}\right)^{kw_0} \left((rn - cw_0 \log n)(cw_0 \log n - 1)\right)^{kw_0} \cdot \\
&\qquad\qquad (cw_0 \log n - kw_0 + 1) \\
&\leq \left(\frac{e^2 w_0}{n^2 k(kw_0 - 1)}\right)^{kw_0} (rn)^{kw_0} (cw_0 \log n)^{kw_0} (cw_0 \log n) \\
&\leq \left(\frac{2e^2 cw_0^2}{(kw_0 - 1)}\right)^{kw_0} cw_0 (\log n)^{kw_0+1} n^{-kw_0}
\end{aligned}$$

Multiplying by the binomial term $\binom{n}{w_0}$ in the original sum, we have that the main summand is:

$$\begin{aligned}
&\leq cw_0 \left(\frac{en}{w_0}\right)^{w_0} \left(\frac{2e^2 cw_0^2}{(kw_0 - 1)}\right)^{kw_0} (\log n)^{kw_0+1} n^{-kw_0} \\
&\leq cw_0 \left(\frac{e}{w_0}\right)^{w_0} \left(\frac{2e^2 cw_0^2}{(kw_0 - 1)}\right)^{kw_0} (\log n)^{kw_0+1} n^{w_0(1-k)}
\end{aligned}$$

Finally, to upper-bound the entire sum, we can first show that the above term is decreasing in w_0. To do this, we will just compute the ratio between consecutive terms.

$$\begin{aligned}
\frac{T(w_0+1)}{T(w_0)} &= \frac{w_0 + 1}{w_0} \frac{ew_0^{w_0}}{(w_0+1)^{w_0+1}} \left(\frac{2e^2 c(w_0+1)^2}{kw_0 + k - 1}\right)^k \cdot \\
&\qquad \left[\left(\frac{w_0+1}{w_0}\right)^2 \frac{kw_0 - 1}{kw_0 + k - 1}\right]^{kw_0} (\log n)^k n^{1-k} \\
&\leq \frac{ew_0^{w_0}}{w_0(w_0+1)^{w_0}} \left(\frac{2e^2 c(w_0+1)^2}{kw_0 + k - 1}\right)^k \left(\frac{w_0+1}{w_0}\right)^{2kw_0} (\log n)^k n^{1-k} \\
&\leq \frac{e}{w_0} e^{2k-1} \left(\frac{2e^2 c(w_0+1)^2}{kw_0}\right)^k (\log n)^k n^{1-k}
\end{aligned}$$

where the last line follows from the identity $(w_0 + 1/w_0)^{w_0} < e$. We will apply one more loose bound to make the equation more legible: $w_0 + 1 \leq 2w_0$, so $(w_0+1)^2 \leq 4w_0^2$:

$$\begin{aligned}
&\leq \frac{e}{w_0} e^{2k-1} \left(\frac{8e^2 cw_0}{k}\right)^k (\log n)^k n^{1-k} \\
&= e^{2k} \left(\frac{8e^2 c}{k}\right)^k w_0^{k-1} (\log n)^k n^{1-k}
\end{aligned}$$

For large enough n, clearly the n^{1-k} will dominate and make this ratio < 1, so our terms are decreasing in w_0.

Going back to our main sum, this means the summand is maximized at $w_0 = 1$:

$$cw_0 \left(\frac{e}{w_0}\right)^{w_0} \left(\frac{2e^2 c w_0^2}{(kw_0 - 1)}\right)^{kw_0} (\log n)^{kw_0+1} n^{w_0(1-k)}$$
$$\leq ce \left(\frac{2e^2 c}{k-1}\right)^{k} \log^{k+1}(n) \frac{1}{n^{k-1}}$$

Since we have the factor of $n^{w_0(1-k)}$, terms from $w_0 = 2$ onward will be asymptotically dominated by the first term, so we can say that

$$\Pr_{\mathbf{\Pi}}[\exists \mathbf{x} \neq 0 \text{ s.t. } \mathsf{wt}(\mathbf{xR\Pi A}) \leq cw_0 \log n | w_0 \leq c'n/\log^2 n]$$
$$= O\left(\frac{\log^{k+1}(n)}{n^{k-1}}\right) = \tilde{O}(n^{1-\frac{r}{2}}).$$

A.2 Comparison with the Binary Field for Stage 2

In this section, we would like to show that the provable failure probability of a uniformly sampled RAA code over $\mathbb{F}_q$ is smaller than the provable failure probability over $\mathbb{F}_2$. Recall that for $\mathbb{F}_2$, $N \in \mathbb{N}$ and $w_2, w_3 \in [N]$, we have

$$A^{N,2}_{w_2,w_3} = \binom{w_3 - 1}{\lceil \frac{w_2}{2} \rceil - 1} \binom{N - w_3}{\lfloor \frac{w_2}{2} \rfloor}.$$

Similar to Lemma 5, for all messages $\mathbf{x}$ with weight $0 < w_0 \leq \frac{n}{c' \log^2 n}$, suppose the weight $\mathsf{wt}(\mathbf{xRP_1A}) = w_2 > cw_0 \log N$, then by the union bound,

$$\Pr_{\mathbf{P}_2}[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{xRP}_1\mathbf{AP}_2\mathbf{A}) \leq \delta N]$$
$$\leq \sum_{w_3=1}^{\delta N} \Pr_{\mathbf{P}_2}[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{xRP}_1\mathbf{AP}_2\mathbf{A}) = w_3]$$
$$\leq \sum_{w_0=1}^{n/c' \log^2 n} \binom{n}{w_0} \sum_{w_2 = cw_0 \log N + 1}^{2\delta N} \sum_{w_3=1}^{\delta N} \frac{A^{N,2}_{rw_0,w_2}}{\binom{N}{rw_0}} \frac{A^{N,2}_{w_2,w_3}}{\binom{N}{w_2}},$$

Where $\mathbf{P}_2$ is just a permutation matrix in the binary case. As the ranges of w_0, w_2 are the same, we drop the summations and $\binom{n}{w_0}$ and focus on comparing

$$\sum_{w_3=1}^{\delta N} \sum_{i=w_2-2w_0-1}^{w_2-1} \frac{\binom{w_2-1}{\lceil \frac{w_1-i}{2} \rceil - 1} \binom{N-w_3}{\lfloor \frac{w_2-i}{2} \rfloor} \binom{w_2 - \lceil \frac{w_1-i}{2} \rceil}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\lceil \frac{w_2-i}{2} \rceil - 1} \binom{N-w_3}{\lfloor \frac{w_2-i}{2} \rfloor} \binom{w_3 - \lceil \frac{w_2-i}{2} \rceil}{i}}{\binom{N}{w_2}} \tag{15}$$

for the finite field case derived from the proof of Lemma 5 with

$$\sum_{w_3=1}^{\delta N} \frac{\binom{w_2-1}{\lceil \frac{w_1}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{w_1}{2} \rfloor}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\lceil \frac{w_2}{2} \rceil - 1}\binom{N-w_3}{\lfloor \frac{w_2}{2} \rfloor}}{\binom{N}{w_2}} \tag{16}$$

derived similarly for the binary case. Recall that r is chosen to be even, so $\lceil w_1/2 \rceil = \lfloor w_1/2 \rfloor = rw_0/2$.

As shown in the proof of Lemma 5, Equation (15) is

$$\begin{aligned} &\leq w_0(2w_0+1)^2(w_2-2w_0)(er)^{rw_0}\left(\frac{eN}{rw_0}\right)^{w_0}\left(\frac{e\delta N}{w_0}\right)^{w_0}\left(\frac{e(1-\delta)N}{w_0}\right)^{w_0}\delta^{w_2} \\ &\leq \frac{ce^{(r+3)w_0}r^{(r-1)w_0}(2+1/w_0)}{w_0^{3w_0-3}}(1-\delta)^{w_0}N^{w_0(3-c\log(1/\delta))}\log N \end{aligned} \tag{17}$$

when $cw_0 \log N < w_2 \leq \delta N/2$. The probability is negligible when $w_2 > \delta N/2$ and we omit that case for simplicity.

As shown in the proof of [BFK+24, Lemma 3.6], Equation (16) is maximized at $w_3 = \delta N$ for $\delta < 1/2$ as well, and is upper-bounded by

$$\begin{aligned} &\leq 1 \cdot \left(4e^2\delta(1-\delta)\right)^{\lfloor w_2/2 \rfloor} \cdot e^2\delta N \\ &\leq e^2\delta N^{1+\frac{cw_0\log(4e^2\delta(1-\delta))}{2}}. \end{aligned} \tag{18}$$

In order for the probability to be less than 1, $4e^2\delta(1-\delta) < 1$ and thus the provable distance δ must be significantly smaller than $1/2$. We upper-bound the term $\frac{A^{N,2}_{rw_0,w_2}}{\binom{N}{rw_0}}$ by 1 in this bound since w_2 can range to $2\delta N$, but under the stricter condition that $\delta < 1/4$, we can achieve a slightly better bound as below:

$$\begin{aligned} &\leq \left(4e^2\delta(1-\delta)\right)^{rw_0/2} \cdot e^2\delta N \cdot \left(4e^2\delta(1-\delta)\right)^{\lfloor w_2/2 \rfloor} \cdot e^2\delta N \\ &\leq e^4\delta^2 N^{2+\frac{cw_0}{2}\log(4e^2\delta(1-\delta))} \cdot \left(4e^2\delta(1-\delta)\right)^{rw_0/2}. \end{aligned} \tag{19}$$

However, this bound is the same asymptotically for the dominating $w_0 = O(1)$ terms. Moreover, the upper bound in Equation (17) above is significantly smaller than the upper bound in Equation (19). To see this, we start with $w_0 = 1$, which is the dominating term in both cases. Equation (17) is $\tilde{O}(N^{3+c\log\delta})$, while Equation (18) is $O(N^{1+\frac{c\log(4e^2\delta(1-\delta))}{2}})$. As $\delta < 1/2$, $\delta < 1-\delta$, thus $N^{1+\frac{c\log(4e^2\delta(1-\delta))}{2}} > N^{1+c(1+\log e+\log\delta)} > N^{3+c\log\delta}$ for any $c \geq 1$. In fact, the inequalities are quite loose. For example, when $\delta = 0.02$ and $c = 6$, the bound of the binary case is only $O(N^{-1.37})$, while the bound of the finite field case is $\tilde{O}(N^{-30.8})$.

For $w_0 > 1$, observe that Equation (17) decays much faster with w_0 than Equation (18). Therefore, the provable upper bound of the failure probability of Stage 2 over a large finite field is significantly smaller than the provable upper bound over the binary field.

B Deferred Lemmas and Proofs for the Low-Weight Case

B.1 Proof of Lemma 5

We will prove two auxiliary lemmas similar to the ones in Section 3.1 that will allow us to upper-bound our summand.

Lemma 8. *In the range $w_2 \in [rw_0/2, 2\delta N]$, the term $S(i, rw_0, w_2)$ is maximized at*

$$w_2^* = \min\left\{2\delta N, \left\lceil \frac{N(i + \lceil \frac{rw_0 - i}{2} \rceil - 1)}{rw_0 - 1} \right\rceil\right\}$$

Proof. We will take the ratio between consecutive values of w_2.

$$\begin{aligned}\frac{S(i, rw_0, w_2+1)}{S(i, rw_0, w_2)} &= \frac{\binom{w_2}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2-1}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2+1-\lceil \frac{rw_0-i}{2} \rceil}{i}}{\binom{w_2-1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N-w_2}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2-\lceil \frac{rw_0-i}{2} \rceil}{i}} \\ &= \frac{w_2(N - w_2 - \lfloor \frac{rw_0-i}{2} \rfloor)}{(N-w_2)(w_2 - \lceil \frac{rw_0-i}{2} \rceil - i + 1)}\end{aligned}$$

For $S(w_2+1) \geq S(w_2)$, we need $w_2 \leq \frac{N(i+\lceil \frac{rw_0-i}{2} \rceil - 1)}{rw_0-1}$. Since $w_2 \leq 2\delta N$, we have our result.

After maximizing over w_2, it suffices to find the maximum of the following over i:

$$\begin{aligned}&\binom{w_2^\star(i) - 1}{\lceil \frac{rw_0-i}{2} \rceil - 1}\binom{N - w_2^\star(i)}{\lfloor \frac{rw_0-i}{2} \rfloor}\binom{w_2^\star(i) - \lceil \frac{rw_0-i}{2} \rceil}{i} \\ &w_2^*(i) = \min\left\{N, 2\delta N, \left\lceil \frac{N(i + \lceil \frac{rw_0 - i}{2} \rceil - 1)}{rw_0 - 1} \right\rceil\right\}\end{aligned}$$

We prove this in the following lemma.

Lemma 9. *For large N, $r \geq 4$ and $w_0 = O(N/\log^2 N)$, over the range $i \in [(r-2)w_0 - 1, rw_0 - 1]$, the maximum occurs at $S((r-2)w_0 - 1, rw_0, w_2^*)$.*

The full proof for the lemma can be found in Appendix Section B.3.

With these two lemmas, we are ready to prove Lemma 5:

Proof. Starting from our bound on $\Pr_{\mathbf{P_1},\mathbf{P_2}}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3]$ for fixed w_0, w_2, w_3, by the union bound on w_0, w_2, w_3, we have

$$\Pr_{\mathbf{P_1},\mathbf{P}_2}[\exists \mathbf{x} \text{ s.t. } \mathsf{wt}(\mathbf{xRP}_1\mathbf{AP}_2\mathbf{A}) \leq \delta N | w_2 > cw_0 \log N]$$

$$\leq \sum_{w_0=1}^{n/c' \log^2 n} \sum_{w_2=cw_0 \log N+1}^{N} \sum_{w_3=1}^{\delta N} \Pr_{\mathbf{P_1},\mathbf{P}_2}[\mathsf{wt}(\mathbf{x}''\mathbf{A}) = w_3]$$

$$\leq \sum_{w_0=1}^{n/c' \log^2 n} \sum_{w_2=cw_0 \log N+1}^{N} \sum_{w_3=1}^{\delta N}$$

$$\left[\sum_{i=(r-2)w_0-1}^{rw_0-1} \sum_{j=w_2-2w_0-1}^{w_2-1} \binom{n}{w_0} \frac{\binom{w_2-1}{\phi-1}\binom{N-w_2}{\phi'}\binom{w_2-\phi}{i}}{\binom{N}{w_1}} \frac{\binom{w_3-1}{\tau-1}\binom{N-w_3}{\tau'}\binom{w_3-\tau}{j}}{\binom{N}{w_2}} + O\left(\frac{1}{q-1}\right) \right]$$

We focus on the binomial coefficient terms. As discussed in Section 2.3, in order for the numerator to be larger than 0, $j \leq 2w_3 - w_2 + 1$, while $j \geq w_2 - 2w_0 - 1$ in the summation. Therefore, the summand is 0 for $w_2 > w_3 + w_0 + 1$. Again, w_2 can be as big as $2w_3$, but those cases happen for $j < w_2 - 2w_0 - 1$ and are in the $O(\frac{1}{q-1})$ terms.

By Lemma 9, we can take $j = w_2 - 2w_0 - 1$ as the maximizer. In Appendix B.2, we prove a similar lemma to Lemma 3 to show that we can take $w_3 = \delta N$ as a maximizer as well. We now have the following:

$$\leq \sum_{w_0=1}^{n/c' \log^2 n} \sum_{w_2=cw_0 \log N+1}^{2\delta N} \delta N \cdot \left[(2w_0+1)^2 \binom{n}{w_0} \cdot \frac{\binom{w_2-1}{w_0}\binom{N-w_2}{w_0}\binom{w_2-w_0-1}{(r-2)w_0-1}}{\binom{N}{w_1}} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0-1}{w_2-2w_0-1}}{\binom{N}{w_2}} \right]$$

We will focus on the summand and use Stirling's inequality.

$$\delta N(2w_0+1)^2 \binom{n}{w_0} \frac{\binom{w_2-1}{w_0}\binom{N-w_2}{w_0}\binom{w_2-w_0-1}{(r-2)w_0-1}}{\binom{N}{w_1}} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0-1}{w_2-2w_0-1}}{\binom{N}{w_2}}$$

$$\leq \delta N(2w_0+1)^2 \binom{n}{w_0} \left(\frac{ew_2}{w_0}\right)^{w_0} \left(\frac{e(N-w_2)}{w_0}\right)^{w_0} \frac{\binom{w_2}{(r-2)w_0-1}}{\binom{N}{rw_0}} (w_2-2w_0) \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$

The last line follows from $r \geq 4$ and $w_2 > (r-2)w_0$, which holds in our summation range as we assumed that $w_2 > cw_0 \log N$. Now, since $w_2 \leq N$, we

have

$$\frac{\binom{w_2}{(r-2)w_0-1}}{\binom{N}{rw_0}} = \frac{(rw_0)!}{((r-2)w_0-1)!\prod_{i=0}^{2w_0-1}(N-(r-2)w_0+1-i)} \prod_{i=0}^{(r-2)w_0} \frac{w_2-i}{N-i}$$
$$\leq \left(\frac{w_2}{N}\right)^{(r-2)w_0-1} \left(\frac{rw_0}{N-rw_0}\right)^{2w_0+1}$$

Continuing with our analysis, we will consider the case where $w_2 = 2\delta N$ is the maximizer.

$$\leq \delta N(2w_0+1)^2 \left(\frac{e2\delta N}{w_0}\right)^{w_0} \left(\frac{eN}{w_0}\right)^{w_0} \left(\frac{2\delta N}{N}\right)^{(r-2)w_0-1} \left(\frac{rw_0}{N-rw_0}\right)^{2w_0+1}$$
$$\cdot (w_2-2w_0)\binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0-1}{w_2-2w_0-1}}{\binom{N}{w_2}}$$
$$\leq \delta(2w_0+1)^2(w_2-2w_0)e^{2w_0}(2\delta)^{(r-1)w_0-1} \left(\frac{r}{1-rw_0/N}\right)^{2w_0+1}$$
$$\cdot \binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$
$$\leq \delta(2w_0+1)^2(w_2-2w_0)r^{2w_0+1}e^{2w_0+1+2rw_0^2/(N-rw_0)}\binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$

We can also consider the other possible maximizer for w_2. Using the upper bound $w_2^* \leq \frac{N(r-1)w_0}{rw_0-1}$, we have

$$\leq \delta N(2w_0+1)^2(w_2-2w_0) \left(\frac{eN(r-1)w_0}{w_0(rw_0-1)}\right)^{w_0} \left(\frac{eN}{w_0}\right)^{w_0}$$
$$\cdot \left(\frac{(r-1)w_0}{rw_0-1}\right)^{(r-2)w_0-1} \left(\frac{rw_0}{N-rw_0}\right)^{2w_0+1} \binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$
$$\leq \delta(2w_0+1)^2(w_2-2w_0)e^{2w_0} \left(\frac{1}{w_0}\right)^{w_0-1} \left(\frac{rw_0}{1-rw_0/N}\right)^{2w_0+1}$$
$$\cdot \binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$
$$\leq \delta(2w_0+1)^2(w_2-2w_0)r^{2w_0+1}e^{2w_0+1+2rw_0^2/(N-rw_0)}\binom{n}{w_0} \frac{\binom{\delta N-1}{w_0}\binom{N-\delta N}{w_0}\binom{\delta N-w_0}{w_2-2w_0}}{\binom{N}{w_2}}$$

With both upper bounds over w_2, we arrive at the same bound and the analysis proceeds the same.

We have the two cases below:

Case 1: $w_2 \leq \delta N/2$.

In this case, it is easy to show that $\frac{\binom{\delta N - w_0}{w_2 - 2w_0}}{\binom{N}{w_2}} \leq \delta^{w_2}$. First, notice that $\binom{\delta N - w_0}{w_2 - 2w_0} \leq \binom{\delta N}{w_2}$ for $w_2 \leq \delta N/2$. It follows that

$$\frac{\binom{\delta N - w_0}{w_2 - 2w_0}}{\binom{N}{w_2}} \leq \frac{\binom{\delta N}{w_2}}{\binom{N}{w_2}} = \prod_{i=0}^{w_2 - 1} \frac{\delta - i/N}{1 - i/N} \leq \delta^{w_2}$$

Using this and Stirling's Approximation $\left(\frac{N}{w}\right)^w \leq \binom{N}{w} \leq \left(\frac{eN}{w}\right)^w$, we can upper-bound our summand by the following:

$$\leq \delta(2w_0+1)^2(w_2 - 2w_0) r^{2w_0+1} e^{2w_0+1+2rw_0^2/(N-rw_0)} \left(\frac{eN}{rw_0}\right)^{w_0} \left(\frac{e\delta N}{w_0}\right)^{w_0} \left(\frac{e(1-\delta)N}{w_0}\right)^{w_0} \delta^{w_2}.$$

As $w_2 > cw_0 \log N$, the equation above is

$$\leq \delta re(2w_0+1)^2(cw_0 \log N - 2w_0) \left(\frac{e^{5+2rw_0/(N-rw_0)} r}{w_0^3}\right)^{w_0} (1-\delta)^{w_0} \delta^{cw_0 \log N} N^{3w_0}$$

$$\leq \frac{\delta c e^{(5+2rw_0/(N-rw_0))w_0+1} r^{w_0+1} (2+1/w_0)^2}{w_0^{3w_0-3}} (1-\delta)^{w_0} N^{w_0(3 - c\log(1/\delta))} \log N. \tag{20}$$

For example, by setting $c \geq \frac{6}{-\log \delta}$, the probability above is $\leq N^{-3w_0} \log N$. Summing over $w_2 \in [cw_0 \log N + 1, \delta N/2]$ introduces at most a factor of N, thus the probability is still $\leq N^{1-3w_0} \log N$.

Finally, summing over all $w_0 \in [1, n/c' \log^2 n]$, the total probability is upper-bounded by $\tilde{O}(N^{-\eta_1})$ for $\eta_1 = -c \log \delta - 4$, where the dominating term is by $w_0 = 1$.

Case 2: $w_2 > \delta N/2$.

As explained above, $w_2 \leq w_3 + w_0 + 1 = \delta N + w_0 + 1$. In this case, as $\delta \leq \frac{1}{2}$ and $w_0 \leq \frac{n}{c' \log^2 n}$, $\frac{\binom{\delta N - w_0}{w_2 - 2w_0}}{\binom{N}{w_2}} \leq \frac{\binom{\delta N - w_0}{(\delta N - w_0)/2}}{\binom{N}{\delta N/2}} \leq \frac{\binom{\delta N}{\delta N/2}}{\binom{N}{\delta N/2}}$, where the numerator is taking the maximum of the combinatorial term at $\binom{\delta N}{\delta N/2}$, and the denominator is taking the minimum at $w_2 = \delta N/2$. By the same argument above, the ratio is $\leq \delta^{\delta N/2}$.

Using this and Stirling's Approximation, we can upper-bound the summand by the following:

$$\leq w_0(2w_0+1)^2(w_2 - 2w_0)(er)^{rw_0} \left(\frac{eN}{rw_0}\right)^{w_0} \left(\frac{e\delta N}{w_0}\right)^{w_0} \left(\frac{e(1-\delta)N}{w_0}\right)^{w_0} \delta^{\delta N/2}$$

$$\leq w_0(2w_0+1)^2(\delta N + 1 - w_0) \left(\frac{e^{r+3} r^{r-1}}{w_0^3}\right)^{w_0} (1-\delta)^{w_0} \delta^{\delta N/2} N^{3w_0}$$

$$\leq w_0(2w_0+1)^2(\delta N + 1) \frac{e^{(r+3)w_0} r^{(r-1)w_0}}{w_0^{3w_0}} (1-\delta)^{w_0} N^{3w_0 - \frac{\delta N \log(1/\delta)}{2 \log N}}$$

As $w_0 \leq \frac{n}{c' \log^2 n}$, the probability is $O(N^{\frac{\delta N \log(1/\delta)}{2 \log N}}) = O(\delta^{\delta N/2})$, which is negligible. This completes the proof together with Case 1. □

B.2 Decreasing Lemma for δN

Lemma 10. *Let $w_2 > 4w_0$. Then, the summand $S(i, w_2, \delta N)$ is decreasing with i on the interval $i \in [w_2 - 2w_0 - 1, w_2 - 1]$.*

Proof. We will split this into cases corresponding to the parity of $w_2 - i$.

Case 1 (odd $w_2 - i$). Let $w_2 - i = 2m + 1$, where m is an integer in the interval $[0, w_0]$. We will consider the ratio between consecutive terms over i and show that it is strictly less than 1. We can use the following two properties of the binomial coefficient: $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$, $\binom{n-1}{k} = \frac{n-k}{n}\binom{n}{k}$.

$$\frac{S(i+1, w_2, \delta N)}{S(i, w_2, \delta N)} = \frac{\binom{\delta N - 1}{m-1}\binom{N-\delta N}{m}\binom{\delta N - m}{i+1}}{\binom{\delta N - 1}{m}\binom{N-\delta N}{m}\binom{\delta N - m - 1}{i}} = \frac{m}{\delta N - m}\frac{\delta N - m}{i+1} = \frac{m}{w_2 - 2m}$$

In order to have $\frac{m}{w_2 - 2m} < 1$, we need $w_2 > 3m$, or $w_2 > 3w_0$.

Case 2 (even $w_2 - i$). Let $w_2 - i = 2m$, where m is now in the interval $[1, w_0]$. We will again compute the ratio between consecutive terms over i:

$$\begin{aligned}
\frac{S(i+1, w_2, \delta N)}{S(i, w_2, \delta N)} &= \frac{\binom{\delta N - 1}{m-1}\binom{N-\delta N}{m-1}\binom{\delta N - m}{i+1}}{\binom{\delta N - 1}{m-1}\binom{N-\delta N}{m}\binom{\delta N - m}{i}} \\
&= \frac{m}{N - \delta N - m + 1}\frac{\delta N - m - i}{i+1} \\
&\leq \frac{\delta N}{(1-\delta)N - w_0 + 1}\frac{w_0}{w_2 - 2w_0 + 1} \\
&\leq \frac{\delta N}{(1-\delta)N/2}\frac{w_0}{w_2 - 2w_0} \\
&= \frac{2\delta}{1-\delta}\frac{w_0}{w_2 - 2w_0},
\end{aligned}$$

where the second to last line follows from $w_0 \leq N/4$, which is the case since $r \geq 4$. Note that $\frac{2\delta}{1-\delta}$ is at most 2 since $\delta \leq 1/2$. If we assume that $w_2 > 4w_0$, then we can say that $w_2 - 2w_0 > 2w_0$, and we can conclude the following:

$$\frac{2\delta}{1-\delta}\frac{w_0}{w_2 - 2w_0} < 2 \cdot \frac{w_0}{2w_0} = 1.$$

Thus the ratio between consecutive terms in both cases is strictly less than 1, so $S(i, w_2, \delta N)$ is decreasing with i. □

B.3 Full Proof of Lemma 9

Lemma. *For large N, $r \geq 4$ and $w_0 = O(N/\log^2 N)$, over the range $i \in [(r-2)w_0 - 1, rw_0 - 1]$, the maximum occurs at $S((r-2)w_0 - 1, rw_0, w_2^\star)$.*

Proof. First, we will split into cases according to the parity of $rw_0 - i$ to deal with the floor and ceiling terms. We will show that the even values are always less than the consecutive odd value.

Even Case. $rw_0 - i$ is even, $= 2k$

$$\binom{w_2^\star(i)-1}{k-1}\binom{N-w_2^\star(i)}{k}\binom{w_2^\star(i)-k}{rw_0-2k} = \binom{rw_0-k-1}{k-1}\binom{N-w_2^\star(i)}{k}\binom{w_2^\star(i)-1}{rw_0-k-1}$$

Odd Case. $rw_0 - i$ is odd $= 2k+1$

$$\binom{w_2^\star(i)-1}{k}\binom{N-w_2^\star(i)}{k}\binom{w_2^\star(i)-k-1}{rw_0-2k-1} = \binom{rw_0-k-1}{k}\binom{N-w_2^\star(i)}{k}\binom{w_2^\star(i)-1}{rw_0-k-1}$$

Notice that $rw_0 - i = 2k, 2k+1$ have the same value of $w_2^*(i)$ because of the ceiling terms. In this form, it is clear to see that the ratio between odd term and previous even term is $\frac{rw_0-2k}{k} \geq r-2 > 1$, since $r \geq 4$ and $w_0 \geq k$. Now we know the maximum must occur at an odd case.

We will consider each possibility for w_2^* separately.

Upper Bound 1. $w_2^* = 2\delta N$

We can take the ratio between consecutive odd terms.

$$\frac{S(rw_0 - 2(k+1) - 1, rw_0, 2\delta N)}{S(rw_0 - 2k - 1, rw_0, 2\delta N)} = \frac{(rw_0-2k-1)(rw_0-2k-2)(N-2\delta N-k)}{(k+1)^2(2\delta N - rw_0 + k + 1)}$$

Notice that since we are looking at the next odd term, $k \leq w_0 - 1$, so $rw_0 \geq 4(k+1)$. Additionally, because $w_2^* = 2\delta N$, we know that $2\delta \leq \frac{i+\left\lceil \frac{rw_0-i}{2} - 1 \right\rceil}{rw_0 - 1}$ with $i = rw_0 - 2(k+1) - 1$. Also, the function $\frac{1-x}{x}$ is decreasing with x. Together with $w_0 = O(N/\log^2 N)$, this gives us the bound

$$\frac{N - 2\delta N - k}{2\delta N - rw_0 + k + 1} \geq (1 - o(1))\frac{k+1}{rw_0 - k - 2}$$

Now, continuing,

$$\begin{aligned}
&\geq (1-o(1))\,\frac{(rw_0-2k-1)(rw_0-2k-2)}{(k+1)(rw_0-k-2)}\\
&\geq (1-o(1))\,\frac{(2k+3)(2k+2)}{(k+1)(3k+2)}\\
&= (1-o(1))\,\frac{4k+6}{3k+2}\\
&> 1
\end{aligned}$$

for large N, where the first line follows because the fraction $\frac{(rw_0-2k-1)(rw_0-2k-2)}{(k+1)(rw_0-k-2)}$ is increasing in rw_0.

Upper Bound 2. $w_2^*(i) = \left\lceil \frac{N(i+\lceil \frac{rw_0-i}{2} \rceil - 1)}{rw_0-1} \right\rceil$

Because $w_0 = O(N/\log^2 N)$, we have $rw_0/N = o(1)$, and thus removing the ceilings will introduce an error of $(1+o(1))$. So, we can write the odd terms as

$$\begin{aligned}
&\binom{rw_0-k-1}{k}\binom{N-w_2^\star(i)}{k}\binom{w_2^\star(i)-1}{rw_0-k-1}\\
&\qquad =(1+o(1))\binom{rw_0-k-1}{k}\binom{\frac{Nk}{rw_0-1}}{k}\binom{\frac{N(rw_0-k-1)}{rw_0-1}-1}{rw_0-k-1}
\end{aligned}$$

First, let's consider bounded w_0. For $N \to \infty$, we can write each odd term with $k \geq 1$ as

$$(1+o(1))\frac{(rw_0-k-1)^{rw_0-k-1}k^k}{(rw_0-1)^{rw_0-1}(k!)^2(rw_0-2k-1)!}N^{rw_0-1}$$

Take the ratio between consecutive odd terms. Note that it could be the case that the previous odd term k has $w_2^* = 2\delta N$. This means that $2\delta N \leq \left\lceil \frac{N(i+\lceil \frac{rw_0-i}{2} \rceil - 1)}{rw_0-1} \right\rceil$, so we can safely replace w_2^* with the larger value for the purpose of this derivation, as it will only increase the kth odd term.

$$\begin{aligned}
&\frac{S_{k+1}}{S_k} := \frac{S(rw_0-2(k+1)-1, rw_0, w_2^*)}{S(rw_0-2k-1, rw_0, w_2^*)}\\
&= (1+o(1))\frac{(rw_0-2k-1)(rw_0-2k-2)}{(rw_0-k-1)(k+1)}\left(1-\frac{1}{rw_0-k-1}\right)^{rw_0-k-2}\left(1+\frac{1}{k}\right)^k\\
&\geq (1-o(1))\frac{(2k+3)(2k+2)}{(3k+3)(k+1)}\frac{1}{e}\left(1+\frac{1}{k}\right)^k\\
&= (1-o(1))\frac{2(2k+3)}{3e(k+1)}\left(1+\frac{1}{k}\right)^k\\
&> 1
\end{aligned}$$

where the third line again follows because the fraction $\frac{(rw_0-2k-1)(rw_0-2k-2)}{(rw_0-k-1)(k+1)}$ is increasing in rw_0. The final line follows from considering $k = 1$ and $k \geq 2$ separately. For $k = 1$, we have $\frac{20}{6e} > 1$. For $k \geq 2$, apply the bound $\left(1 + \frac{1}{k}\right)^k > \frac{ek}{k+1}$ and the result follows when $k^2 > 3$.

When $k = 0$, we can explicitly compute that case, and a direct comparison with the $k = 1$ case shows the ratio is > 1 for large N.

Now let's consider $w_0 \to \infty$. For fixed $k \geq 1$, using the exact binomial ratios and the fact that $\frac{N}{rw_0-1} \to \infty$, we have

$$\frac{S_{k+1}}{S_k} = (1 + o(1))\frac{rw_0}{e}\frac{(k+1)^{k-1}}{k^k}.$$

Since $rw_0 \to \infty$, this tends to infinity, so the ratio is > 1 for every fixed $k \geq 1$. When $k = 0$, we compute directly:

$$\frac{S_1}{S_0} = (1 + o(1))\frac{rw_0 - 2}{e} > 1$$

Finally, let's consider $k \to \infty$. We will use a variant of Stirling's formula to bound our odd terms. We will bound the latter two binomial coefficients as:

$$\begin{aligned}
&\binom{\frac{Nk}{rw_0-1}}{k}\binom{\frac{N(rw_0-k-1)}{rw_0-1} - 1}{rw_0 - k - 1} \\
&\qquad = (1 + o(1))\frac{1}{2\pi}\left(\frac{\left(\frac{N}{rw_0-1}\right)^{\frac{N}{rw_0-1}}}{\left(\frac{N}{rw_0-1} - 1\right)^{\frac{N}{rw_0-1}-1}}\right)^{rw_0-1} \frac{1}{\sqrt{k(rw_0 - k - 1)}}
\end{aligned}$$

Notice that the middle term is independent of k, so when we take the ratio of consecutive odd terms, it will disappear.

$$\begin{aligned}
\frac{S_{k+1}}{S_k} &= (1 + o(1))\frac{\binom{rw_0-k-2}{k+1}}{\binom{rw_0-k-1}{k}}\sqrt{\frac{k(rw_0 - k - 1)}{(k+1)(rw_0 - k - 2)}} \\
&= (1 + o(1))\frac{(rw_0 - 2k - 1)(rw_0 - 2k - 2)}{(rw_0 - k - 1)(k+1)}\sqrt{\frac{k(rw_0 - k - 1)}{(k+1)(rw_0 - k - 2)}} \\
&\geq (1 - o(1))\frac{2(2k+3)}{3(k+1)}\sqrt{\frac{k}{k+1}}
\end{aligned}$$

where the last line follows from the same bounding arguments we used in the previous cases. Now, as $k \to \infty$, the last line tends to $4/3 > 1$.

Together, these cases show that the maximum of i occurs at the end point $(r-2)w_0 - 1$. □

C Deferred Lemmas and Proofs for the High-Weight Case

C.1 Full Proof of Lemma 6

Proof. As previously stated, we can write

$$w_0 = \alpha N^a \qquad w_2 = \beta N^b \qquad w_3 = \rho N^c \qquad i = \gamma_1 N^{e_1} \qquad j = \gamma_2 N^{e_2}$$

where $1 - \frac{2 \log \log N}{\log N} \leq a \leq b \leq c \leq 1$ and $e_1, e_2 \leq a$.

Using the bound from [KZKC08, AR09], we can bound each binomial coefficient in the summand as:

$$\left(\frac{N}{w}\right)^w \frac{1}{\varphi_N(w)} \leq \binom{N}{w} \leq \left(\frac{N}{w}\right)^w \varphi_w(w),$$

where $\varphi_N(w) = \exp\left(\frac{w(w-1)}{2N}\right)$. Following the proof in [AR09], we will consider the summand as the product of two terms, corresponding to each accumulation. The first accumulation, which we will denote as A_1, can be bounded as follows:

$$\begin{aligned}
A_1 :=& \binom{n}{w_0} \frac{\binom{N-w_2}{i}\binom{w_2-1}{i-1}\binom{w_2-i}{w_1-2i}}{\binom{N}{w_1}} \leq \\
& \left(\frac{N}{r\alpha N^a}\right)^{\alpha N^a} \left(\frac{N-\beta N^b}{\gamma_1 N^{e_1}}\right)^{\gamma_1 N^{e_1}} \left[\frac{\gamma_1 N^{e_1}}{\beta N^b}\left(\frac{\beta N^b}{\gamma_1 N^{e_1}}\right)^{\gamma_1 N^{e_1}}\right] \\
& \left(\frac{\beta N^b - \gamma_1 N^{e_1}}{r\alpha N^a - 2\gamma_1 N^{e_1}}\right)^{r\alpha N^a - 2\gamma_1 N^{e_1}} \left(\frac{N}{r\alpha N^a}\right)^{-r\alpha N^a} \varphi_{\alpha N^a}(\alpha N^a) \\
& (\varphi_{\gamma_1 N^{e_1}}(\gamma_1 N^{e_1}))^2 \varphi_{r\alpha N^a - 2\gamma_1 N^{e_1}}(r\alpha N^a - 2\gamma_1 N^{e_1}) \frac{1}{\varphi_N(r\alpha N^a)} \\
=& \left(\frac{N}{r\alpha N^a}\right)^{(\alpha - r\alpha)N^a} \left(\frac{N-\beta N^b}{\gamma_1 N^{e_1}}\right)^{\gamma_1 N^{e_1}} \left(\frac{\beta N^b}{\gamma_1 N^{e_1}}\right)^{\gamma_1 N^{e_1}} \\
& \left(\frac{\beta N^b - \gamma_1 N^{e_1}}{r\alpha N^a - 2\gamma_1 N^{e_1}}\right)^{r\alpha N^a - 2\gamma_1 N^{e_1}} \phi_{A_1}(N)
\end{aligned}$$

where $\phi_{A_1}(N)$ is the product of all the φ terms, and the last equality follows from $a \geq e_1$. Note that we can trivially write $\phi_{A_1}(N) = \exp(o(N^a \ln N))$, which is a very loose upper bound that will be used in some cases of our analysis. The second accumulation, which we will denote as A_2, can be bounded similarly:

$$A_2 := \frac{\binom{N-w_3}{j}\binom{w_3-1}{j-1}\binom{w_3-j}{w_2-2j}}{\binom{N}{w_2}} \leq$$
$$\left(\frac{N-\rho N^c}{\gamma_2 N^{e_2}}\right)^{\gamma_2 N^{e_2}} \left[\frac{\gamma_2 N^{e_2}}{\rho N^c}\left(\frac{\rho N^c}{\gamma_2 N^{e_2}}\right)^{\gamma_2 N^{e_2}}\right] \left(\frac{N}{\beta N^b}\right)^{-\beta N^b}$$
$$\left(\frac{\rho N^c - \gamma_2 N^{e_2}}{\beta N^b - 2\gamma_2 N^{e_2}}\right)^{\beta N^b - 2\gamma_2 N^{e_2}} (\varphi_{\gamma_2 N^{e_2}}(\gamma_2 N^{e_2}))^2$$
$$\varphi_{\beta N^b - 2\gamma_2 N^{e_2}}(\beta N^b - 2\gamma_2 N^{e_2}) \frac{1}{\varphi_N(\beta N^b)}$$
$$= \left(\frac{N-\rho N^c}{\gamma_2 N^{e_2}}\right)^{\gamma_2 N^{e_2}} \left(\frac{\rho N^c}{\gamma_2 N^{e_2}}\right)^{\gamma_2 N^{e_2}} \left(\frac{N}{\beta N^b}\right)^{-\beta N^b}$$
$$\left(\frac{\rho N^c - \gamma_2 N^{e_2}}{\beta N^b - 2\gamma_2 N^{e_2}}\right)^{\beta N^b - 2\gamma_2 N^{e_2}} \phi_{A_2}(N)$$

Now, we must consider 5 cases corresponding to which variables are $o(N)$:

- **Case 1:** $1 - \frac{2\log\log N}{\log N} \leq a = b \leq c < 1$
- **Case 2:** $1 - \frac{2\log\log N}{\log N} \leq a = b < c \leq 1$
- **Case 3:** $1 - \frac{2\log\log N}{\log N} \leq a < b \leq c < 1$
- **Case 4:** $1 - \frac{2\log\log N}{\log N} \leq a < b < c = 1$
- **Case 5:** $1 - \frac{2\log\log N}{\log N} \leq a < b = c = 1$

We show that in all cases, $N^2 \cdot A_1 \cdot A_2 = \exp(-O(N/\log N))$.

We bound the terms in five cases.

Case 1. $1 - \frac{2\log\log N}{\log N} \leq a = b \leq c < 1$

We will calculate the order of both terms using the fact that $a = b$ and $\phi_{A_1}(N) = \exp(o(N^a \ln N))$.

$$A_1 \leq (O(1)N^{1-a})^{(\alpha - r\alpha)N^a} (O(1)N^{a-e_1}(N^{1-a} - \beta))^{\gamma_1 N^{e_1}} (O(1)N^{a-e_1})^{\gamma_1 N^{e_1}}$$
$$\left(\frac{\beta N^{a-e_1} - \gamma_1}{r\alpha N^{a-e_1} - 2\gamma_1}\right)^{r\alpha N^a - 2\gamma_1 N^{e_1}} \exp(o(N^a \ln N))$$
$$= \exp((1-a)(\alpha - r\alpha + \gamma_1^*)N^a \ln N + o(N^a \ln N))$$

where $\gamma_1^* = \gamma_1$ when $a = e_1$ and 0 otherwise. This can be seen by considering the subcases $a = e_1$ and $a > e_1$ separately. Then, in particular, $\gamma_1^* \leq \gamma_1 \leq \alpha$ as we are considering $0 \leq i \leq w_0$. Note that the constants and lower order terms get absorbed by the $o(N^a \ln N)$. Finally, we can write

$$A_1 \leq \exp(-(r-2)(1-a)\alpha N^a \ln N + o(N^a \ln N))$$

Since we have $r \geq 4$, we have shown that A_1 will decay to 0 as $N \to \infty$.

We can bound A_2 similarly, using $\gamma_2^* = \gamma_2$ when $b = e_2$ and 0 otherwise, and the bound $\gamma_2^* \leq \alpha$. Because $a = b$, we also have the bound $\beta \geq r\alpha/2$, so $\gamma_2^* \leq 2\beta/r \leq \beta/2$.

$$\begin{aligned}
A_2 &\leq (O(1)N^{c-e_2}(N^{1-c} - \rho))^{\gamma_2 N^{e_2}} (O(1)N^{c-e_2})^{\gamma_2 N^{e_2}} (O(1)N^{1-b})^{-\beta N^b} \\
&\quad (O(1)N^{c-b} + O(1))^{\beta N^b - 2\gamma_2 N^{e_2}} \exp(o(N^b \ln N)) \\
&= \exp((1-c)(\gamma_2^* - \beta)N^b \ln N + o(N^b \ln N)) \\
&\leq \exp(-(1-c)(\beta/2)N^b \ln N + o(N^b \ln N))
\end{aligned}$$

where we use the fact that $(N^{c-b})^{\beta N^b} = (N^{b-c})^{-\beta N^b}$ to obtain the second equality. Since $c < 1$ and $a = b \geq 1 - \frac{2 \log \log N}{\log N}$, it follows that $N^2 \cdot A_1 \cdot A_2 = \exp(-O(N^a \ln N)) = \exp(-O(N/\log N))$, which is negligible.

Case 2. $1 - \frac{2 \log \log N}{\log N} \leq a = b < c \leq 1$

As in the previous case, we have that

$$A_1 = \exp(-(r-2)(1-a)\alpha N^a \ln N + o(N^a \ln N))$$

Since $c = 1$ and $a = b$, the bound on the second term becomes the following:

$$A_2 = \exp(o(N^a \ln N))$$

Then, in particular,

$$N^2 \cdot A_1 \cdot A_2 = \exp(-(r-2)(1-a)\alpha N^a \ln N + o(N^a \ln N))$$

Case 3. $1 - \frac{2 \log \log N}{\log N} \leq a < b \leq c < 1$

We can derive a bound on A_1 without the assumption that $a = b$ in a similar way as before.

$$\begin{aligned}
A_1 &\leq (O(1)N^{1-a})^{(\alpha - r\alpha)N^a} (O(1)N^{b-e_1}(N^{1-b} - \beta))^{\gamma_1 N^{e_1}} (O(1)N^{b-e_1})^{\gamma_1 N^{e_1}} \\
&\quad \left(\frac{\beta N^{b-e_1} - \gamma_1}{r\alpha N^{a-e_1} - 2\gamma_1}\right)^{r\alpha N^a - 2\gamma_1 N^{e_1}} \exp(o(N^a \ln N)) \\
&= \exp([(1-a)(\alpha - r\alpha + \gamma_1^*) + (b-a)r\alpha]N^a \ln N + o(N^a \ln N)) \\
&\leq \exp(\alpha[-(1-a)(r-2) + (b-a)r]N^a \ln N + o(N^a \ln N))
\end{aligned}$$

The derivation for A_2 remains the same as in the first case.

$$A_2 = \exp(-(1-c)(\beta/2)N^b \ln N + o(N^b \ln N))$$

Because $b > a$, we have that

$$N^2 \cdot A_1 \cdot A_2 = \exp(-(1-c)(\beta/2)N^b \ln N + o(N^b \ln N))$$

Case 4. $1 - \frac{2\log\log N}{\log N} \le a < b < c = 1$

The bound on A_1 remains the same as in the previous case. To bound A_2, we will now need to consider the coefficients.

$$A_2 \le \left(\frac{1-\rho}{\gamma_2} N^{1-e_2}\right)^{\gamma_2 N^{e_2}} \left(\frac{\rho}{\gamma_2} N^{1-e_2}\right)^{\gamma_2 N^{e_2}} \left(\frac{1}{\beta} N^{1-b}\right)^{-\beta N^b} \left(\left(\frac{\rho - o(1)}{\beta - o(1)}\right) N^{1-b}\right)^{\beta N^b - 2\gamma_2 N^{e_2}} \cdot \phi_{A_2}(N)$$

where

$$\begin{aligned}
\phi_{A_2}(N) &= \frac{(\varphi_{\gamma_2 N^{e_2}}(\gamma_2 N^{e_2}))^2 \varphi_{\beta N^b - 2\gamma_2 N^{e_2}}(\beta N^b - 2\gamma_2 N^{e_2})}{\varphi_N(\beta N^b)} \\
&= \exp\left((\gamma_2 N^{e_2} - 1) + (\beta N^b - 2\gamma_2 N^{e_2} - 1)/2 - \beta N^{b-1}(\beta N^b - 1)/2\right) \\
&\le \exp\left((\beta N^b - \beta^2 N^{2b-1} + \beta N^{b-1})/2\right) \\
&\le \exp\left(\beta N^b/2 + o(1)\right)
\end{aligned}$$

where the last line follows from $b \le 1$.

Since $a < b$, we have $e_2 < b$. Then, the first two terms in the product are $\exp(o(N^b))$.

$$\begin{aligned}
A_2 &= \exp\Big((\beta N^b - 2\gamma_2 N^{e_2})\left[(1-b)\ln N + \ln(\rho/\beta) + o(1)\right] \\
&\qquad - \beta N^b\left[(1-b)\ln N - \ln\beta\right] + \beta N^b/2 + o(N^b)\Big) \\
&= \exp((1/2 + \ln(\rho))\beta N^b + o(N^b))
\end{aligned}$$

Note that because $\rho \le \delta \le \frac{1}{2}$, the exponent is negative, thus $N^2 \cdot A_1 \cdot A_2$ is still $\exp(-O(N/\log N))$.

Case 5: $1 - \frac{2\log\log N}{\log N} \le a < b = c = 1$

This case is the same as the previous one, except we have

$$\begin{aligned}
\phi_{A_2}(N) &= \frac{(\varphi_{\gamma_2 N^{e_2}}(\gamma_2 N^{e_2}))^2 \varphi_{\beta N - 2\gamma_2 N^{e_2}}(\beta N - 2\gamma_2 N^{e_2})}{\varphi_N(\beta N)} \\
&= \exp\left((\gamma_2 N^{e_2} - 1) + (\beta N - 2\gamma_2 N^{e_2} - 1)/2 - \beta(\beta N - 1)/2\right) \\
&\le \exp\left((1-\beta)\beta N/2\right)
\end{aligned}$$

so our bound becomes

$$A_2 = \exp(((1-\beta)/2 + \ln(\rho))\beta N + o(N))$$

□

C.2 Full Proof of Lemma 7

Lemma. *For $r = 4$ and $\delta = 1/2$, when w_0, w_2, w_3 are all linear in N, the summand $S(i, j, w_0, w_2, w_3) = \exp(-O(N))$.*

Proof. Recall

$$S(i,j,w_0,w_2,w_3) = \binom{n}{w_0} \frac{\binom{N-w_2}{i}\binom{w_2-1}{i-1}\binom{w_2-i}{w_1-2i}}{\binom{N}{w_1}} \frac{\binom{N-w_3}{j}\binom{w_3-1}{j-1}\binom{w_3-j}{w_2-2j}}{\binom{N}{w_2}},$$

where, for $r = 4$, $w_1 = 4w_0$ and $N = 4n$.

Removing dependence on w_3. For the first accumulation, using the basic facts about binomial coefficients, we have

$$\binom{N-w_2}{i} \le \binom{N-w_2+1}{i}, \quad \text{and} \quad \binom{w_2-1}{i-1} \le \binom{w_2+1}{i}.$$

Moreover, we also have

$$\binom{w_2-i}{w_1-2i} \le \binom{w_2}{w_1-i}.$$

Simply fix i additional elements in a set of size w_2. Every (w_1-2i)-subset of the remaining w_2-i elements can be mapped injectively to a (w_1-i)-subset of the full w_2-element set by adjoining these i fixed elements. Therefore,

$$\frac{\binom{N-w_2}{i}\binom{w_2-1}{i-1}\binom{w_2-i}{w_1-2i}}{\binom{N}{w_1}} \le \frac{\binom{N-w_2+1}{i}\binom{w_2+1}{i}\binom{w_2}{w_1-i}}{\binom{N}{w_1}}.$$

We can do the same for the second accumulation term. Thus, defining

$$A(t) := \binom{N-t+1}{j}\binom{t+1}{j}\binom{t}{w_2-j},$$

we have

$$\frac{\binom{N-w_3}{j}\binom{w_3-1}{j-1}\binom{w_3-j}{w_2-2j}}{\binom{N}{w_2}} \le \frac{A(w_3)}{\binom{N}{w_2}}.$$

We now show that $A(t)$ is non-decreasing for $t < N/2$. Indeed,

$$\frac{A(t+1)}{A(t)} = \frac{t+2}{t+2-j}\frac{N-t+1-j}{N-t+1}\frac{t+1}{t+1-w_2+j}.$$

The last term is ≥ 1 as $w_2 \ge j$. For the first two terms, we have $\frac{t+2}{t+2-j}\frac{N-t+1-j}{N-t+1} \ge 1$ if and only if $(t+2)(N-t+1-j) \ge (t+2-j)(N-t+1)$. This is equivalent to the condition $j(N-2t-1) \ge 0$, which holds whenever $t < N/2$. Thus, $A(w_3) \le A(N/2)$ for every $w_3 \le N/2$.

Now, we have

$$S(i,j,w_0,w_2,w_3) \leq \binom{N/4}{w_0} \frac{\binom{N-w_2+1}{i}\binom{w_2+1}{i}\binom{w_2}{4w_0-i}}{\binom{N}{4w_0}} \frac{\binom{N/2+1}{j}^2\binom{N/2}{w_2-j}}{\binom{N}{w_2}}.$$

As in the sublinear case, we write

$$w_0 = \alpha N, \qquad w_2 = \beta N, \qquad i = \gamma_1 N, \qquad j = \gamma_2 N.$$

Applying Stirling's Approximation to the binomial coefficients gives

$$S(i,j,w_0,w_2,w_3) \leq \text{poly}(N) \cdot 2^{N\Psi_4(\alpha,\gamma_1,\gamma_2,\beta)},$$

where

$$\begin{aligned}\Psi_4(\alpha,\gamma_1,\gamma_2,\beta) = &-\frac{3}{4}H(4\alpha) + \beta H\left(\frac{\gamma_1}{\beta}\right) + (1-\beta)H\left(\frac{\gamma_1}{1-\beta}\right) \\ &+ \beta H\left(\frac{4\alpha-\gamma_1}{\beta}\right) + H(2\gamma_2) + \frac{1}{2}H(2(\beta-\gamma_2)) - H(\beta).\end{aligned}$$

since we are fixing $r = 4$. In the region under consideration, we have $i+j \leq w_0-1$.

Removing dependence on γ_1. We now maximize

$$\Psi_4(\alpha,\gamma_1,\gamma_2,\beta)$$

over the feasible region

$$0 < \alpha \leq \frac{1}{6}, \qquad \gamma_1,\gamma_2 \geq 0, \qquad \gamma_1+\gamma_2 \leq \alpha,$$

subject to

$$4\alpha-\gamma_1 \leq \beta, \qquad \gamma_1 \leq 1-\beta, \qquad \beta-\gamma_2 \leq \frac{1}{2}.$$

For fixed α, γ_2, and β, differentiating with respect to γ_1 gives

$$\frac{\partial\Psi_4}{\partial\gamma_1} = \log\left(\frac{(\beta-\gamma_1)(1-\beta-\gamma_1)(4\alpha-\gamma_1)}{\gamma_1^2(\beta-4\alpha+\gamma_1)}\right).$$

Throughout the feasible region, we can show that this derivative is positive. First, $\beta-\gamma_1 \geq (4\alpha-\gamma_1)-\gamma_1 \geq 2\alpha$. Next, $1-\beta-\gamma_1 \geq 1-(1/2+\gamma_2)-\gamma_1 \geq 1/2-\alpha \geq 1/3$. Since $4\alpha-\gamma_1 \geq 3\alpha$, the numerator is at least $2\alpha^2$. For the denominator, we have $\gamma_1^2 \leq \alpha^2$ and $\beta-4\alpha+\gamma_1 \leq 1/2+\gamma_2-4\alpha+\gamma_1 \leq 1/2$. So,

$$\frac{\partial\Psi_4}{\partial\gamma_1} \geq \log 4 > 0$$

Therefore Ψ_4 is increasing with γ_1 until the boundary $\gamma_1+\gamma_2 = \alpha$, where the maximum occurs. Substituting $\gamma_1 = \alpha-\gamma_2$ into our equation, we have

$$\beta H\left(\frac{\gamma_1}{\beta}\right) + (1-\beta)H\left(\frac{\gamma_1}{1-\beta}\right) \leq H(2\gamma_1)$$

and

$$H(2\gamma_1) + H(2\gamma_2) \leq 2H(\gamma_1 + \gamma_2) = 2H(\alpha).$$

Additionally, $4\alpha - \gamma_1 = 4\alpha - (\alpha - \gamma_2) = 3\alpha + \gamma_2$. Therefore,

$$\Psi_4(\alpha, \gamma_1, \gamma_2, \beta) \leq V(\alpha, \gamma_2, \beta),$$

where

$$V(\alpha, \gamma_2, \beta) = -\frac{3}{4}H(4\alpha) + 2H(\alpha) + \beta H\left(\frac{3\alpha + \gamma_2}{\beta}\right) + \frac{1}{2}H\left(2(\beta - \gamma_2)\right) - H(\beta).$$

Removing dependence on β. The feasible region for the remaining variables is

$$0 \leq \gamma_2 \leq \alpha \leq \frac{1}{6}, \qquad 3\alpha + \gamma_2 \leq \beta \leq \frac{1}{2} + \gamma_2.$$

We find the maximum of V by computing the second partial derivative with respect to β:

$$\frac{\partial V}{\partial \beta} = \log\left(\frac{\beta^2(1 - 2\beta + 2\gamma_2)}{2(\beta - \gamma_2)(\beta - 3\alpha - \gamma_2)(1 - \beta)}\right)$$

$$\begin{aligned}\frac{\partial^2 V}{\partial \beta^2} &= \frac{2}{\beta} - \frac{1}{\beta - 3\alpha - \gamma_2} - \frac{1}{\beta - \gamma_2} - \frac{2}{1 - 2\beta + 2\gamma_2} + \frac{1}{1 - \beta} \\ &< \frac{2}{\beta} - \frac{1}{\beta} - \frac{1}{\beta} - \frac{2}{2 - 2\beta} + \frac{1}{1 - \beta} = 0\end{aligned}$$

where the second line follows from $\gamma_2 \leq \alpha \leq \frac{1}{6} < \frac{1}{2}$. This implies that V is concave in β and the unique maximum occurs at $\frac{\partial V}{\partial \beta} = 0$. If we set $x := \beta - \gamma_2$, this holds when

$$(1 + 2(3\alpha + \gamma_2))x^2 - 2(3\alpha + \gamma_2)(1 - \gamma_2)x - \gamma_2^2 = 0$$

. The unique, positive solution to this quadratic equation in x is

$$x^*(\alpha, \gamma_2) = \frac{(3\alpha + \gamma_2)(1 - \gamma_2) + \sqrt{(3\alpha + \gamma_2)^2(1 - \gamma_2)^2 + \gamma_2^2(1 + 2(3\alpha + \gamma_2))}}{1 + 2(3\alpha + \gamma_2)},$$

which means the maximizing value of β is $\beta^*(\alpha, \gamma_2) = x^*(\alpha, \gamma_2) + \gamma_2$.

Removing dependence on γ_2. We upper bounded our equation by $V(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))$ subject to $0 \leq \gamma_2 \leq \alpha$. Now, we differentiate with respect to γ_2:

$$\frac{dV(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))}{d\gamma_2} = \frac{\partial V}{\partial \gamma_2} + \frac{\partial V}{\partial \beta}\frac{\partial \beta^*}{\partial \gamma_2} = \frac{\partial V}{\partial \gamma_2},$$

as $\frac{\partial V}{\partial \beta}(\alpha, \gamma_2, \beta^*) = 0$. For a fixed β, we thus have

$$\frac{dV(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))}{d\gamma_2} = \log\left(\frac{\beta - 3\alpha - \gamma_2}{3\alpha + \gamma_2}\right) + \log\left(\frac{2(\beta^* - \gamma_2)}{1 - 2(\beta^* - \gamma_2)}\right)$$

For the second term, using the condition $\frac{\partial V}{\partial \beta} = 0$, we obtain

$$\begin{aligned}
&= \log\left(\frac{\beta - 3\alpha - \gamma_2}{3\alpha + \gamma_2}\right) + \log\left(\frac{\beta^2}{(\beta - 3\alpha - \gamma_2)(1-\beta)}\right) \\
&= \log\left(\frac{\beta^2}{(3\alpha + \gamma_2)(1-\beta)}\right).
\end{aligned}$$

We will show the function is convex, so we will compute the second derivative:

$$\frac{d^2 V(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))}{(d\gamma_2)^2} = \frac{\partial \beta^*}{\partial \gamma_2}\left(\frac{2}{\beta} + \frac{1}{1-\beta}\right) - \frac{1}{3\alpha + \gamma_2}$$

Computing $\frac{\partial \beta^*}{\partial \gamma_2}$ implicitly gives us a fraction with numerator $\frac{1}{\beta - 3\alpha - \gamma_2} + \frac{1}{\beta^* - \gamma_2} + \frac{2}{1 - 2(\beta^* - \gamma_2)}$ and denominator $\frac{1}{\beta - 3\alpha - \gamma_2} + \frac{1}{\beta^* - \gamma_2} + \frac{2}{1 - 2(\beta^* - \gamma_2)} - \frac{2}{\beta} - \frac{1}{1-\beta}$. We now show that

$$\frac{\partial \beta^*}{\partial \gamma_2} \geq \frac{\beta^*}{3\alpha + \gamma_2}.$$

Since the denominator is positive, this inequality is equivalent to

$$\begin{aligned}
&\beta^*\left(\frac{2}{\beta^*} + \frac{1}{1-\beta^*}\right) \\
&\quad - (\beta^* - 3\alpha - \gamma_2)\left(\frac{1}{\beta^* - 3\alpha - \gamma_2} + \frac{1}{\beta^* - \gamma_2} + \frac{2}{1 - 2(\beta^* - \gamma_2)}\right) \geq 0.
\end{aligned}$$

Using the stationarity condition $\frac{\partial V}{\partial \beta}(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2)) = 0$, the left-hand side simplifies to

$$\frac{3\alpha - (3\alpha + \gamma_2)\beta^*}{(\beta^* - \gamma_2)(1 - \beta^*)}.$$

The denominator is positive throughout the interior of the feasible region. For the numerator, since $0 \leq \gamma_2 \leq \alpha$, we have $3\alpha + \gamma_2 \leq 4\alpha$. Also, from $\beta^* \leq \frac{1}{2} + \gamma_2$ and $\gamma_2 \leq \alpha \leq 1/6$, we obtain $\beta^* \leq \frac{1}{2} + \alpha \leq \frac{2}{3}$. Therefore,

$$(3\alpha + \gamma_2)\beta^* \leq 4\alpha \cdot \frac{2}{3} = \frac{8}{3} \cdot \alpha < 3\alpha,$$

and so the partial derivative $\frac{\partial \beta^*}{\partial \gamma_2} \geq \frac{\beta^*}{3\alpha + \gamma_2}$. Substituting this into the expression for the second derivative gives

$$\begin{aligned}
\frac{d^2 V(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))}{(d\gamma_2)^2} &\geq \frac{\beta^*}{3\alpha + \gamma_2}\left(\frac{2}{\beta^*} + \frac{1}{1-\beta^*}\right) - \frac{1}{3\alpha + \gamma_2} \\
&= \frac{1}{3\alpha + \gamma_2}\left(1 + \frac{\beta^*}{1-\beta^*}\right) \\
&> 0.
\end{aligned}$$

So, $V\left(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2)\right)$ is convex in γ_2. Since $0 \leq \gamma_2 \leq \alpha$, its maximum is attained at one of the endpoints. Therefore,

$$V\left(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2)\right) \leq \max\left\{V\left(\alpha, 0, \beta^*(\alpha, 0)\right), V\left(\alpha, \alpha, \beta^*(\alpha, \alpha)\right)\right\}.$$

We now have to bound these two endpoints. Define $E_0(\alpha) := V\left(\alpha, 0, \beta^*(\alpha, 0)\right)$ and $E_\alpha(\alpha) := V\left(\alpha, \alpha, \beta^*(\alpha, \alpha)\right)$ for convenience.

Endpoint 1: E_0. From our equation for $\beta^*(\alpha, \gamma_2)$, we obtain $\beta^*(\alpha, 0) = \frac{6\alpha}{1+6\alpha}$. Plugging this into our equation for V, we obtain

$$\begin{aligned} E_0(\alpha) = & -\frac{3}{4}H(4\alpha) + 2H(\alpha) + \frac{6\alpha}{1+6\alpha}H\left(\frac{1+6\alpha}{2}\right) \\ & + \frac{1}{2}H\left(\frac{12\alpha}{1+6\alpha}\right) - H\left(\frac{6\alpha}{1+6\alpha}\right). \end{aligned}$$

Differentiating with respect to α gives

$$E_0'(\alpha) = \log\left(\frac{64\alpha(1-\alpha)^2(1-6\alpha)^3}{(1-4\alpha)^3(1+6\alpha)^3}\right).$$

To show that $E_0(\alpha)$ is decreasing on the region, we need to show that this equation is always positive on the region $0 < \alpha < 1/6$, or equivalently, the denominator is larger than the numerator. We can just compute this:

$$\begin{aligned} & (1-4\alpha)^3(1+6\alpha)^3 - 64\alpha(1-\alpha)^2(1-6\alpha)^3 \\ & \quad = (1-2\alpha)\left[3\left(72\alpha^2 - 17\alpha + \frac{28}{51}\right)^2 + \frac{65}{17}\alpha^2 + \frac{83}{867}\right] > 0 \end{aligned}$$

Now, we know that $E_0(\alpha)$ is strictly decreasing with α on $(0, \alpha)$. At $\alpha = 0$, from the equation for E_0 it is easy to verify that $E_0(\alpha) = 0$ as $H(0) = 0$. That means that for every $\alpha > 0$, $E_0(\alpha) < 0$ by continuity and we have shown negativity.

Endpoint 2: E_α. Here, we use the convexity of our function $V(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2)$ in γ_2 on $0 \leq \gamma_2 \leq \alpha$ to have the bound $E_\alpha(\alpha) \leq E_0(\alpha) + \alpha \cdot \frac{\partial}{\partial \gamma_2} V(\alpha, \gamma_2, \beta^*(\alpha, \gamma_2))$. At $\gamma_2 = \alpha$,

$$\frac{\partial}{\partial \gamma_2} V(\alpha, \alpha, \beta^*(\alpha, \alpha)) = \log\left(\frac{(\beta^*(\alpha, \alpha)^2)}{4\alpha(1-\beta^*(\alpha, \alpha))}\right)$$

from the derivations we obtained earlier. We explicitly calculate β^*.

$$\beta^*(\alpha, \alpha) = \frac{\alpha\left(5 + 4\alpha + \sqrt{17 - 24\alpha + 16\alpha^2}\right)}{1+8\alpha} \leq \frac{10\alpha}{1+9\alpha}$$

for $0 < \alpha \leq 1/6$. The expression inside of log is increasing with β^* on the region $0 < \beta^* < 1$, so we have $\frac{\beta^*(\alpha,\alpha)^2}{4\alpha(1-\beta^*(\alpha,\alpha))} \leq \frac{25\alpha}{(1+9\alpha)(1-\alpha)}$. Therefore,

$$E_\alpha(\alpha) \leq E_0(\alpha) + \alpha \log\left(\frac{25\alpha}{(1+9\alpha)(1-\alpha)}\right) = U(\alpha).$$

Showing this upper bound is negative is a bit tricky since the second term in the sum is not negative throughout the whole region. We split into subregions.

Region $0 < \alpha \leq 1/32$**.** Computing the derivative of the term inside of the log,

$$\frac{d}{d\alpha}\left(\frac{25\alpha}{(1+9\alpha)(1-\alpha)}\right) = \frac{1}{\alpha(1+9\alpha)} + \frac{1}{1-\alpha} < 0$$

for $0 < \alpha < 1$. We can see that the error term is increasing and upper bound it by its value at $\alpha = 1/32$. We have $\frac{32}{(1+9(1/32))} + \frac{1}{1-1/32} = \frac{800}{1271} < 1$ which means the whole error term is negative on the region. Since we know $E_0(\alpha)$ is negative on the region we have that the full upper bound $U(\alpha)$ is negative.

Region $1/32 < \alpha \leq 1/16$**.** We again use the monotonicity of each term to get the upper bound

$$U(\alpha) \leq E_0\left(\frac{1}{32}\right) + \frac{1}{16}\log\left(\frac{16}{15}\right)$$

We can use entropy bounds on the term for $E_0(\alpha)$ then compute explicitly

$$U(\alpha) \leq \frac{1}{32}\left(\log 2 - \frac{19}{16}\right) + \frac{1}{16}\left(\frac{1}{15}\right) < 0.$$

Region $1/16 < \alpha \leq 1/6$**.** If we show that U is concave on this region, i.e. $U''(\alpha) < 0$, we can use a tangent line at a convenient value of $\alpha = 1/10$ to upper bound U.

$$U(\alpha) \leq U\left(\frac{1}{10}\right) + U'\left(\frac{1}{10}\right)\left(\alpha - \frac{1}{10}\right)$$

It is simple to get the estimates $U\left(\frac{1}{10}\right) < -\frac{1}{180}$ and $0 < U'\left(\frac{1}{10}\right) < \frac{1}{20}$ directly from the equations. If $\alpha \leq 1/10$, then $\alpha - \frac{1}{10} \leq 0$, and so $U(\alpha) \leq U\left(\frac{1}{10}\right) < 0$. If $1/10 \leq \alpha \leq 1/6$, then $\alpha - \frac{1}{10} \leq \frac{1}{15}$, so $U(\alpha) < -\frac{1}{180} + \frac{1}{20}\cdot\frac{1}{15} = -\frac{1}{450} < 0$. Therefore we have shown $E_\alpha(\alpha)$ is negative on the region if U is concave. Differentiating $U(\alpha)$ twice gives

$$\begin{aligned} U''(\alpha) = &\left(\frac{2}{\alpha} - \frac{18}{1+6\alpha} - \frac{9}{1+9\alpha} - \frac{9}{(1+9\alpha)^2}\right) \\ &+ \left(\frac{\alpha}{(1-\alpha)^2} - \frac{6}{(1-6\alpha)(1-4\alpha)}\right). \end{aligned}$$

We can bound the second line on our region $1/16 < \alpha \leq 1/6$ as follows:

$$\frac{\alpha}{(1-\alpha)^2} - \frac{6}{(1-6\alpha)(1-4\alpha)} \leq \frac{6}{25} - \frac{64}{5} = -\frac{314}{25}$$

The first line is decreasing on the interval which can be seen by computing the derivative. We bound it above by its evaluation on $1/16$ which is $\frac{65056}{6875}$ to conclude that $U''(\alpha) < 0$ and U is concave on this region.

We have concluded that the value of both endpoints remain negative on the region. □

D Polynomial Commitment based on Error-Correcting Codes

In this section, we present the construction of the polynomial commitment scheme (PCS) we used in our evaluations. The description mostly follows the prior works of [GLS+23, BFK+24] verbatim.

A PCS allows the prover to commit to a polynomial, and later prove to the verifier that its evaluation at a particular point is computed correctly. It consists of the following algorithms:

- $\mathsf{pp} \leftarrow \mathsf{PC.setup}(1^\lambda, \mathcal{F})$: Given the security parameter and a family of polynomials, output the public parameters pp.
- $\mathsf{com} \leftarrow \mathsf{PC.commit}(\phi, \mathsf{pp})$: On input a polynomial ϕ and the public parameter pp, output the commitment $\mathcal{C}$.
- $(y, \pi) \leftarrow \mathsf{PC.open}(\phi, \mathsf{pp}, x)$: On input the polynomial, the public parameters, and the point to be evaluated at, output $y = \phi(x)$, and the proof.
- $\{0, 1\} \leftarrow \mathsf{PC.verify}(\mathsf{pp}, \mathsf{com}, x, y, \pi)$: Verify the proof against the claim $\phi(x) = y$, output 1 if verification passes, otherwise 0.

A PCS scheme satisfies the following two properties: (1) Completeness: the verification always outputs 1 if the evaluation and the proof are computed honestly; (2) Soundness: for any probabilistic polynomial time (PPT) adversary generating an incorrect evaluation, the probability of the verification outputting 1 is negligible. See [GLS+23] for the formal definitions.

Construction. We present the construction of PCS proposed in [GLS+23] based on any error-correcting code with constant relative distance. The construction uses the Merkle tree [Mer87], a data structure that allows one to commit to a vector of messages by a single hash value h, and later open any message a_i with a proof of correctness. We denote the algorithm as:

- $h \leftarrow \mathsf{Merkle.Commit}(\mathbf{a})$.
- $(a_i, \pi_i) \leftarrow \mathsf{Merkle.Open}(\mathbf{a}, i)$.
- $\{0, 1\} \leftarrow \mathsf{Merkle.Verify}(\pi_\mathsf{i}, a_i, h)$.

A polynomial evaluation can be expressed as a tensor product. Here we use a multilinear polynomial as an example, but the construction works for univariate polynomials as well. Given a multilinear polynomial ϕ, its evaluation on input vector $x_0, x_1, ..., x_{\log m-1}$ is:

$$\phi(x_0, x_1, ..., x_{\log m-1}) = \sum_{i_0=0}^{1} \sum_{i_1=0}^{1} ... \sum_{i_{\log N-1}=0}^{1} c_{i_0 i_1 ... i_{\log m-1}} x_0^{i_0} x_1^{i_1} ... x_{\log m-1}^{i_{\log m-1}}.$$

The degree of each variable is either 0 or 1 by the definition of a multilinear polynomial, and thus there are m monomials and coefficients with $\log m$ variables. Let $i_0 i_1 ... i_{\log m-1}$ be the binary representation of number i, i.e., $i = \sum_{j=0}^{\log m-1} 2^j i_j$. We use c to denote the coefficients where $c[i] = c_{i_0 i_1 ... i_{\log m-1}}$.

Protocol 1 Polynomial Commitment Scheme [GLS+23]

Public input: The evaluation point $\mathbf{x}$, parsed as a tensor product $\mathbf{x} = \mathbf{x}_0 \otimes \mathbf{x}_1$;
Private input: The polynomial ϕ, the coefficients of ϕ are denoted by c.
Let C be the $[N, n, \delta N]$-linear code, $E_C : \mathbb{F}_q^n \to \mathbb{F}_q^N$ be the encoding function, $m = n \times n$. If m is not a perfect square, we can pad it to the next perfect square. We use a python style notation to select the i-th column of a matrix $\mathsf{mat}[:, i]$.

function PC.Commit(ϕ)
  Parse c as a $n \times n$ matrix C. The prover computes the encoding C_1 locally. In particular, C_1 is a $n \times N$ matrix by encoding each row of the matrix C using E_C.
  for $i \in [N]$ **do**
    Compute the Merkle tree root $\mathsf{Root}_i = \mathsf{Merkle.Commit}(\mathsf{C}_1[:, \mathsf{i}])$.
  Compute a Merkle tree root $\mathsf{Root} = \mathsf{Merkle.Commit}([\mathsf{Root}_0, ..., \mathsf{Root}_{N-1}])$ and output Root as the commitment.
function PC.Open($\phi, \mathbf{x}$)
  $\mathcal{P}$ receives a random vector $\mathbf{r} \in \mathbb{F}_q^n$ from $\mathcal{V}$, and computes $\mathbf{y}_r = \sum_{i=0}^{n-1} \mathbf{r}[i]\mathsf{C}[i]$, $\mathbf{y}_1 = \sum_{i=0}^{n-1} \mathbf{x}_0[i]\mathsf{C}[i]$.
  $\mathcal{P}$ sends $\mathbf{y}_r, \mathbf{y}_1$ to $\mathcal{V}$.
  $\mathcal{V}$ randomly samples $t \in [N]$ indices as an array I and sends it to $\mathcal{P}$.
  for $i \in I$ **do**
    $\mathcal{P}$ sends $\mathsf{C}_1[:, i]$ and $(\mathsf{Root}_i, \pi_i) \leftarrow \mathsf{Merkle.Open}([\mathsf{Root}_0, ..., \mathsf{Root}_{N-1}], i)$ to $\mathcal{V}$.
function PC.Verify($\pi_{\mathbf{x}}, \mathbf{x}, y = \phi(\mathbf{x}), \mathsf{Root}$)
  Compute $\mathbf{c}_r = E_C(\mathbf{y}_r)$ and $\mathbf{c}_1 = E_C(y_1)$.
  for $\forall i \in I$ **do**
    Check if $\mathbf{c}_r[i] == \langle \mathbf{r}, \mathsf{C}_1[:, i] \rangle$.
    Check if $\mathbf{c}_1[i] == \langle \mathbf{x}_0, \mathsf{C}_1[:, i] \rangle$.
    Compute $\mathsf{Root}_i = \mathsf{Merkle.Commit}\mathsf{C}_1[:, i]$); run $\mathsf{Merkle.Verify}(\pi_i, \mathsf{Root}_i, \mathsf{Root})$.
  Check if $y == \langle \mathbf{x}_1, \mathbf{y}_1 \rangle$.
  Output 1 if all if the above conditions hold, otherwise, output 0.

Similarly, we define $X_i = x_0^{i_0} x_1^{i_1} ... x_{\log m-1}^{i_{\log m-1}}$. For $n = \sqrt{m}$, define vectors $\mathbf{x}_0, \mathbf{x}_1$ as $\mathbf{x}_0 = \{X_0, X_1, ..., X_{n-1}\}, \mathbf{x}_1 = \{X_{0\times n}, X_{1\times n}, X_{2\times n}, ..., X_{(n-1)\times n}\}$. Then we have $X = \mathbf{x}_0 \otimes \mathbf{x}_1$. The polynomial evaluation is reduced to a tensor product $\phi(x_0, x_1, ..., x_{\log m-1}) = \langle c, \mathbf{x}_0 \otimes \mathbf{x}_1 \rangle$.

The polynomial commitment scheme is presented in Protocol 1, where the message length of the error-correcting code is set to $n = \sqrt{m}$, the squre-root of the size of the polynomial m.

Finally, zkSNARKs can be constructed by combining the polynomial commitment scheme various polynomial interactive oracle proof (PIOP) protocols (e.g., [GWC19, XZZ+19, Set20, GLS+23, CBBZ22]). See [Tha22] for a survey. We omit the formal descriptions in this paper.