Electronic Colloquium on Computational Complexity, Report No. 91 (2024)

When Do Low-Rate Concatenated Codes Approach The Gilbert–Varshamov Bound?

Dean Doron* Jonathan Mosheiff† Mary Wootters‡

Abstract

The *Gilbert–Varshamov* (GV) bound is a classical existential result in coding theory. It implies that a random linear binary code of rate ε^2 has relative distance at least $\frac{1}{2} - O(\varepsilon)$ with high probability. However, it is a major challenge to construct *explicit* codes with similar parameters.

One hope to derandomize the Gilbert–Varshamov construction is with code concatenation: We begin with a (hopefully explicit) outer code $\mathcal{C}_{\text{out}}$ over a large alphabet, and concatenate that with a small binary random linear code $\mathcal{C}_{\text{in}}$. It is known that when we use *independent* small codes for each coordinate, then the result lies on the GV bound with high probability, but this still uses a lot of randomness. In this paper, we consider the question of whether code concatenation with a *single* random linear inner code $\mathcal{C}_{\text{in}}$ can lie on the GV bound; and if so what conditions on $\mathcal{C}_{\text{out}}$ are sufficient for this.

We show that first, there *do* exist linear outer codes $\mathcal{C}_{\text{out}}$ that are "good" for concatenation in this sense (in fact, *most* linear codes codes are good). We also provide two sufficient conditions for $\mathcal{C}_{\text{out}}$, so that if $\mathcal{C}_{\text{out}}$ satisfies these, $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ will likely lie on the GV bound. We hope that these conditions may inspire future work towards constructing explicit codes $\mathcal{C}_{\text{out}}$.

*Ben-Gurion University of the Negev. deand@bgu.ac.il. Supported in part by NSF-BSF grant #2022644.

†Ben-Gurion University of the Negev. mosheiff@bgu.ac.il. Supported by an Alon Fellowship.

‡Stanford University. marykw@stanford.edu. Partially supported by NSF grants CCF-2231157 and CCF-2133154.

ISSN 1433-8092

1 Introduction

An *error correcting code* (or just a *code*) is a subset $\mathcal{C} \subseteq \Sigma^n$, for some alphabet Σ. We think of a code $\mathcal{C}$ being used to encode messages in Σ^k for $k = \log_{|\Sigma|} |\mathcal{C}|$. That is, for any $m \in \Sigma^k$, we can identify m with a codeword $\mathcal{C}(m) \in \mathcal{C}$.[1] The idea is that encoding m into the codeword $\mathcal{C}(m)$ will introduce redundancy that can later be used to correct errors. In this work we focus on *linear* codes $\mathcal{C}$, which are codes where $\Sigma = \mathbb{F}$ is a finite field and $\mathcal{C} \subseteq \mathbb{F}^n$ is a linear subspace of $\mathbb{F}^n$.

Two important properties of error correcting codes are the *rate* R and the *relative distance* δ. For a code $\mathcal{C} \subseteq \Sigma^n$, the rate is defined as $R = \frac{\log_{|\Sigma|} |\mathcal{C}|}{n} = \frac{k}{n}$, and it quantifies how large the code is. The rate is between 0 and 1, and typically we want it to be as close to 1 as possible; this means that the encoding map does not introduce much redundancy. The (relative) distance of $\mathcal{C} \subseteq \Sigma^n$ is defined as $\delta = \frac{1}{n} \min_{c \neq c' \in \mathcal{C}} \Delta(c, c')$, where $\Delta(\cdot, \cdot)$ is Hamming distance. Again, the relative distance is between 0 and 1, and again we typically want it to be as close to 1 as possible; this means that the code can correct many worst-case errors.

These two quantities—rate and distance—are in tension. The larger the rate is, the smaller the distance must be. For binary codes (that is, codes where $\Sigma = \mathbb{F}_2$), it is a major open question to pin down the best trade-off possible between rate and distance. However, we know that good trade-offs are possible: The best known possibility result in general is the *Gilbert–Varshamov* (GV) bound (Theorem 2.1).

In this paper we focus on *low rate* codes. In this parameter regime, the GV bound implies that there *exist* binary linear codes with relative distance $\frac{1-\varepsilon}{2}$ and rate $\Omega(\varepsilon^2)$, for small $\varepsilon > 0$. In fact, Varshamov's proof shows that a random binary linear code achieves this with high probability.

Constructing such codes explicitly, hopefully accompanied by an efficient decoding algorithm, has been subject to extensive and fruitful research in the past decades (e.g., [NN93, ABN+92, AGHP92, BT13, GI05, Ta-17, BD22]), with several exciting breakthroughs in recent years. These breakthroughs include explicit constructions of codes with distance $\delta = \frac{1-\varepsilon}{2}$ and rate $R = \Omega(\varepsilon^{2+o(1)})$, even with efficient algorithms (see Section 1.1). However, there are still open questions. For example, we do not know how to attain $\delta = \frac{1-\varepsilon}{2}$ and $R = \Omega(\varepsilon^2)$ (without any $o(1)$ term) explicitly, and we do not have explicit constructions approaching the GV bound with rates bounded away from zero. Motivated by these questions, we consider *concatenated codes*, possibly with some randomness, which we discuss next.

Concatenated Codes, and Our Question. A natural candidate for explicit (for low randomness) codes on the GV bound are *concatenated linear codes*. These codes are built out of two ingredients: a (hopefully explicit) linear outer code $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ with dimension k for some large q; and a smaller inner binary linear code $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$, with dimension $k_0 = \log_2 q$.

[1] Here and throughout the paper, we will abuse notation and use $\mathcal{C}$ both as the code itself (a subset of Σ^n) and also as an encoding map $\mathcal{C} \colon \Sigma^k \to \Sigma^n$.

We define the concatenated code $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0 \cdot n}$ by first encoding a message $m \in \mathbb{F}_q^k$ (which can also be thought of as $m \in \mathbb{F}_2^{k_0 \cdot k}$) with $\mathcal{C}_{\text{out}}$. Then, we encode each symbol of the resulting codeword using $\mathcal{C}_{\text{in}}$. That is, for a message m,

$$\mathcal{C}(m) = (\mathcal{C}_{\text{in}}(\mathcal{C}_{\text{out}}(m)_1), \mathcal{C}_{\text{in}}(\mathcal{C}_{\text{out}}(m)_2), \cdots, \mathcal{C}_{\text{in}}(\mathcal{C}_{\text{out}}(m)_n)) \in \mathbb{F}_2^{n_0 \cdot n}.$$

It is not hard to see that the rate of $\mathcal{C}$ is the product of the rates of $\mathcal{C}_{\text{in}}$ and $\mathcal{C}_{\text{out}}$, and that the distance of $\mathcal{C}$ is *at least* the product of the distances of $\mathcal{C}_{\text{in}}$ and $\mathcal{C}_{\text{out}}$.

The natural approach to constructing a good concatenated code is to choose $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$ with the best known trade-offs: Since $\mathcal{C}_{\text{out}}$ is over a large alphabet, we know explicit constructions of codes with optimal rate-distance trade-off[2]; and if n_0 is sufficiently small, we can find a $\mathcal{C}_{\text{in}}$ on the GV bound either deterministically by brute force or else with low randomness, depending on the size of n_0.

However, in general this approach will not achieve the GV bound. If we do not assume any additional properties of $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$, and simply use the concatenation properties, then setting the parameters so that $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ has distance $\frac{1-\varepsilon}{2}$, the rate of $\mathcal{C}$ will be at most roughly ε^3. This is known as the *Zyablov bound* [Zya71] (see also [GRS]). As we discuss more in Section 1.1, concatenation has been a popular approach to obtain fully explicit codes with good rate-distance trade-offs, but none of these constructions are known to beat the Zyablov bound.

Instead of using a *single* inner code, several works have focused on a related construction originally due to Thommesen [Tho83], which uses multiple inner codes. More precisely, this construction uses i.i.d. random linear inner codes for each coordinate. It can be shown [Tho83] that the resulting code does lie on the GV bound with high probability, and if $\mathcal{C}_{\text{out}}$ is chosen appropriately there are even efficient decoding algorithms for it [GI04, Rud07, HRZW19]. However, this approach relies heavily on the fact that the inner codes are independent, and as a result uses a lot of randomness.

This state of affairs motivates the following question (also asked in the title of this paper):

Question 1. *Are there concatenated linear codes $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ (with a* single *random linear inner code $\mathcal{C}_{\text{in}}$) that meet the GV bound with high probability over $\mathcal{C}_{\text{in}}$?[3] If so, are there sufficient conditions on $\mathcal{C}_{\text{out}}$ that will guarantee this?*

In this paper, we show that *yes*, there are concatenated codes that meet the GV bound, and we also give two sufficient conditions on $\mathcal{C}_{\text{out}}$ for this to hold. Our existential result is non-constructive, but it is our hope that our sufficient conditions will lead to explicit constructions of appropriate $\mathcal{C}_{\text{out}}$-s, which would lead to explicit (or at least pseudo-random, depending on the alphabet size of $\mathcal{C}_{\text{out}}$) concatenated codes on the GV bound.

[2]For codes over large alphabets, the best possible trade-off is the *Singleton bound*, or $R = 1 - \delta$. This is achievable, for example, by Reed–Solomon codes.

[3]Of course, if the length of either the inner code or the outer code is 1, the this question reduces to the non-concatenated setting; we are interested in parameter regimes where n_0 is non-trivial.

Remark 1 (Motivation for Question 1)**.** *Above, we have motivated Question 1 as an avenue towards explicit or pseudo-random binary codes on the GV bound, and indeed this is our original motivation. But we point out that Question 1 is also interesting in its own right. Concatenated codes are a classical construction, going back to the 1960's [For65], and have been used in many different settings over the decades. It seems like a fundamental question to understand when these codes can attain the GV bound.*

Remark 2 (Focus on Linear Codes)**.** *In Question 1 and in this paper, we focus on* linear *codes. This is because if we used, say, a uniformly random non-linear code as the inner code, it would require exponentially more randomness than a random linear inner code, so this does not seem like a hopeful avenue for derandomization. We note however that the question is much easier for non-linear codes. For example, suppose that $\mathcal{C}_{\text{out}}$ is a Reed–Solomon code of rate ε so that each symbol is additionally tagged with its evaluation point: that is, the symbol corresponding to $\alpha \in \mathbb{F}_q$ is $(\alpha, f(\alpha)) \in \mathbb{F}_q^2$. For the inner code, we use a completely random (non-linear) code of rate ε. Then since all of the symbols in each outer codeword are different by construction, each codeword is essentially uniformly random, and it is not hard to show that the result is close to the GV bound in the sense that a code of rate $O(\varepsilon^2)$ will have distance $1/2 - O(\varepsilon)$ with high probability. This same argument will not work when $\mathcal{C}_{\text{in}}$ is linear, since the different symbols of codewords of $\mathcal{C}_{\text{out}}$ will still have $\mathbb{F}_2$-linear relationships.*

Our Contributions. Our main results are:

1. **Existence of concatenated codes on the GV bound.** We answer the first part of Question 1: there *are* concatenated codes $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ that achieve the GV bound, in a wide variety of parameter regimes. In particular, we show that *most* codes $\mathcal{C}_{\text{out}}$ are actually good:

   Theorem 1.1 (Informal; Theorem 4.2)**.** *Suppose that $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ and $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ are random linear codes of rate ε, so that $q \geq 2^{\Omega(\varepsilon^{-3})}$. Then $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ has rate ε^2, and with high probability, the relative distance of $\mathcal{C}$ is at least $1/2 - O(\varepsilon)$.*

   While Theorem 1.1 seems intuitive (in the sense that a random linear code lies on the GV bound with high probability, so why not concatenated random linear codes?), to the best of our knowledge it has not appeared in the literature before, and the proof was not obvious (to us). One challenge is that if $\mathcal{C}_{\text{out}}$ is a random linear code, then the symbols of any codeword $c \in \mathcal{C}_{\text{out}}$ are not independent, and the inner codewords that appear in the concatenation will have complicated relationships with each other. In particular, this means that the natural strategy of "show that each non-zero codeword has high weight with high probability and union bound" that is used to establish the Gilbert–Varshamov bound will not work in this setting, as we do not have enough concentration.

2. **Sufficient conditions for $\mathcal{C}_{\text{out}}$.** Our existence result above uses a random linear code as the outer code, which does not help in the quest for explicit constructions. However, our proof techniques inspire two sufficient conditions on $\mathcal{C}_{\text{out}}$. That is, if $\mathcal{C}_{\text{out}}$

satisfies these conditions, then $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ will meet the GV bound with high probability when $\mathcal{C}_{\text{in}}$ is a random linear code. Our hope is that formalizing these will lead to explicit constructions in the future.

We give an overview and intuition for our two sufficient conditions here. We note that both conditions are only sufficient when the alphabet size q for $\mathcal{C}_{\text{out}}$ is suitably large (exponential in $1/\operatorname{poly}(\varepsilon)$); see Theorem 5.1 and Theorem 6.2 for details.

- **Sufficient Condition 1: A soft-decoding-like condition on $\mathcal{C}_{\text{out}}^{\perp}$.** Our first sufficient condition, formalized in Theorem 5.1, is a soft-list-decoding-like condition on $\mathcal{C}_{\text{out}}^{\perp}$. More precisely, we define a distribution $\mathcal{D}$[4] on the alphabet $\mathbb{F}_q$; the condition is that

$$\Pr_{x \sim \mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}] \leq \frac{1}{q^k}(1+\Delta) \tag{1}$$

  for some small Δ. Note that $1/q^k$ is the probability that a completely random vector is in $\mathcal{C}_{\text{out}}^{\perp}$, so this condition is saying that if the coordinates of x are drawn i.i.d. from the same distribution $\mathcal{D}$, then x not much more likely to be in $\mathcal{C}^{\perp}$ than in a uniformly random vector. We show that if this holds, then $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ lies on the GV bound with high probability over the choice of a random linear inner code $\mathcal{C}_{\text{in}}$.

  It's not hard to see (Remark 8) that this condition holds in expectation for a random linear code $\mathcal{C}_{\text{out}}$, and in particular there exist linear codes $\mathcal{C}_{\text{out}}$ that have this property.

  This condition is reminiscent of $\mathcal{C}_{\text{out}}^{\perp}$ being list-decodable from soft information (e.g., [KV03]). In soft-list-decoding, one typically gets a distribution $\mathcal{D}_i$ for each $i \in [n]$, interpreted as giving "soft information" about the i'th symbol. If one can show that a vector drawn from $\mathcal{D}_1 \times \cdots \times \mathcal{D}_n$ is unlikely to be in the code, this implies that there are not too many codewords that are likely given the soft information we hare received. However, there are several differences between existing work on soft list-decoding and our work, notably that our distribution $\mathcal{D}$ is a particular one and is the same for all i, and also there are some differences in the parameter settings.

  This condition can also be seen as a soft form of *list-recovery*, where we have the same list in each coordinate.[5] In more detail, if the support of $\mathcal{D}$ is concentrated on a small set S (which ours is for reasonable settings of n_0, ε, see Remark 7), then the condition in Theorem 5.1 is related to asking that the number of codewords that lie in the combinatorial rectangle given by $S \times S \times \cdots \times S$ is about what it should be. Unfortunately, the definition of "small" here does

[4] The distribution $\mathcal{D}$ is intuitively defined as follows. Let $\mathcal{C}_{\text{in}}$ be the inner code, and suppose that it has a generator matrix $G_0 \in \mathbb{F}_2^{n_0 \times k_0}$. Then to sample from $\mathcal{D}$, we take a random sparse linear combination of the rows of G_0 (over $\mathbb{F}_2$), and interpret the result in $\mathbb{F}_2^{k_0}$ as an element of $\mathbb{F}_q$, which we return.

[5] Informally, a code $\mathcal{C} \subseteq \Sigma^n$ is said to be list-recoverable if for any small sets $S_1, \ldots, S_n \subseteq \Sigma$, there are not too many codewords $c \in \mathcal{C}$ so that $c_i \in S_i$ for many values of i.

not seem to be small enough for existing constructions of list-recoverable codes (for example folded RS codes or multiplicity codes) to yield any results.

- **Sufficient Condition 2: $\mathcal{C}_{\text{out}}$ has good min-entropy.** Our second sufficient condition, formalized in Theorem 6.2, requires the codewords of $\mathcal{C}_{\text{out}}$ to be "smooth", meaning, roughly, that every nonzero codeword has a fairly uniform distribution of symbols from $\mathbb{F}_q$. To illustrate why a smoothness condition is desirable, let us consider two extreme cases.

  The bad extreme is when there exists a codeword c that is supported on very few symbols, say even on a single symbol. If $c = (\sigma, \sigma, \ldots, \sigma)$ for some $\sigma \in \mathbb{F}_q$, then the relative weight of $c \circ \mathcal{C}_{\text{in}}$, for a random binary inner code $\mathcal{C}_{\text{in}}$ of rate ε, might be $\frac{1}{2} - \Omega(\sqrt{\varepsilon})$, much worse than the $\frac{1}{2} - O(\varepsilon)$ that we would want for the GV bound.

  The good (possibly unrealistic) extreme is where each nonzero codeword of $\mathcal{C}_{\text{out}}$ has a symbol distribution that is *uniform* over $\mathbb{F}_q$. In this case it is not hard to see that $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ will be close to the GV bound with high probability over a random linear code $\mathcal{C}_{\text{in}}$. (For this, all we need is that $\mathcal{C}_{\text{in}}$ has about the "right" weight distribution, which a random linear code will have with high probability).

  The natural question is thus *how smooth* the codewords of $\mathcal{C}_{\text{out}}$ should be in order for $\mathcal{C}$ to have distance $\frac{1}{2} - O(\varepsilon)$. In Section 6, we quantify this by the *smooth min-entropy* of the codewords' empirical distributions on symbols. We show in Theorem 6.2 that if this smooth min-entropy is large enough for all $c \in \mathcal{C}_{\text{out}}$, then $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is likely to lie near the GV bound when $\mathcal{C}_{\text{in}}$ is a random linear binary code.

  How large is "large enough"? For this informal discussion, we give one example of the parameter settings from Theorem 6.2: It is enough for every non-zero codeword $c \in \mathcal{C}_{\text{out}}$ to have a symbol distribution that has $\Theta(\varepsilon n)$ copies of the same symbol (say, the zero symbol), while the remaining symbols in c are uniformly distributed over a set of size only $q^{1-\varepsilon}$. By some metrics this is still a fairly "spiky" distribution, but it is "smooth enough" for our purposes.

  Note that while our soft-decoding-like condition considers $\mathcal{C}_{\text{out}}^{\perp}$, our smooth min-entropy condition here considers $\mathcal{C}_{\text{out}}$ itself.

1.1 Related Work

Explicit Concatenated Codes. Concatenation (with a single inner code) has been a common approach to obtain explicit codes close to the GV bound. Here we mention a few such places this comes up. Choosing $\mathcal{C}_{\text{out}}$ to be the Reed–Solomon code, and $\mathcal{C}_{\text{in}}$ to be the Hadamard code, gets a code of length $O(k^2/\varepsilon^2)$ for any dimension k [AGHP92], and replacing Reed–Solomon with the Hermitian code gets length $O((k/\varepsilon)^{5/4})$ [BT13]. Choosing a different AG code for $\mathcal{C}_{\text{out}}$ can result in non-vanishing rate and in fact approach rate ε^3 (see [Ta-17]). Moreover, concatenating Reed–Solomon with the Wozencraft ensemble gives the *Justesen* code [Jus72], having constant relative rate and constant relative distance.

Note that none of these concatenation-based constructions thus far have beat the Zyablov bound.

Non-Concatenation-Based Explicit Constructions. As mentioned above, there have been several breakthroughs in the past few years obtaining explicit constructions of binary codes near the GV bound, and even efficient algorithms for them. In a breakthrough result, Ta-Shma [Ta-17] constructed *explicit* linear codes of relative distance $\frac{1-\varepsilon}{2}$ having rate $\varepsilon^{2+o(1)}$. Ta-Shma's codes are also ε-*balanced*, i.e., $\Delta(x,y) \in \left[\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right]$, and thus give rise to explicit ε-biased sample spaces, which are ubiquitous in pseudorandomness and derandomization. Works that followed gave efficient *decoding* of Ta-Shma codes and their variants [AJQ+20, JQST20, JST21, RR23, JST23] (see also [BD22] for a different, randomized, construction that slightly improves upon the rate of [Ta-17], and admits efficient decoding). We note that these codes are graph-based, and do not in general have a concatenated structure.

Results with Multiple i.i.d. Inner Codes. Thommesen showed that when the outer code is a Reed–Solomon code, and it is concatenated with n different random linear codes, one for each coordinate, chosen independently, then the resulting code lies on the GV bound with high probability [Tho83]. Guruswami and Indyk devised efficient decoding algorithms for these codes, based on *list-recoverability* of the outer code [GI04]. That work used a Reed–Solomon code as the outer code, which is list-recoverable up to the Johnson bound. Later, Rudra [Rud07] observed that the parameters could be improved by swapping out the Reed–Solomon code for a code that can be list-recovered up to capacity, for example a Folded Reed–Solomon code. Later work obtained nearly-linear-time decoding algorithms by swapping out the outer code for a capacity-achieving list-recoverable code with near-linear-time list-recovery algorithms [HRZW19, KRRZ+20].

We also mention the work of Guruswami and Rudra [GR10], who show that the same construction (a list-recoverable code concatenated with n different i.i.d. random linear codes) is list-*decodable* up to capacity with high probability. In the results [GI04, Rud07, HRZW19, KRRZ+20] mentioned above, list-recovery of the outer code was needed for *algorithms*, not the combinatorial result (which follows already from [Tho83]). In contrast, in [GR10], the list-recoverability of the outer code is needed for the combinatorial result itself. In that sense, the flavor is similar to our sufficient condition in Section 5, although the techniques are very different, and in our work we only use one inner code.

Further Low-randomness Constructions of Binary Codes on GV Bound. If one's goal is to explicitly construct a binary code that achieves that GV bound, at least two types of partial results may be considered as subgoals. In the first class of results, one seeks explicit codes whose rate vs. distance tradeoff is as close to the GV bound as possible. This includes the works discussed in the first two paragraphs of Section 1.1 above. A second path is to seek codes that fully attain the GV bound, and strive to minimize the amount of randomness used in their construction.

Varshamov's classic result [Var57] is that a random linear code likely achieves the GV bound. Constructing such a code of length n and rate R requires sampling either a random generating matrix or a random parity-check matrix, and thus $O(\min\{R, 1-R\} \cdot n^2)$ random bits are needed. Two classical elementary constructions—the Wozencraft ensemble [Mas63] and the random Toeplitz Matrix construction (e.g., [GRS, Exercise 4.6]) —are able to reduce the needed randomness to $O(n)$.

So far, no codes achieving the GV bound using $o(n)$ randomness are known. Moreover, there is a certain natural obstacle, which we now describe, that needs to be tackled before sublinear randomness can be achieved. Say that a random code $\mathcal{C} \subseteq \mathbb{F}_2^n$ is *uniform* if every $x \in \mathbb{F}_2^n \setminus \{0\}$ appears in the code with the same probability, namely, $p_{R,n} = \frac{2^{Rn}-1}{2^n-1}$. It is not hard to prove via a union bound that a uniform linear code achieves the GV bound with high probability (this is exactly Varshamov's observation). To the best of our knowledge, every known GV-bound construction to date, including the linear randomness constructions mentioned above, is uniform. Unfortunately, a uniform code ensemble with sublinear randomness cannot exist as long as R is bounded away from 1. Indeed, to have events that occur with probability $p_{R,n}$, at least $\log_2 \frac{1}{p_{R,n}} \approx (1-R)n$ random bits are required. Therefore, a code construction obtaining the GV bound with sublinear randomness would have to do so without being uniform (see also [MRSY24, Section 5]). We have hope that our sufficient conditions in Theorems 5.1 and 6.2 could be attained by non-uniform codes. For example, as discussed above, the soft-decoding-like condition of Theorem 5.1 is reminiscent of results on soft-list-decoding and soft-list-recovery, which in different parameter regimes can even be achieved by deterministic codes.

A related line of work [GM22, PP24, MRSY24] attempts to construct codes that enjoy a broad class of desirable combinatorial properties similar to those of random linear codes using as little randomness as possible. Such properties include not just the GV bound, but also list decodability up to the *Elias bound* (see [MRSY24]), list recoverability, and, more generally, *local similarity* (see [MRSY24, Definition 2.14]) to a random linear code.

1.2 Technical Overview

In this section we give an overview of the main technical ideas. This section also serves as an outline of the paper.

Section 3: A moment-based framework. In Section 3, we set up a framework that will be useful for the results in Section 4 and Section 5. We describe this approach here.

Suppose that we are trying to encode a message $m \in \mathbb{F}_q^k$ with our concatenated code $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$, to obtain $\mathcal{C}(m) = w \in \mathbb{F}_2^{n \cdot n_0}$. Each symbol of w is indexed by some $\alpha \in [n]$ and some $\beta \in [n_0]$; this symbol is equal to

$$(\mathcal{C}_{\text{in}}(\mathcal{C}_{\text{out}}(m)_\alpha))_\beta = \langle \mathcal{C}_{\text{out}}(m)_\alpha, b_\beta \rangle,$$

where b_β is the β'th row for a generator matrix $G_0 \in \mathbb{F}_2^{n_0 \times k_0}$ for $\mathcal{C}_{\text{in}}$, and where the $\langle \cdot, \cdot \rangle$ notation denotes the dot product over $\mathbb{F}_2$. This motivates the definition of a variable $X_m \in$

$\mathbb{R}$ defined by

$$X_m = \sum_{\alpha \in [n]} \sum_{\beta \in [n_0]} (-1)^{\langle \mathcal{C}_{\text{out}}(m)_\alpha, b_\beta \rangle}.$$

Indeed, X_m is the bias of $w = \mathcal{C}(m)$; the weight of w is at least $\frac{1}{2} - O(\varepsilon N)$ if and only if X_m is at most $O(\varepsilon N)$. Thus, to show that the code $\mathcal{C}$ has distance at least $\frac{1}{2} - O(\varepsilon N)$, it suffices to show that

$$\max_{m \in \mathbb{F}_q^k \setminus \{0\}} X_m = O(\varepsilon N).$$

Our strategy will be to consider a large moment of X_m over the choice of a random nonzero message m:

$$\mathbb{E}_{m \sim \mathbb{F}_q^k \setminus \{0\}}[X_m^r]$$

for some appropriate r. If we can show that this is smaller than $(c\varepsilon N)^r / q^k$, then Markov's inequality will imply that

$$\Pr_{m \sim \mathbb{F}_q^k \setminus \{0\}}[X_m \geq c\varepsilon N] \leq \frac{\mathbb{E}_{m \sim \mathbb{F}_q^k \setminus \{0\}}[X_m^r]}{(c\varepsilon N)^r} < \frac{1}{q^k},$$

and in particular that there are no messages m so that $X_m \geq c\varepsilon N$.

In Lemma 3.3, we take a Fourier transform in order to re-write $\mathbb{E}[X_m^r]$ as a quantity involving $\mathcal{C}_{\text{out}}^\perp$. This quantity can be thought of as follows. For every integer-valued matrix[6] $V \in \mathbb{Z}_{\geq 0}^{n_0 \times n}$ with entries that sum to r, we consider a vector $g_V \in \mathbb{F}_q^n$ defined by considering the matrix $G_0^T \cdot V \in \mathbb{F}_2^{k_0 \times n}$ and then treating it as a vector $g_V \in \mathbb{F}_q^n$ by identifying each of the columns in $\mathbb{F}_2^{k_0}$ with elements of $\mathbb{F}_q$. Then the quantity in Lemma 3.3 has to do with the number of these vectors g_V that are in $\mathcal{C}_{\text{out}}^\perp$. The exact expression doesn't matter too much for this informal discussion; instead we explain below how we use this re-writing to prove Theorem 4.2 and Theorem 5.1.

Section 4: Most codes $\mathcal{C}_{\text{out}}$ are good. Theorem 4.2 informally says that if $\mathcal{C}_{\text{out}}$ is a random linear code, then with high probability $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is near the GV bound. In the proof, we use our framework from Section 3, and show that with high probability over $\mathcal{C}_{\text{out}}$, the moment $\mathbb{E}_m[X_m^r]$ is small for an appropriate r. To do this, we need to count the number of matrices V described above that are likely to land in $\mathcal{C}_{\text{out}}^\perp$. Since $\mathcal{C}_{\text{out}}$ is a random linear code, so is $\mathcal{C}_{\text{out}}^\perp$, and so the probability of any particular *non-zero* g_V landing in it is small (about $1/q^k$), while of course the probability that 0 is contained in $\mathcal{C}_{\text{out}}^\perp$ is 1. Thus, the challenge is understanding how many g_V-s are actually zero. There are two ways that a matrix V as described above could lead to $g_V = 0$: Either $V = 0 \bmod 2$, or else V is non-zero mod 2 but $G_0^T V = 0$. The first case can be counted straightforwardly. For the second, we leverage the weight distribution that the inner code $\mathcal{C}_{\text{in}}$ is likely to have. We note that this is the only place (in any of our arguments) that we need $\mathcal{C}_{\text{in}}$ to be a random linear code: We just need it to have approximately the "right" weight distribution.

[6] In the actual quantity, the entries of this matrix are ordered, and we denote it $\mathcal{V}$ instead of V; we ignore the ordering in this discussion for simplicity.

Section 5: A soft-decoding-like sufficient condition. The expression that we get for $\mathbb{E}_m[X_r^m]$ in Lemma 3.3 directly inspires our soft-decoding-like sufficient condition in Theorem 5.1. One can view the task of counting the matrices V so that $g_V \in \mathcal{C}_{\text{out}}^\perp$ as choosing a random V and asking about the probability that $g_V \in \mathcal{C}_{\text{out}}^\perp$. If the columns of V were independent, then this would be the same as choosing the coordinates of g_V i.i.d. from some distribution $\mathcal{D}$. Thus we would get a requirement on $\Pr_{x\sim\mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^\perp]$, similar to the condition in Equation (1) that we end up with.

Of course, the coordinates are not independent (because the total weight of V is fixed to be r), but this can be solved. In more detail, we choose r to be a Poisson random variable, which in this setting makes the columns of V independent. One hiccup is that the "Poisson-ized" distribution turns out to be meaningfully different than the original distribution, in the sense that it is much more likely that $g_V = 0$ in the Poisson-ized version. This means that the "natural" soft-decoding-like condition that one would get out of this is not realizable: The probability that $g_V \in \mathcal{C}_{\text{out}}^\perp$ is much bigger than we want it to be, for *any* $\mathcal{C}_{\text{out}}$, just because g_V is too likely to be zero. Fortunately, this seems to be the only obstacle: as in Equation (1), we separate out the $g_V = 0$ term (using the analysis from Section 4) to arrive at a condition that *is* realizable. We explain why the condition is realizable—that is, why there exists a $\mathcal{C}_{\text{out}}$ that meets it—in Remark 8.

Section 6: A smoothness condition on $\mathcal{C}_{\text{out}}$. For our second sufficient condition, we depart from our moment-based framework and work from first principles. Our main theorem in Section 6 is Theorem 6.2, which informally says that if the elements of $\mathcal{C}_{\text{out}}$ have "smooth" enough distributions of symbols, in the sense that they each have large enough min-entropy, that $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ will lie near the GV bound with high probability. The basic idea is to consider a *worst-case* assignment of symbols in $\mathbb{F}_q$ to codewords in $\mathcal{C}_{\text{in}}$; this assignment need not be linear and can depend on a particular codeword $c \in \mathcal{C}_{\text{out}}$. Such a worst-case assignment would simply assign the lowest-weight codewords in $\mathcal{C}_{\text{in}}$ to the most frequent symbols in a codeword $c \in \mathcal{C}_{\text{out}}$. Using the weight distribution that $\mathcal{C}_{\text{in}}$ is likely to have, along with the min-entropy assumption, we can show that this worst-case assignment will *still* result in codewords $w \in \mathcal{C}$ of weight at least $\frac{1}{2} - O(\varepsilon)$.

We note that, unlike our sufficient condition from Section 5, we don't have a proof of feasibility for our smoothness condition. That is, as far as we know, there may not be any linear code $\mathcal{C}_{\text{out}}$ that is smooth in this sense. However, as a proof of concept we mention in Remark 9 that a random linear code will have a similar property with high probability. Moreover, we find it plausible that codewords of *algebraically structured* codes (say, Folded Reed–Solomon codes, Folded Multiplicity, or even large sub-codes of plain Reed–Solomon codes), would satisfy this property, even if a random code does not.

2 Preliminaries

Notation. For a vector $x \in \mathbb{F}^n$, and $\alpha \in [n]$, we use x_α to denote that α'th entry of x. For $x, y \in \mathbb{F}_2^N$, we define $\langle x, y\rangle = \sum_{\alpha=1}^N x_\alpha y_\alpha \in \mathbb{F}_2$. For fields $\mathbb{F}_q$ where $q > 2$ is a power

of 2, we use $\langle x, y \rangle = \mathrm{Tr}(\sum_\alpha x_\alpha y_\alpha)$, where $\mathrm{Tr}\colon \mathbb{F}_q \to \mathbb{F}_2$ is the *field trace* defined below. (In particular it is *not* the standard dot product over $\mathbb{F}_q$!) The reason for this is explained later, but informally it is because we will think of elements of $\mathbb{F}_q$ as vectors in $\mathbb{F}_2^{k_0}$ when $q = 2^{k_0}$, and our notation matches this.

Codes, Linear Codes, and Random Linear Codes. For a finite field $\mathbb{F}$, a code $\mathcal{C} \subseteq \mathbb{F}^n$ is called *linear* if $\mathcal{C}$ is a linear subspace of $\mathbb{F}^n$. The dimension of $\mathcal{C}$ as a subspace is called the *dimension* of the code. For a linear code $\mathcal{C} \subseteq \mathbb{F}^n$, we define the dual code $\mathcal{C}^\perp \subseteq \mathbb{F}^n$ by

$$\mathcal{C}^\perp = \left\{ g \in \mathbb{F}^n \,:\, \sum_{\alpha \in [n]} g_\alpha \cdot c_\alpha = 0 \; \forall c \in \mathcal{C} \right\}.$$

We say that a code $\mathcal{C} \subseteq \mathbb{F}^n$ is a *random linear code* of dimension k if $\mathcal{C}$ is chosen uniformly among all subspaces of $\mathbb{F}_2^n$ of dimension k.

Gilbert–Varshamov Bound. In this paper, we study codes that approach the *Gilbert–Varshamov Bound*, which states that there exist codes of distance δ and rate R approaching $1 - h_2(\delta)$. Here, h_2 is the binary entropy function, given by

$$h_2(x) = x \log\left(\frac{1}{x}\right) + (1-x)\log_2\left(\frac{1}{1-x}\right).$$

Theorem 2.1 (GV Bound, [Gil52, Var57])**.** *Let $\delta \in [0, 1/2)$ and let $\eta \in (0, 1 - h_2(\delta)]$. Then for any $n > 1/\eta$, there exists a (linear) code code $\mathcal{C} \subseteq \mathbb{F}_2^n$ with rate*

$$R \leq 1 - h_2(\delta) - \eta$$

and relative distance at least δ.

We are interested in the parameter regime where the rate R of the code is very small and the distance δ is very large. More precisely, we will focus on the setting where $\delta = 1/2 - O(\varepsilon)$. It is not hard to see (for example, from the Taylor expansion of the entropy function) that in this case

$$1 - h_2(1/2 - \varepsilon) = \Theta(\varepsilon^2)$$

as $\varepsilon \to 0$. Thus, our goal will be the following.

Goal 1 (What we mean by "approaching the GV bound" for low-rate codes)**.** *In this paper, we say that a family of low-rate codes $\mathcal{C}_{N,\varepsilon} \subseteq \mathbb{F}_2^N$ "approaches the GV bound" if $\mathcal{C}_{N,\varepsilon}$ has rate $\Omega(\varepsilon^2)$ and distance $1/2 - O(\varepsilon)$, where the asymptotic notation is as $\varepsilon \to 0$ and as $N \to \infty$.*

Concatenated Codes and Our Default Parameters. Throughout the paper, we use $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ and $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ as our outer and inner (linear) codes, both of rate ε. We will let $k = \varepsilon n$ and $k_0 = \varepsilon n_0 = \log(q)$ throughout. As mentioned in the introduction, we sometimes abuse notation and write $\mathcal{C}_{\text{out}} \colon \mathbb{F}_q^k \to \mathbb{F}_q^n$ and $\mathcal{C}_{\text{in}} \colon \mathbb{F}_2^{k_0} \to \mathbb{F}_2^{n_0}$ to represent an arbitrary encoding map for these codes.

Let $N = n_0 \cdot n$ and $K = k_0 \cdot k$. Abusing notation as noted above, the concatenated code $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^N$ is given by the encoding map $\mathcal{C} \colon \mathbb{F}_2^K \to \mathbb{F}_q^N$ defined as follows. For a message $m \in \mathbb{F}_2^K$, we interpret m as an element of $\mathbb{F}_q^k$ and let $c = \mathcal{C}_{\text{out}}(m)$; then $\mathcal{C}(m)$ is defined as

$$\mathcal{C}(m) = (\mathcal{C}_{\text{in}}(c_1), \mathcal{C}_{\text{in}}(c_2), \cdots, \mathcal{C}_{\text{in}}(c_n)) \in \mathbb{F}_2^N.$$

As mentioned in the introduction, we'll take our inner code $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ to be a random linear code of dimension $k_0 = \varepsilon n_0$. The important property we will need of $\mathcal{C}_{\text{in}}$ is that it have about the right weight distribution, which we formalize in the following property.

Definition 2.2 (τ-niceness of the inner code)**.** *Fix parameters $0 < \tau < \varepsilon$. We say that the inner code $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{k_0}$ is* τ-nice *if for any $i \in \{1, \ldots, n_0\}$, the number of $c \in \mathcal{C}_{\text{in}}^{\perp}$ so that* $\mathsf{weight}(c) = i$ *is at most*

$$\binom{n_0}{i} \cdot 2^{-n_0(\varepsilon - \tau)}.$$

(Notice that we omit the "ε" from the name "τ-nice," because ε will be the same ε throughout the paper; it is the rate of both $\mathcal{C}_{\text{in}}$ and $\mathcal{C}_{\text{out}}$).

Lemma 2.3. *Let $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ be a random linear code of dimension $k_0 = \varepsilon n_0$. Then with probability at least $1 - n_0 \cdot 2^{-\tau n_0}$, $\mathcal{C}_{\text{in}}$ is τ-nice.*

Proof: Let E_i denote the event that $\mathcal{C}_{\text{in}}^{\perp}$ has at most $\binom{n_0}{i} \cdot 2^{-n_0(\varepsilon-\tau)}$ codewords of weight i. Note that the expected number of such codewords is at most $\binom{n_0}{i} 2^{-n_0 \varepsilon}$, so by Markov's inequality,

$$\Pr_{\mathcal{C}_{\text{in}}}[E_i] \geq 1 - 2^{-\tau n_0}.$$

By a union bound,

$$\Pr[\mathcal{C}_{\text{in}} \text{is } \tau\text{-nice}] \geq 1 - n_0 2^{-\tau n_0},$$

as claimed. ∎

We also need the following observation about random binary linear codes.

Lemma 2.4 (Negative correlation of $x \in \mathcal{C}$ and $y \in \mathcal{C}$)**.** *Let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be a random linear code and let $x, y \in \mathbb{F}_2^n \setminus \{0\}$ such that $x \neq y$. Then,*

$$\Pr_{\mathcal{C}}[x, y \in \mathcal{C}] \leq \Pr_{\mathcal{C}}[x \in \mathcal{C}] \cdot \Pr_{\mathcal{C}}[y \in \mathcal{C}].$$

Proof: A random linear code of dimension Rn contains a given d-dimensional linear subspace of $\mathbb{F}_2^n$ with probability $\prod_{i=0}^{d-1} \frac{2^{Rn}-2^i}{2^n-2^i}$. Since x and y span a 2-dimensional space, we have

$$\Pr[x, y \in \mathcal{C}] = \frac{\left(2^{Rn}-1\right)\cdot\left(2^{Rn}-2\right)}{(2^n-1)\cdot(2^n-2)} \leq \left(\frac{2^{Rn}-1}{2^n-1}\right)^2 = \Pr[x \in \mathcal{C}]\cdot\Pr[y \in \mathcal{C}].$$

■

$\mathbb{F}_q$ as a vector space over $\mathbb{F}_2$, and Fourier Analysis. Let $q = 2^{k_0}$ be a power of 2; we will set q like this for the rest of the paper. It is not hard to see that $\mathbb{F}_q$ is a vector space over $\mathbb{F}_2$ of dimension k_0. In particular, we can identify elements of $\mathbb{F}_2^{k_0}$ with $\mathbb{F}_q$ by simply writing elements of $\mathbb{F}_q$ out in any basis of $\mathbb{F}_q$ over $\mathbb{F}_2$. For convenience, we will choose a particular basis, which interacts nicely with the *trace map*, defined by

$$\mathrm{Tr}(x) = x + x^2 + \cdots + x^{2^{k_0-1}}.$$

It is well-known that Tr is both $\mathbb{F}_2$-linear and also that its image is indeed $\mathbb{F}_2$. In that sense, $\mathrm{Tr}(\cdot,\cdot) : \mathbb{F}_q \times \mathbb{F}_q \to \mathbb{F}_2$ behaves a bit like a dot product, and in fact this can be made formal.

In more detail, for any $k_0 > 0$ there always exists a *self-dual* basis $\nu_1, \ldots, \nu_{k_0}$ of $\mathbb{F}_q = \mathbb{F}_{2^{k_0}}$ over $\mathbb{F}_2$ (e.g., [SL80, JMV90]); that is, this basis has the property that $\mathrm{Tr}(\nu_i\nu_j) = \mathbf{1}[i=j]$. In particular, this means that if we choose such a basis in order to identify $\mathbb{F}_q$ with $\mathbb{F}_2^{k_0}$, we have

$$\mathrm{Tr}(\alpha\cdot\beta) = \langle\alpha,\beta\rangle$$

for any $\alpha, \beta \in \mathbb{F}_q \sim \mathbb{F}_2^{k_0}$, where on the left hand side we treat α and β as elements of $\mathbb{F}_q$, and on the right hand side we treat them as elements of $\mathbb{F}_2^{k_0}$.[7]

The reason we want to do our identification between $\mathbb{F}_q$ and $\mathbb{F}_2^{k_0}$ is because it will make the notation a bit easier for *Fourier transforms*. For a function $\varphi\colon \mathbb{F}_q^n \to \mathbb{R}$, we define the Fourier transform of φ, denoted $\hat{\varphi}\colon \mathbb{F}_q^n \to \mathbb{R}$, by

$$\hat{\varphi}(\omega) = \frac{1}{q^n}\sum_{x\in\mathbb{F}_q^n} \varphi(x)(-1)^{\mathrm{Tr}(\langle\omega,x\rangle)}.$$

Notice that if we were to replace the elements of $\mathbb{F}_q$ with their corresponding elements of $\mathbb{F}_2^{k_0}$, and treat $\omega \in \mathbb{F}_q^n$ as $\omega \in \mathbb{F}_2^{k_0 n}$ in the natural way, by the above correspondence we can write this as

$$\hat{\varphi}(\omega) = \frac{1}{2^{k_0 n}}\sum_{x\in\mathbb{F}_2^{k_0 n}} \varphi(x)(-1)^{\langle\omega,x\rangle}.$$

Thus, we will drop the trace notation for the rest of the paper, and just use the above definition. Not only does this simplify the notation for Fourier transforms, but it allows

[7] Indeed, if we write $\alpha = \sum_{i=1}^{k_0}\alpha_i\nu_i$ and $\beta = \sum_{j=1}^{k_0}\beta_j\nu_j$ as elements of $\mathbb{F}_q$, and hence as $(\alpha_1,\ldots,\alpha_{k_0})$ and $(\beta_1,\ldots,\beta_{k_0})$ as elements of $\mathbb{F}_2^{k_0}$, then by $\mathbb{F}_2$-linearity, $\mathrm{Tr}(\alpha\cdot\beta) = \sum_{i,j}\alpha_i\beta_j\mathrm{Tr}(\nu_i\cdot\nu_j) = \sum_i \alpha_i\beta_i = \langle\alpha,\beta\rangle$.

us to move back and forth between elements of $\mathbb{F}_q$ and elements of $\mathbb{F}_2^{k_0}$, which we will want to do anyway as we have to identify elements of $\mathbb{F}_q$ as messages for $\mathcal{C}_{\text{in}}$ in $\mathbb{F}_2^{k_0}$.

Next, we record a few useful facts about the Fourier transform. The first is that the Fourier transform can be inverted: For any function $\varphi\colon \mathbb{F}_q^n \to \mathbb{R}$ and any x,

$$\varphi(x) = \sum_{\omega \in \mathbb{F}_q^n} \hat{\varphi}(\omega)(-1)^{\langle x,\omega\rangle}. \tag{2}$$

The second is that for any $\mathbb{F}_q$-subspace $V \subseteq \mathbb{F}_q^n$, and for any $g \in \mathbb{F}_q^n$, we have

$$\widehat{\mathbf{1}_V}(g) = \frac{|V|}{q^n}\mathbf{1}_{V^\perp}(g). \tag{3}$$

3 A Useful Moment Computation

In this section, we define a random variable X_m that quantifies how "bad" a message m is for our concatenated code, and we prove (Lemma 3.3) that the moments of X_m are well-behaved.

We use the notation set up in Section 2. Namely, we fix $\varepsilon > 0$, an integer n, and a power of two $q = 2^{k_0}$. We consider the concatenated $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ for linear codes $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ of dimension $k = \varepsilon n$ and $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ of dimension $k_0 = \varepsilon n_0$. We let $N = n \cdot n_0$.

In this section, we think of both $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$ as *fixed*. The only randomness will be in choosing a random message $m \in \mathbb{F}_q^k \setminus \{0\}$. Before we begin, we introduce one more definition.

Definition 3.1. *For a code $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ of dimension k_0, let $G_0 \in \mathbb{F}_2^{n_0 \times k_0}$ be an (arbitrary) generator matrix for $\mathcal{C}_{\text{in}}$. That is,*

$$\mathcal{C}_{\text{in}} = \{G_0 \cdot x \,:\, x \in \mathbb{F}_2^{k_0}\}.$$

Let $\Omega \subseteq \mathbb{F}_2^{k_0}$ denote the set of rows of G_0, so $|\Omega| = n_0$. By identifying $\mathbb{F}_2^{k_0}$ with $\mathbb{F}_q$ (as described in Section 2, using a self-dual basis), we may also treat Ω as a subset of $\mathbb{F}_q$. We say that Ω is the set derived from $\mathcal{C}_{\text{in}}$.

Remark 3. *Since G_0 can be an arbitrary generator matrix, there is some freedom in defining the set Ω derived from $\mathcal{C}_{\text{in}}$. It won't matter for the results in this paper, but it may matter for instantiating our sufficient condition in Section 5.*

Next, we define the random variable whose moments we want to bound.

Definition 3.2. *For a message of the outer code $m \in \mathbb{F}_q^k$, define*

$$X_m \triangleq \sum_{\alpha \in [n]} \sum_{b \in \Omega} (-1)^{\langle \mathcal{C}_{\text{out}}(m)_\alpha, b\rangle}.$$

Observe that X_m is the bias (namely, the difference between the zeros and ones) of the codeword $\mathcal{C}(m) = \mathcal{C}_{\text{out}}(m) \circ \mathcal{C}_{\text{in}}$. In particular, if $X_m = O(\varepsilon N)$ for all nonzero m, this will imply that the distance of $\mathcal{C}$ is least $1/2 - O(\varepsilon)$. Thus, our goal will be to show that X_m is small. Towards that end, we will compute the r-th moment of X_m (over the randomness of a random nonzero message m), for a suitable large r. Finally, given an ordered list $\mathcal{V}$ of pairs in $[n] \times \Omega$, for every $\alpha \in [n]$ we denote by $\mathcal{V}_\alpha$ the *multiset* $\mathcal{V}_\alpha = \{b : (\alpha, b) \in \mathcal{V}\}$ (so b appears twice in $\mathcal{V}_\alpha$ if (α, b) appears twice in $\mathcal{V}$).

The following lemma characterizes the r-th moment of X_m in terms of the dual code $\mathcal{C}_{\text{out}}^\perp$.

Lemma 3.3. *Suppose that Ω is derived from $\mathcal{C}_{\text{in}}$. Using the notation above, for every $r \geq 1$, it holds that*

$$\mathop{\mathbb{E}}_{m \sim \mathbb{F}_q^k \setminus \{0\}} [X_m^r] = \frac{1}{q^k - 1} \sum_{\mathcal{V} \in ([n] \times \Omega)^r} \left(q^k \cdot \mathbf{1}[g_{\mathcal{V}} \in \mathcal{C}_{\text{out}}^\perp] - 1 \right),$$

where $\mathcal{V}$ ranges over all tuples $((\alpha_1, b_1), \ldots, (\alpha_r, b_r))$, and $g_{\mathcal{V}} \in \mathbb{F}_q^n$ is such that $(g_{\mathcal{V}})_\alpha = \sum_{b \in \mathcal{V}_\alpha} b$.

Remark 4 ($\mathbb{F}_q$-linear codes over $\mathbb{F}_q^s$.)**.** *Several constructions of potential outer codes (for example, Folded RS codes [GR08] or univariate multiplicity codes [Kop15, GW13]) are not linear over their alphabets but instead are linear over a subfield. That is, the alphabets for these codes are $\mathbb{F}_q^s$ for some $s > 1$, and the codes are $\mathbb{F}_q$-linear. An inspection of the proof shows that Lemma 3.3 still holds in this case: The only thing that changes is that we need to define the dual code $\mathcal{C}_{\text{out}}^\perp$ by first "unfolding" the original code to treat it as a subspace of $(\mathbb{F}_q)^{sn}$, taking the dual, and "re-folding." Intuitively, the proof goes through because the definition of the Fourier transform is the same whether the relevant vectors lie in $(\mathbb{F}_q^s)^n$ or $(\mathbb{F}_q)^{sn}$. Similarly, Lemma 3.3 holds when $\mathcal{C}_{\text{out}}$ is any $\mathbb{F}_2$-linear code (rather than $\mathbb{F}_q$-linear).*

More generally, it is not hard to see that all of the results in the paper go through for $\mathbb{F}_q$-linear codes over $\mathbb{F}_q^s$ (or even any $\mathbb{F}_2$-linear codes) $\mathcal{C}_{\text{out}}$. Indeed, the results in Section 4 and Section 5 essentially rely only on Lemma 3.3; while the results in Section 6 are separate but are already stated for $\mathbb{F}_2$-linear codes and can easily be seen to extend to larger alphabets.

Proof of Lemma 3.3: First, we can write, for any $m \in \mathbb{F}_q^k$,

$$X_m^r = \sum_{\mathcal{V} = \langle (\alpha_1, b_1), \ldots, (\alpha_r, b_r) \rangle} \prod_{j=1}^{r} (-1)^{\langle \mathcal{C}_{\text{out}}(m)_{\alpha_j}, b_j \rangle}$$

Taking the expectation over a random $m \in \mathbb{F}_q^k \setminus \{0\}$, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{m \sim \mathbb{F}_q^k \setminus \{0\}} [X_m^r] &= \frac{1}{q^k - 1} \sum_{m \in \mathbb{F}_q^k \setminus \{0\}} \sum_{\mathcal{V}} \prod_{j=1}^{r} (-1)^{\langle \mathcal{C}_{\text{out}}(m)_{\alpha_j}, b_j \rangle} \\ &= \frac{1}{q^k - 1} \sum_{m \in \mathbb{F}_q^k \setminus \{0\}} \sum_{\mathcal{V}} \prod_{\alpha \in [n]} (-1)^{\left\langle \mathcal{C}_{\text{out}}(m)_\alpha, \sum_{b \in \mathcal{V}_\alpha} b \right\rangle}, \end{aligned} \tag{4}$$

where we use the convention that $\sum_{b\in\mathcal{V}_\alpha} b$ is the zero vector whenever $\mathcal{V}_\alpha$ is empty.

Now, for any $\alpha \in [n]$, let

$$\varphi_\alpha(x) = (-1)^{\left\langle x, \sum_{b\in\mathcal{V}_\alpha} b\right\rangle}.$$

Taking the Fourier transform over $\mathbb{F}_2^{k_0}$, we get for every w,

$$\begin{aligned}\widehat{\varphi_\alpha}(w) &= \frac{1}{q}\sum_{x\in\mathbb{F}_2^{k_0}} (-1)^{\left\langle x, \sum_{b\in\mathcal{V}_\alpha} b\right\rangle}\cdot(-1)^{\langle w,x\rangle} \\ &= \frac{1}{q}\sum_{x\in\mathbb{F}_2^{k_0}} (-1)^{\left\langle x, w+\sum_{b\in\mathcal{V}_\alpha} b\right\rangle} = \mathbf{1}\left[w = \sum_{b\in\mathcal{V}_\alpha} b\right].\end{aligned}$$

Let us abbreviate $\overline{m} = \mathcal{C}_{\text{out}}(m)$. Plugging the above back to Equation (4), and using the inverse Fourier transform, we get

$$\begin{aligned}\mathop{\mathbb{E}}_{m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r] &= \frac{1}{q^k-1}\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}\sum_{\mathcal{V}}\prod_{\alpha\in[n]}\varphi_\alpha(\overline{m}_\alpha) \\ &= \frac{1}{q^k-1}\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}\sum_{\mathcal{V}}\prod_{\alpha\in[n]}\left(\sum_{w\in\mathbb{F}_2^{k_0}}\widehat{\varphi_\alpha}(w)\cdot(-1)^{\langle w,\overline{m}_\alpha\rangle}\right) \\ &= \frac{1}{q^k-1}\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}\sum_{\mathcal{V}}\sum_{g\colon[n]\to\mathbb{F}_2^{k_0}}\left(\prod_{\alpha\in[n]}\widehat{\varphi_\alpha}(g_\alpha)\right)\left(\prod_{\alpha\in[n]}(-1)^{\langle g_\alpha,\overline{m}_\alpha\rangle}\right).\end{aligned}$$

Moving the sum over m inside, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r] &= \frac{1}{q^k-1}\sum_{\mathcal{V}}\sum_{g\in\mathbb{F}_q^n}\left(\prod_{\alpha\in[n]}\widehat{\varphi_\alpha}(g_\alpha)\right)\left(\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}(-1)^{\sum_{\alpha\in[n]}\langle g_\alpha,\overline{m}_\alpha\rangle}\right) \\ &= \frac{1}{q^k-1}\sum_{\mathcal{V}}\sum_{g\in\mathbb{F}_q^n}\left(\prod_{\alpha\in[n]}\widehat{\varphi_\alpha}(g_\alpha)\right)\left(\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}(-1)^{\langle g,\overline{m}\rangle}\right), \qquad (5)\end{aligned}$$

where in the second equation we have treated $g, \overline{m}$ as elements of $\mathbb{F}_2^{k_0 n}$ in the natural way. Next, we observe that

$$\sum_{m\in\mathbb{F}_q^k\setminus\{0\}}(-1)^{\langle g,\overline{m}\rangle} = \begin{cases} q^k-1 & g\in\mathcal{C}_{\text{out}}^\perp, \\ -1 & \text{otherwise.}\end{cases} \qquad (6)$$

Indeed, we have

$$\sum_{m\in\mathbb{F}_q^k}(-1)^{\langle g,\overline{m}\rangle} = q^n\cdot\widehat{\mathbf{1}_{\mathcal{C}_{\text{out}}}}(g) = |\mathcal{C}_{\text{out}}|\cdot\mathbf{1}_{\mathcal{C}_{\text{out}}^\perp}(g) = q^k\cdot\mathbf{1}_{\mathcal{C}_{\text{out}}^\perp}(g),$$

and then we subtract off the zero term, where the penultimate inequality follows from Equation (3).

Thus, given Equation (6), we can write Equation (5) as

$$\mathop{\mathbb{E}}_{m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r] = \frac{1}{q^k-1}\sum_{\mathcal{V}}\sum_{g\in\mathbb{F}_q^n}\left(\prod_{\alpha\in[n]}\widehat{\varphi_\alpha}(g_\alpha)\right)\left(q^k\cdot\mathbf{1}\left[g\in\mathcal{C}_{\text{out}}^\perp\right]-1\right). \tag{7}$$

Recalling our expression for $\widehat{\varphi_\alpha}$, observe that

$$\prod_{\alpha\in[n]}\widehat{\varphi_\alpha}(g_\alpha) = \mathbf{1}\left[\forall\alpha\in[n], g_\alpha = \sum_{b\in\mathcal{V}_\alpha} b\right].$$

Thus, in Equation (7), the only nonzero term in the sum over the g-s is $g_{\mathcal{V}}$, which was indeed defined as $g_{\mathcal{V}}(\alpha) = \sum_{b\in\mathcal{V}_\alpha} b$. We can then conclude that

$$\mathop{\mathbb{E}}_{m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r] = \frac{1}{q^k-1}\sum_{\mathcal{V}}\left(q^k\cdot\mathbf{1}\left[g_{\mathcal{V}}\in\mathcal{C}_{\text{out}}^\perp\right]-1\right).$$

■

4 Most Linear Codes $\mathcal{C}_{\text{out}}$ Work Well

In this section we show that there *exist* low-rate concatenated linear codes approaching the GV bound, and in fact most codes $\mathcal{C}_{\text{out}}$ will work when concatenated with a random linear inner code $\mathcal{C}_{\text{in}}$. In more detail, keeping the notation as Section 2 and Section 3, we show that $\mathcal{C} = \mathcal{C}_{\text{out}}\circ\mathcal{C}_{\text{in}}$ has distance $\frac{1}{2}-O(\varepsilon)$, with high probability, when both $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$ are random linear codes of rate ε. Before our proof, we will set further notation that will be useful in later sections as well.

Definition 4.1. *For a sequence of tuples $\mathcal{V} = ((\alpha_1,b_1),\dots,(\alpha_r,b_r))$, we define $V = V(\mathcal{V}) \in \mathbb{N}^{n\times n_0}$ to be the "unordered" version of $\mathcal{V}$, that is, $V[\alpha,b]$ is the number of times that the pair (α,b) appears in $\mathcal{V}$. Further, let $B = B(\mathcal{V})$ simply be $V \bmod 2$, where the modulo is taken element-wise. We refer to the number of 1-s in the matrix as the* weight *of B, denoted $\|B\|$.*

Theorem 4.2. *There exist constants $c,\bar{c},\tilde{c} > 0$ such that the following holds. Fix any integer $k>0$, any $\varepsilon>0$ sufficiently small (in terms of $\tilde{c}$), and any power-of-two $q = 2^{k_0}$. Let $n_0 = k_0/\varepsilon$ and $n = k/\varepsilon$, and let $N = n_0 n$. Suppose that $q \geq 2^{\tilde{c}/\varepsilon^3}$. Let $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ be a linear code of dimension k_0 that is τ-nice, for $\tau = 1/\sqrt{n_0}$. Let $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ be an independent random linear code of dimension k. Then, with probability at least $1-2^{-\varepsilon^2 N/\bar{c}}$ over the choice of $\mathcal{C}_{\text{out}}$, the relative distance of $\mathcal{C} = \mathcal{C}_{\text{out}}\circ\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^N$ is at least $\frac{1}{2}-c\cdot\varepsilon$.*

Remark 5 (τ-niceness of inner code)**.** *Note that, by Lemma 2.3, a random linear code $\mathcal{C}_{\text{in}}$ is τ-nice for $\tau = 1/\sqrt{n_0}$ with probability at least $1-2^{-\Omega(\sqrt{n_0})}$. Thus, Theorem 4.2 implies that $\mathcal{C}_{\text{out}}\circ\mathcal{C}_{\text{in}}$ approaches the GV bound with high probability over a random $\mathcal{C}_{\text{in}}$ and a random $\mathcal{C}_{\text{out}}$.*

Proof: Pick $r = \varepsilon^2 N$. More precisely, we will choose r to be the smallest even integer that is larger than $\varepsilon^2 N$; we assume without loss of generality that $r = \varepsilon^2 N$, which will only affect the constants in the theorem statement, and will make the computations much more readable. Lemma 3.3 implies that for any fixed $\mathcal{C}_{\text{in}}$ (and thus any fixed Ω, the set derived from $\mathcal{C}_{\text{in}}$, as in Definition 3.1),

$$\mathop{\mathbb{E}}_{m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r] = \frac{1}{q^k-1}\sum_{\mathcal{V}\in([n]\times\Omega)^r}\left(q^k\cdot\mathbf{1}[g_{\mathcal{V}}\in\mathcal{C}_{\text{out}}^{\perp}]-1\right). \tag{8}$$

In order to make the dependence on $\mathcal{C}_{\text{out}}$ and $\mathcal{C}_{\text{in}}$ more explicit, we will write $X_m^r(\mathcal{C}_{\text{out}},\Omega)$.

Our strategy will be to take the expectation of Equation (8) over the randomness in $\mathcal{C}_{\text{out}}$, and use Markov's inequality. To that end, note that for every sequence of tuples $\mathcal{V} = \mathcal{V}(\Omega)$, it holds that

$$\Pr_{\mathcal{C}_{\text{out}}}\left[g_{\mathcal{V}}\in\mathcal{C}_{\text{out}}^{\perp}\right] = \begin{cases} 1 & g_{\mathcal{V}}=0,\\ \frac{q^{n-k}-1}{q^n-1} & \text{otherwise.}\end{cases}$$

This is since $\mathcal{C}_{\text{out}}^{\perp}$ is a random linear code of dimension $n-k$. Therefore, upon taking the expectation over $\mathcal{C}_{\text{out}}$, for any fixed Ω, Equation (8) becomes

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{C}_{\text{out}},m\sim\mathbb{F}_q^k\setminus\{0\}}[X_m^r(\mathcal{C}_{\text{out}},\Omega)] &= \frac{1}{q^k-1}\sum_{\mathcal{V}}\left(q^k\Pr_{\mathcal{C}_{\text{out}}}\left[g_{\mathcal{V}}\in\mathcal{C}_{\text{out}}^{\perp}\right]-1\right)\\ &= \sum_{\mathcal{V}}\left(\mathbf{1}[g_{\mathcal{V}}=0]-\frac{1}{q^n-1}\cdot\mathbf{1}[g_{\mathcal{V}}\neq 0]\right)\\ &\leq \sum_{\mathcal{V}}\mathbf{1}[g_{\mathcal{V}}=0]. \end{aligned}\tag{9}$$

Thus, we are left with bounding the number of $\mathcal{V}$-s for which $g_{\mathcal{V}}=0$, recalling that $(g_{\mathcal{V}})_\alpha = \sum_{b\in\mathcal{V}_\alpha} b$.

Using the notation of Definition 4.1, each $\mathcal{V}$ gives rise to $V(\mathcal{V})\in\mathbb{N}^{n\times n_0}$ and $B(\mathcal{V})\in\mathbb{F}_2^{n\times n_0}$. Note that $g_{\mathcal{V}}=0$ if and only if every row of $B(\mathcal{V})$ is a left kernel vector of $G_0\in\mathbb{F}_2^{n_0\times k_0}$, where G_0 is the generating matrix of $\mathcal{C}_{\text{in}}$. That is, if and only if every row of $B(\mathcal{V})$ belongs to $\mathcal{C}_{\text{in}}^{\perp}$.

We also need the following claim, whose proof we defer.

Claim 4.3. *Fix a linear code $\mathcal{C}_{\text{in}}\subseteq\mathbb{F}_2^{n_0}$ of dimension k_0 that is τ-nice for $\tau = 1/\sqrt{n_0}$. Let $\varepsilon \triangleq \frac{k_0}{n_0}$ and let $r\in\mathbb{N}$. Let $G_0\in\mathbb{F}_2^{n_0\times k_0}$ be a matrix whose left kernel is $\mathcal{C}_{\text{in}}^{\perp}$, and denote*

$$W = \{\mathcal{V}\in([n]\times\Omega)^r : B(\mathcal{V})\cdot G_0 = 0\}.$$

Then

$$|W| \leq (8N)^r\cdot(r/N)^{r/2}\cdot 2^{\frac{2N}{\sqrt{n_0}}+\log_2 N}\cdot\max\left\{1,\left(\frac{re}{\varepsilon^2 N}\right)^{r/2}\right\}.$$

Now we finish the proof, given Claim 4.3. Let $\tau = 1/\sqrt{n_0}$ as in the theorem statement. Let $\mathcal{V} \in ([n]) \times \Omega)^r$ be such that $g_{\mathcal{V}} = 0$. Recall that every row of $B(\mathcal{V})$ must belong to $\mathcal{C}_{\text{in}}^{\perp}$, so, in particular, $\mathcal{V}$ must belong to W. Hence, for a large enough n_0, Equation (9) and Claim 4.3 yield

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{C}_{\text{out}}, m \sim \mathbb{F}_q^k \setminus \{0\}} [X_m^r(\mathcal{C}_{\text{out}}, \Omega)] &\leq \sum_{\mathcal{V}} \mathbf{1}[g_{\mathcal{V}} = 0] \\
&= |W| \\
&\leq (8N)^r \cdot (re/N)^{r/2} \cdot 2^{2N/\sqrt{n_0} + \log_2 N} \\
&\leq (8N)^r \cdot (e\varepsilon^2)^{r/2} \cdot 2^{2N\varepsilon^2/\sqrt{\tilde{c}}} \cdot N \\
&= \left(8\sqrt{e} \cdot 2^{2/\sqrt{\tilde{c}}} \cdot N\varepsilon\right)^r \cdot N \\
&\leq \left(32\sqrt{e} \cdot N\varepsilon\right)^r \cdot N. \qquad (10)
\end{aligned}$$

Above, in the third line we have used the fact that $r = \varepsilon^2 N$ to replace the maximum in Claim 4.3 with $e^{r/2}$; in the fourth line we have plugged in $r = \varepsilon^2 N$ and also the fact that our assumption $q \geq 2^{\tilde{c}/\varepsilon^3}$ implies that $n_0 \geq \tilde{c}/\varepsilon^4$; and in the last line we have assumed without loss of generality that $\tilde{c} \geq 1$.

By Markov's inequality,

$$\begin{aligned}
&\Pr_{\mathcal{C}_{\text{out}}} \left[\exists m \in \mathbb{F}_q^k \setminus \{0\},\ X_m(\mathcal{C}_{\text{out}}, \Omega) > c\varepsilon N\right] \\
&\qquad = \Pr_{\mathcal{C}_{\text{out}}} \left[\exists m \in \mathbb{F}_q^k \setminus \{0\},\ X_m^r(\mathcal{C}_{\text{out}}, \Omega) > (c\varepsilon N)^r\right] \\
&\qquad \leq \sum_{m \in \mathbb{F}_q^k \setminus \{0\}} \Pr_{\mathcal{C}_{\text{out}}} [X_m^r(\mathcal{C}_{\text{out}}, \Omega) > (c\varepsilon N)^r] \\
&\qquad \leq \sum_{m \in \mathbb{F}_q^k \setminus \{0\}} \frac{\mathbb{E}_{\mathcal{C}_{\text{out}}} [X_m^r(\mathcal{C}_{\text{out}}, \Omega)]}{(c\varepsilon N)^r} \\
&\qquad = \frac{\left(q^k - 1\right) \cdot \mathbb{E}_{\mathcal{C}_{\text{out}}, m \sim \mathbb{F}_q^k \setminus \{0\}} [X_m^r(\mathcal{C}_{\text{out}}, \Omega)]}{(c\varepsilon N)^r} \\
&\qquad \leq \frac{2^r \cdot \mathbb{E}_{\mathcal{C}_{\text{out}}, m \sim \mathbb{F}_q^k \setminus \{0\}} [X_m^r(\mathcal{C}_{\text{out}}, \Omega)]}{(c\varepsilon N)^r},
\end{aligned}$$

where in the last line we have used that $q^k = 2^{N\varepsilon^2} = 2^r$. Plugging in (10), we see that this expression is at most

$$\left(\frac{64\sqrt{e} \cdot N\varepsilon}{c\varepsilon N}\right)^r \cdot N = \left(\frac{64\sqrt{e}}{c}\right)^r \cdot N.$$

Thus if we choose $c = 128\sqrt{e}$, we conclude that

$$\begin{aligned}
\Pr_{\mathcal{C}_{\text{out}}} [\exists m \in \mathbb{F}_q^k \setminus \{0\}, X_m(\mathcal{C}_{\text{out}}, \Omega) > c\varepsilon N] &\leq N \cdot 2^{-r} \\
&\leq N \cdot 2^{-N\varepsilon^2}.
\end{aligned}$$

As we are assuming that $n_0 \geq \tilde{c}/\varepsilon^4$, we have that

$$n_0 \leq 2^{n_0 \varepsilon^2}$$

as long as ε is sufficiently small relative to $\tilde{c}$, and in particular we have

$$N = n \cdot n_0 \leq 2^{n \cdot n_0 \varepsilon^2} = 2^{N\varepsilon^2}$$

as we always have that $n \geq 1$. Thus, the above reads

$$\Pr_{\mathcal{C}_{\text{out}}}[\exists m \in \mathbb{F}_q^k \setminus \{0\}, X_m(\mathcal{C}_{\text{out}}, \Omega) > c\varepsilon N] \leq 2^{-N\varepsilon^2/2},$$

which proves the claim after observing that without loss of generality we may take $\bar{c} \geq 2$. ■

Finally, we prove Claim 4.3.

Proof of Claim 4.3: Let $W' = \{V(\mathcal{V}) : \mathcal{V} \in W\}$. Since $V(\mathcal{V})$ preserves all the information in $\mathcal{V}$ up to ordering, we have $|W| \leq r! \cdot |W'|$. We turn to bound $|W'|$. Let $V \in W'$. Write $V = B + E$ where B is a $\{0,1\}$ matrix and E is a matrix whose entries are all even. By abuse of notation we also think of B as a matrix in $\mathbb{F}_2^{n \times n_0}$.

Write $p = \frac{r}{N}$. Recall that the sum of entries of $V \in W'$ is Np. For $0 \leq m \leq Np$, write W'_m for the set of matrices $V = B + E \in W'$ where the weight of B is exactly m. We proceed to bound $|W'_m|$ by separately counting the number of ways to choose E and B.

Choosing E: Given a choice of m, the matrix E has weight $Np - m$. Each entry of E is even, so each non-zero entry is at least 2. Thus, there are at most $\frac{Np-m}{2}$ non-zero entries. The number of ways to choose these entries is thus at most

$$\sum_{t=0}^{\frac{Np-m}{2}} \binom{N}{t} \leq 2^{N \cdot h_2(\frac{Np-m}{2N})}.$$

Subject to this choice, the number of possible matrices E is at most

$$\binom{\frac{Np-m}{2} + Np - m - 1}{Np - m} \leq 2^{\frac{3Np}{2}}.$$

Choosing B: Let $I = \{i \in [n] : B_i \neq 0\}$, where B_i stands for the i-th row of B. The number of ways to choose I is at most 2^n. Write $|I| = \gamma \cdot n$ for some $\gamma \in [0,1]$. Given m and the choice of I, we claim that there are at most

$$2^{N\gamma \cdot \left(h_2\left(\frac{m}{\gamma N}\right) - \varepsilon + \frac{1}{\sqrt{n_0}}\right)}$$

ways to choose the matrix B.

Indeed, recall that every row of B must lie in $\mathcal{C}_{\text{in}}^{\perp}$. Let $B' \in \mathbb{F}_2^{n \times n_0}$ be a random matrix sampled uniformly from all matrices with weight m, whose set of non-zero rows is I. The number of such matrices is at most $\binom{n_0 \cdot |I|}{m} = \binom{\gamma N}{m} \leq 2^{\gamma N \cdot h_2\left(\frac{m}{\gamma N}\right)}$. Conditioning on the weight of each row of B', the fact that $\mathcal{C}_{\text{in}}$ is $(1/\sqrt{n_0})$-nice implies that

$$\Pr\left[B'_i \in \mathcal{C}_{\text{in}}^{\perp} \mid |B'_i| = w\right] = \frac{|\{x \in \mathcal{C}_{\text{in}}^{\perp} : |x| = w\}|}{\binom{n_0}{w}} \leq 2^{-n_0\varepsilon + \sqrt{n_0}}$$

for all $i \in I$. Since the rows of B'_i are independent under this conditioning, the probability that every row of B' lies in $\mathcal{C}_{\text{in}}^{\perp}$ is at most $2^{N\gamma \cdot \left(-\varepsilon + \frac{1}{\sqrt{n_0}}\right)}$. The desired bound on the number of choices for B' follows.

Obtaining the conclusion: Overall, writing $m = \alpha N$, we conclude that

$$\begin{aligned}
\frac{\log_2(|W'_m|)}{N} &\leq \max_{\gamma \in [0,1]} \left\{ \frac{3p}{2} + \frac{n}{N} + h_2\left(\frac{p-\alpha}{2}\right) + \gamma \cdot \left(h_2\left(\frac{\alpha}{\gamma}\right) - \varepsilon + \frac{1}{\sqrt{n_0}} \right) \right\} \\
&\leq \max_{\gamma \in [0,1]} \left\{ \frac{3p}{2} + \frac{n}{N} + \frac{1}{\sqrt{n_0}} + \left(h_2\left(\frac{p-\alpha}{2}\right) + \gamma \cdot \left(h_2\left(\frac{\alpha}{\gamma}\right) - \varepsilon \right) \right) \right\} \\
&\leq \max_{\gamma \in [0,1]} \left\{ \frac{3p}{2} + \frac{1}{\sqrt{n_0}} + \frac{n}{N} + \frac{(p-\alpha)}{2} \log_2\left(\frac{2}{p-\alpha}\right) + (p-\alpha) + \alpha \log_2\left(\frac{\gamma}{\alpha}\right) + 2\alpha - \gamma\varepsilon \right\} \\
&\leq \frac{5p}{2} + \frac{1}{\sqrt{n_0}} + \frac{n}{N} + \frac{(p-\alpha)}{2} \log_2\left(\frac{2}{p-\alpha}\right) + \alpha\left(\log_2\left(\frac{1}{\varepsilon}\right) + 1 - \frac{\ln\ln 2 + 1}{\ln 2} \right) \\
&\leq 3p + \frac{1}{\sqrt{n_0}} + \frac{n}{N} + \frac{(p-\alpha)}{2} \log_2\left(\frac{1}{p-\alpha}\right) + \alpha \cdot \log_2\left(\frac{1}{\varepsilon}\right).
\end{aligned}$$

Here, the third inequality is since $h_2(x) \leq -x \log_2 x + 2x$ for all $x \in [0,1]$. The fourth inequality is by observing that $\gamma = \frac{\alpha}{\varepsilon \ln 2}$ maximizes the expression.

To bound this last expression, we observe[8] that it is maximized at

$$\alpha = \max\{p - \varepsilon^2/e, 0\}.$$

[8] Indeed, the derivative with respect to α of this expression is

$$\frac{1}{2\ln(2)}\left(1 + \ln\left(\frac{p-\alpha}{\varepsilon^2}\right)\right).$$

We consider two cases, one where $p < \varepsilon^2/e$ and one where $p \geq \varepsilon^2/e$. When $p < \varepsilon^2/e$,

$$\ln\left(\frac{p-\alpha}{\varepsilon^2}\right) < \ln(1/e) = -1,$$

and in particular the derivative is negative for all $\alpha \in [0, p)$. This means that the maximum is attained at $\alpha = 0$. On the other hand, if $p \geq \varepsilon^2/e$, the derivative is non-negative at zero, and vanishes when $\alpha = p - \varepsilon^2/e$, so the maximum is attained there.

Plugging this in, we claim that

$$\frac{\log_2(|W'_m|)}{N} \leq 3p + \frac{1}{\sqrt{n_0}} + \frac{n}{N} + \frac{p}{2}\log_2\left(\max\left\{\frac{1}{p}, \frac{e}{\varepsilon^2}\right\}\right).$$

In more detail, if $p < \varepsilon^2/e$, then plugging in $\alpha = 0$ yields the $1/p$ term inside the maximum. On the other hand, if $p \geq \varepsilon^2/e$, then plugging in $\alpha = p - \varepsilon^2/e$, we have

$$\begin{aligned}
\frac{(p-\alpha)}{2}\log_2\left(\frac{1}{p-\alpha}\right) + \alpha \cdot \log_2\left(\frac{1}{\varepsilon}\right) &= \frac{\varepsilon^2}{2e}\log_2\left(\frac{e}{\varepsilon^2}\right) + \left(p - \frac{\varepsilon^2}{e}\right)\log_2\left(\frac{1}{\varepsilon}\right) \\
&= \frac{\varepsilon^2}{2\ln(2)e} + \frac{p}{2}\log_2\left(\frac{1}{\varepsilon^2}\right) \\
&\leq \frac{p}{2\ln(2)} + \frac{p}{2}\log_2\left(\frac{1}{\varepsilon^2}\right) \qquad \text{as } p \geq \varepsilon^2/e \\
&= \frac{p}{2}\log_2\left(\frac{e}{\varepsilon^2}\right),
\end{aligned}$$

which gives us the e/ε^2 term inside the maximum. Therefore,

$$\begin{aligned}
|W| &\leq (Np)! \sum_{m=0}^{Np} |W'_m| \\
&\leq N^{Np+1} \cdot 2^{3Np+N/\sqrt{n_0}+n} \cdot p^{Np/2} \cdot \left(\max\{1, p/\varepsilon^2\}\right)^{Np/2} \\
&\leq (8N)^r \cdot \left(\frac{r}{N}\right)^{r/2} \cdot 2^{2N/\sqrt{n_0}+\log_2(N)} \cdot \left(\max\left\{1, \frac{r}{\varepsilon^2 N}\right\}\right)^{r/2},
\end{aligned}$$

where in the last line we have plugged in $p = r/N$ and also used $n = \frac{N}{n_0} \leq \frac{N}{\sqrt{n_0}}$. ∎

5 A Soft-Decoding Sufficient Condition for $\mathcal{C}_{\text{out}}$

In this section we give a sufficient condition on $\mathcal{C}_{\text{out}}$ for $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ to approach the GV bound. This condition is similar to a *soft-decoding* condition, except where the distributions for each coordinate are the same. That is, for a particular distribution $\mathcal{D}$ on $\mathbb{F}$ (defined in Theorem 5.1 below), we imagine choosing a random word $x \in \mathbb{F}_q^n$ so that $x_\alpha \sim \mathcal{D}$ for $\alpha \in [n]$ are all i.i.d. Then we ask about the probability that x lies in $\mathcal{C}_{\text{out}}^{\perp}$. If this probability is close to what it "should" be (for, say, a random linear code) then $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ will approach the GV bound with high probability over $\mathcal{C}_{\text{in}}$. In more detail, we prove the following theorem.

Theorem 5.1 (Sufficient soft-decoding condition for $\mathcal{C}_{\text{out}}$)**.** *There are constants $\tilde{c}, c > 0$ so that the following holds. Let $\varepsilon > 0$. Suppose that $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ is a binary linear code of dimension $k_0 = \varepsilon n_0$ that is τ-nice for $\tau = 1/\sqrt{n_0}$. Further assume that $n_0 \geq 64/\varepsilon^4$. Let $\Omega \subseteq \mathbb{F}_q$ be the set defined from $\mathcal{C}_{\text{in}}$ as per Definition 3.1.*

Let $Y \in \mathbb{F}$ be the random variable given by

$$Y = \sum_{b \in \Omega} \zeta_b \cdot b,$$

where $\zeta_b \sim \mathsf{Ber}\left(\frac{1-e^{-2\tilde{c}\varepsilon^2}}{2}\right)$ are i.i.d. Bernoulli random variables, and let $\mathcal{D}$ be the distribution of Y. Suppose that $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ is a linear code of dimension $k = \varepsilon n$ that satisfies

$$\Pr_{x \sim \mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}] \leq \frac{1}{q^k}(1+\Delta), \tag{11}$$

where

$$\Delta \leq \left(\frac{c\varepsilon}{2}\right)^{\tilde{c}\varepsilon^2 N + 100\sqrt{\tilde{c}\varepsilon^2 N}}.$$

Then, $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is a binary linear code of rate ε^2 with relative distance at least $\frac{1-c\varepsilon}{2}$.

Remark 6 (τ-niceness of inner code)**.** *Note that, by Lemma 2.3, a random linear code $\mathcal{C}_{\text{in}}$ is τ-nice for $\tau = 1/\sqrt{n_0}$ with probability at least $1 - 2^{-\Omega(\sqrt{n_0})}$. Thus, Theorem 5.1 implies that if $\mathcal{C}_{\text{out}}$ satisfies Equation (11), then $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ approaches the GV bound with high probability over a random $\mathcal{C}_{\text{in}}$.*

Remark 7 (The distribution $\mathcal{D}$ is somewhat concentrated.)**.** *We note that $(1 - e^{-2\tilde{c}\varepsilon^2})/2 = \Theta(\varepsilon^2)$ for small $\varepsilon > 0$. In particular, the random variables ζ_b in the distribution $\mathcal{D}$ that appears in Theorem 5.1 is $\mathsf{Ber}(p)$ for $p = \Theta(\varepsilon^2)$. This means that a "typical" draw from $\mathcal{D}$ will be a sum of $\Theta(\varepsilon^2 n_0)$ elements of Ω, so $\mathcal{D}$ has most of its mass on about $q^{O(\varepsilon \log(1/\varepsilon))}$ elements of $\mathbb{F}_q$, out of q. In this sense, the condition in Theorem 5.1 is reminiscent of a list-recovery-type condition on $\mathcal{C}_{\text{out}}^{\perp}$, where the input lists are all the same and have size about $\ell = q^{O(\varepsilon \log(1/\varepsilon))}$.*

Remark 8 (Codes satisfying Equation (11) exist)**.** *While the eventual goal is to explicitly construct a code $\mathcal{C}_{\text{out}}$ that satisfies Equation (11), as a proof of concept we remark that such codes do exist. Indeed, imagine taking a random linear code $\mathcal{C}_{\text{out}}$ of dimension $k = \varepsilon n$. It is not hard to see that*

$$\mathop{\mathbb{E}}_{\mathcal{C}_{\text{out}}}\left[\Pr_{x \sim \mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}]\right] = q^{-k} - q^{-n}.$$

In particular, there exists a linear code $\mathcal{C}_{\text{out}}$ of dimension k so that

$$\Pr_{x \sim \mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}] \leq \frac{1}{q^k},$$

satisfying the condition of Theorem 5.1 with $\Delta = 0$. (Notice that in the theorem, it is okay if $\Pr_{x \sim \mathcal{D}^n}[x \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}]$ is smaller than $1/q^k$, it just should not be much larger).

We prove the theorem at the end of the section, after putting a few preliminaries in place. For the rest of the section, let $\tilde{c}$ and c be the constants in Theorem 5.1.

For a message $m \in \mathbb{F}^k \setminus \{0\}$, say that m is *bad* if $|X_m| \geq c\varepsilon N$, where X_m is as defined in Section 3. Thus, if there are no bad messages, then the conclusion of Theorem 5.1 holds.

For any $r \geq 0$, define

$$\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) \triangleq \frac{q^k}{q^k - 1} \cdot \frac{1}{(c\varepsilon N)^r} \cdot \sum_{\mathcal{V} \in (\Lambda \times \Omega)^r} \left(q^k \cdot \mathbf{1}[g_{\mathcal{V}} \in \mathcal{C}_{\text{out}}^{\perp}] - 1\right).$$

Observation 5.2. *For any r, the number of bad messages $m \in \mathbb{F}^k \setminus \{0\}$ is at most $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}})$.*

Proof: By Lemma 3.3, we have

$$\mathop{\mathbb{E}}_{m \sim \mathbb{F}^k \setminus \{0\}}[X_m^r] = \frac{1}{q^k - 1} \sum_{\mathcal{V} \in ([n] \times \Omega)^r} (q^k \cdot \mathbf{1}[g_{\mathcal{V}} \in \mathcal{C}_{\text{out}}^{\perp}] - 1).$$

Markov's inequality along with the definition of $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}})$ implies that

$$\Pr_{m \sim \mathbb{F}^k \setminus \{0\}}[X_m \geq C\varepsilon N] \leq \frac{1}{q^k} \mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}).$$

Thus, the total number of of bad m-s is bounded by $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}})$, as desired. ∎

We may re-write the sum over $\mathcal{V} \in ([n] \times \Omega)^r$ as an expectation over a corresponding distribution $\mathcal{D}_r$ on $g \in \mathbb{F}_q^n$. That is, to sample from $\mathcal{D}_r$, we choose a sequence $\mathcal{V}$ of pairs $(\alpha_i, b_i) \in [n] \times \Omega$ for $i \in [r]$ independently and uniformly at random; let $g \in \mathbb{F}_q^n$ be given by

$$g_\alpha = \sum_{b \in \mathcal{V}_\alpha} b$$

for $\alpha \in [n]$, as we did in Section 3. Thus, the above becomes

$$\begin{aligned} \mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) &= \frac{q^k}{q^k - 1} \cdot \left(\frac{N}{c\varepsilon N}\right)^r \cdot \mathbb{E}_{g \sim \mathcal{D}_r}\left(q^k \cdot \mathbf{1}[g \in \mathcal{C}_{\text{out}}^{\perp}] - 1\right) \\ &= \frac{q^k}{q^k - 1} \cdot \left(\frac{1}{c\varepsilon}\right)^r \cdot \left(q^k \Pr_{g \sim \mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^{\perp}] - 1\right). \end{aligned} \tag{12}$$

This inspires a nice condition for $\mathcal{C}_{\text{out}}$; we want $\Pr_{g \sim \mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^{\perp}]$ to be about $1/q^k$. This is reminiscent of a soft-decoding problem, but one difference is that coordinates g_α for $\alpha \in [n]$ are not independent. In order to *make* them independent, as they are in the statement of Theorem 5.1, we will first choose r randomly from an Poisson distribution. That is, we will choose

$$r \sim \mathsf{Poi}(\tilde{c}\varepsilon^2 N).$$

We first observe that choosing r at random like this and then choosing $g \sim \mathcal{D}_r$ results in the product distribution $\mathcal{D}^n$ in Theorem 5.1.

Claim 5.3. *Suppose that $r \sim \mathsf{Poi}(\tilde{c}\varepsilon^2 N)$ and $g \sim \mathcal{D}_r$. The the joint distribution of the g_α is given by*

$$g_\alpha = \sum_{b \in \Omega} \zeta_{\alpha,b} \cdot b,$$

where

$$\zeta_{\alpha,b} \sim \mathsf{Ber}\left(\frac{1 - e^{-2\tilde{c}\varepsilon^2}}{2}\right)$$

are i.i.d. Bernoulli random variables. In particular, the g_α are all independent and identically distributed for each $\alpha \in [n]$.

Proof: We may view the process of choosing the (α_i, b_i) as dropping r balls into N bins, with each bin corresponding to a tuple $(\alpha, b) \in [n] \times \Omega$. Since $r \sim \mathsf{Poi}(\tilde{c}\varepsilon^2 N)$, the occupancy $Y_{\alpha,b}$ of each bin (α, b) is independent, and distributed as

$$Y_{\alpha,b} \sim \mathsf{Poi}(\tilde{c}\varepsilon^2).$$

Notice that $Y_{\alpha,b}$ is the number of times that b appears in the multiset $\mathcal{V}_\alpha$. Thus, by definition, each g_α is of the form

$$g_\alpha = \sum_{b \in \Omega} Y_{\alpha,b} b = \sum_{b \in \Omega} \zeta_{\alpha,b} b,$$

where

$$\zeta_{\alpha,b} = Y_{\alpha,b} \bmod 2.$$

Above, we can replace $Y_{\alpha,b}$ with $\zeta_{\alpha,b}$ since we are working over a field of characteristic two.

Since the $Y_{\alpha,b}$ are all independent, so are the $\zeta_{\alpha,b}$. Further, $\zeta_{\alpha,b}$ is a Bernoulli random variable, and the probability that it is equal to one is the probability that $Y_{\alpha,b}$ is odd. Since $Y_{\alpha,b} \sim \mathsf{Poi}(\tilde{c}\varepsilon^2)$, this is

$$\Pr[\zeta_{\alpha,b} = 1] = \Pr[Y_{\alpha,b} \text{ odd}] = \sum_{j=0}^{\infty} \frac{(\tilde{c}\varepsilon^2)^{2j} e^{-\tilde{c}\varepsilon^2}}{(2j)!} = e^{-\tilde{c}\varepsilon^2}\left(\frac{e^{\tilde{c}\varepsilon^2} - e^{-\tilde{c}\varepsilon^2}}{2}\right) = \frac{1}{2}\left(1 - e^{-2\tilde{c}\varepsilon^2}\right).$$

This proves the claim. ∎

Given Claim 5.3, we return to (12), and break it up into two terms, one corresponding

to the event that $g = 0$, and one corresponding to the event that $g \neq 0$. We have

$$\begin{aligned}
\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) &= \frac{q^k}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp] - 1\right) \\
&= \frac{q^k}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \Pr_{g\sim\mathcal{D}_r}[g=0] \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp \,|\, g=0] - 1\right) \\
&\qquad + \frac{q^k}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \Pr_{g\sim\mathcal{D}_r}[g\neq 0] \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp \,|\, g\neq 0] - 1\right) \\
&= q^k\left(\frac{1}{c\varepsilon}\right)^r \cdot \Pr_{g\sim\mathcal{D}_r}[g=0] \\
&\qquad + \frac{q^k}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \Pr_{g\sim\mathcal{D}_r}[g\neq 0] \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp \,|\, g\neq 0] - 1\right) \\
&= q^k\left(1+\frac{1}{q^k-1}\right)\left(\frac{1}{c\varepsilon}\right)^r \cdot \Pr_{g\sim\mathcal{D}_r}[g=0] \qquad (13) \\
&\qquad + \frac{q^k}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp \setminus \{0\}] - 1\right) \qquad (14)
\end{aligned}$$

Above, in the second-to-last equality we simplified the first summand by noting that $\Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^\perp \,|\, g = 0] = 1$ and canceling the $q^k - 1$ terms. In the final equality, we distributed $\Pr_{g\sim\mathcal{D}_r}[g \neq 0] = 1 - \Pr_{g\sim\mathcal{D}_r}[g = 0]$ inside the sum in the second summand, and moved a resulting $\frac{1}{q^k-1}\left(\frac{1}{c\varepsilon}\right)^r \Pr_{g\sim\mathcal{D}_r}[g = 0]$ term to the first summand. We handle the two terms (13) and (14) separately. The first one we will show is small; and the second we will show is small *if* $\mathcal{C}_{\text{out}}$ has a particular soft-decoding-like guarantee.

Claim 5.4. *Suppose that the constants $c, \tilde{c}$ satisfy*

$$\tilde{c} \geq 4\ln(2) \qquad \textit{and} \qquad c \geq 72\tilde{c},$$

and that n is sufficiently large. Suppose that $r \sim \mathsf{Poi}(\tilde{c}\varepsilon^2 N)$ as above, and further suppose that $n_0 \geq 64/\varepsilon^4$. Then the expectation over r of the term (13) satisfies

$$\mathbb{E}_r[(13)] \leq 2^{-\varepsilon^2 N/2}.$$

Proof: First, observe that for fixed r, the probability that $g \sim \mathcal{D}_r$ is zero is precisely what is bounded in Claim 4.3, and we have

$$\Pr_{g\sim\mathcal{D}_r}[g=0] \leq 8^r \cdot \left(\frac{r}{N}\right)^{r/2} \cdot 2^{\frac{2N}{\sqrt{n_0}}+\log_2(N)} \cdot \left(\max\left\{1, \frac{er}{\varepsilon^2 N}\right\}\right)^{r/2}.$$

Let $\lambda = \tilde{c}\varepsilon^2 N$ be the mean of the Poisson distribution that r is drawn from. Taking the

expectation over r and plugging the result of Claim 4.3, the expected value of (13) is:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{r}\left[q^k\left(\frac{q^k}{q^k-1}\right)\left(\frac{1}{c\varepsilon}\right)^r \Pr_{g\sim\mathcal{D}_r}[g=0]\right] \\
&\quad\le q^k 2^{3N/\sqrt{n_0}}\sum_{r\ge 0}\left(\frac{1}{c\varepsilon}\right)^r 8^r\left(\frac{r}{N}\right)^{r/2}\left(\max\left\{1,\frac{er}{\varepsilon^2 N}\right\}\right)^{r/2}\Pr[\mathsf{Poi}(\lambda)=r] \\
&\quad= q^k 2^{3N/\sqrt{n_0}}\sum_{r\ge 0}\left(\frac{1}{c\varepsilon}\right)^r 8^r\left(\frac{r}{N}\right)^{r/2}\left(\max\left\{1,\frac{er}{\varepsilon^2 N}\right\}\right)^{r/2}\left(\frac{\lambda^r e^{-\lambda}}{r!}\right) \\
&\quad= e^{-\lambda}q^k 2^{3N/\sqrt{n_0}}\sum_{r\ge 0}\frac{1}{r!}\left(\frac{8\lambda\sqrt{r/N}}{c\varepsilon}\right)^r\left(\max\left\{1,\frac{er}{\varepsilon^2 N}\right\}\right)^{r/2} \\
&\quad= e^{-\lambda}q^k 2^{3N/\sqrt{n_0}}\left(\sum_{r=0}^{\varepsilon^2 N/e}\frac{1}{r!}\left(\frac{8\lambda\sqrt{r/N}}{c\varepsilon}\right)^r + \sum_{r>\varepsilon^2 N/e}\frac{1}{r!}\left(\frac{8\sqrt{e}\lambda r}{c\varepsilon^2 N}\right)^r.\right)
\end{aligned}$$

Consider each of the two summations above. The first one is bounded by

$$\begin{aligned}
\sum_{r=0}^{\varepsilon^2 N/e}\frac{1}{r!}\left(\frac{8\lambda\sqrt{r/N}}{c\varepsilon}\right)^r &\le \sum_{r=0}^{\infty}\frac{1}{r!}\left(\frac{8\lambda\sqrt{\varepsilon^2/e}}{c\varepsilon}\right)^r \\
&= \exp\left(\frac{8}{c\sqrt{e}}\cdot\lambda\right).
\end{aligned}$$

Meanwhile, the second term is bounded by

$$\begin{aligned}
\sum_{r>\varepsilon^2 N/e}\frac{1}{r!}\left(\frac{8\sqrt{e}\lambda r}{c\varepsilon^2 N}\right)^r &\le \sum_{r>\varepsilon^2 N/e}\left(\frac{8e\sqrt{e}\lambda r}{c\varepsilon^2 r N}\right)^r \qquad \text{using } r!\ge (r/e)^r \\
&= \sum_{r>\varepsilon^2 N/e}\left(\frac{8e\sqrt{e}\tilde{c}}{c}\right)^r \\
&\le 2^{-\varepsilon^2 N/e} < 1
\end{aligned}$$

provided that $c \ge 16e\sqrt{e}\tilde{c}$. Then, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{r}\left[q^k\left(\frac{1}{c\varepsilon}\right)^r \Pr_{g\sim\mathcal{D}_r}[g=0]\right] &\le e^{-\lambda}q^k 2^{3N/\sqrt{n_0}}(\exp(8\lambda/c)+1) \\
&\le 2e^{-\lambda}q^k 2^{3N/\sqrt{n_0}}\exp(8\lambda/c) \\
&\le \exp\big(\ln(2)\big(4N/\sqrt{n_0}+\varepsilon^2 N\big) - \lambda(1-8/c)\big).
\end{aligned}$$

Above, in the final line we used the fact that $q^k = 2^{\varepsilon^2 N}$. Plugging in $\lambda = \tilde{c}\varepsilon^2 N$, we get

$$\mathop{\mathbb{E}}_{r}\left[q^k\left(\frac{1}{c\varepsilon}\right)^r \Pr_{g\sim\mathcal{D}_r}[g=0]\right] \le \exp\left(\varepsilon^2 N(\ln(2)-\tilde{c}(1-8/c)) + \frac{4\ln(2)N}{\sqrt{n_0}}\right).$$

Provided that

$$\tilde{c} \geq 4\ln(2) \qquad \text{and} \qquad c \geq 16,$$

(which both follow from the assumptions in the theorem statement) we get

$$\begin{aligned}\mathbb{E}_r\left[q^k\left(\frac{1}{c\varepsilon}\right)^r \Pr_{g\sim\mathcal{D}_r}[g=0]\right] &\leq \exp_2\left(-\varepsilon^2 N\left(1-\frac{4}{\varepsilon^2\sqrt{n_0}}\right)\right)\\ &\leq \exp_2\left(-\varepsilon^2 N/2\right),\end{aligned}$$

finally using our assumption that $n_0 \geq 64/\varepsilon^4$. ∎

Finally we are ready to prove Theorem 5.1.

Proof of Theorem 5.1: The rate of the code $\mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ follows by definition, so we only need to establish the bound on the relative distance. By Observation 5.2, the number of bad messages is at most $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}})$ for any r, so it suffices to show that there exists an r so that $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) < 1$. We will choose $r \sim \mathsf{Poi}(\lambda)$, where $\lambda = \tilde{c}\varepsilon^2 N$, as above. As above, for any r we can write

$$\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) = (13) + (14),$$

and Claim 5.4 implies that

$$\mathbb{E}_r[(13)] \leq 2^{-\varepsilon^2 N/2}.$$

In particular, by Markov's inequality, with probability at least $1 - 2^{-\varepsilon^2 N/4}$ over the choice of r, we have

$$(13) \leq 2^{-\varepsilon^2 N/4}. \tag{15}$$

Now we turn our attention to the term (14):

$$(14) = \frac{q^k}{q^k - 1}\left(\frac{1}{c\varepsilon}\right)^r \cdot \left(q^k \Pr_{g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}] - 1\right).$$

Note that, by Claim 5.3, when $r \sim \mathsf{Poi}(\lambda)$, the distribution $\mathcal{D}_r$ is the same as the distribution $\mathcal{D}^n$ from the statement of the theorem. Thus, by the assumption of the theorem,

$$q^k \Pr_{r\sim\mathsf{Poi}(\lambda), g\sim\mathcal{D}_r}[g \in \mathcal{C}_{\text{out}}^{\perp} \setminus \{0\}] - 1 \leq \Delta \leq \left(\frac{c\varepsilon}{2}\right)^{\lambda + 100\sqrt{\lambda}}.$$

Further, by a Chernoff bound for Poisson random variables (e.g., [MU17]),

$$\Pr[r \geq \lambda + 100\sqrt{\lambda}] \leq \exp\left(\frac{-100^2\lambda}{\lambda + 100\sqrt{\lambda}}\right) \leq \exp(-100^2/2)$$

for sufficiently large N. Thus, union bounding over the event that Equation (15) occurs and that $r \leq \lambda + 100\sqrt{\lambda}$, we see that with probability at least

$$1 - 2^{-\varepsilon^2 N/2} - \exp(-100^2/2) > 0,$$

over the choice of r, we have

$$\begin{aligned}\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) &\leq 2^{-\varepsilon^2 N/4} + \frac{q^k}{q^k - 1}\left(\frac{1}{c\varepsilon}\right)^{\lambda+100\sqrt{\lambda}} \cdot \left(\frac{c\varepsilon}{2}\right)^{\lambda+100\sqrt{\lambda}} \\ &\leq 2^{-\varepsilon^2 N/4} + 2^{-\tilde{c}\varepsilon^2 N} \\ &< 1.\end{aligned}$$

As this is (much) less than 1 for sufficiently large N, we conclude that in particular there exists an r so that $\mathcal{B}_r(\mathcal{C}_{\text{out}}, \mathcal{C}_{\text{in}}) < 1$, which proves the theorem. ∎

6 A High Min-Entropy Sufficient Condition for $\mathcal{C}_{\text{out}}$

In this section, we give a second sufficient condition under which $\mathcal{C}_{\text{out}}$ will be "good" for concatenation with a random linear inner code. This second sufficient condition, informally, says that the codewords of $\mathcal{C}_{\text{out}}$ should have "mildly flat" symbol distributions.

In more detail, given a word $c \in \mathbb{F}_q^n$, let $\mathcal{D}_c$ denote the empirical distribution of symbols in c. That is, for $\sigma \in \mathbb{F}_q$,

$$\Pr_{\mathcal{D}_c}[\sigma] = \Pr_{\alpha \sim [n]}[c_\alpha = \sigma].$$

(For example, if $c = (\sigma, \sigma, \ldots, \sigma)$, then $\boldsymbol{c}$ is the distribution on $\mathbb{F}_q$ with 100% of the mass on σ; and if $n = q$ and c has one of each different symbol in $\mathbb{F}_q$, then $\boldsymbol{c}$ is uniform on $\mathbb{F}_q$). Given a word $c \in \mathbb{F}_q^n$, the *min-entropy* of the distribution $\mathcal{D}_c$ is given by

$$H_\infty(c) \triangleq -\log_2 \max_\sigma \Pr_{\mathcal{D}_c}[\sigma].$$

Observe that $H_\infty(c) \leq \log_2(q)$.

Our condition will be about a smoothed notation of min-entropy, which informally allows a small η-fraction of the mass of $\mathcal{D}_c$ to have high min-entropy. Formally, for some smoothness parameter $\eta > 0$, we define the *smoothed min-entropy* H_∞^η by

$$H_\infty^\eta(\mathcal{D}_c) \triangleq \max_{\mathcal{D}_{c'}: \Delta_{\text{TV}}(\mathcal{D}_c, \mathcal{D}_{c'}) \leq \eta} H_\infty(\mathcal{D}_{c'}),$$

where for two distributions $\mathcal{D}_c$, $\mathcal{D}_{c'}$, Δ_{TV} is the total variation distance.

Before we state our main theorem in this section (Theorem 6.2 below), we state a Lemma that we will eventually apply to the inner code $\mathcal{C}_{\text{in}}$.

Lemma 6.1. *For any $n \in \mathbb{N}$, $\frac{8}{\sqrt{n}} \leq \gamma \leq \frac{1}{3}$, and integers $\frac{24 \log n}{\gamma} \leq k \leq \frac{n}{5}$ and $2^{2k/3} \leq T \leq 2^{(1-\gamma)k}$, a random linear code $\mathcal{C} \subseteq \mathbb{F}_2^n$ of dimension k satisfies the following with probability $1 - 2^{-\Omega(\gamma k + \gamma^2 n + \sqrt{n})}$.*

For $0 \leq j \leq n$, let Δ_j be the number of codewords of $\mathcal{C}$ of Hamming weight j, and denote by $j^\star$ the minimal j for which $\sum_{i=0}^{j} \Delta_i \geq T$. Then,

1. *It holds that* $j^\star \geq \alpha n$, *for some* $\alpha \geq h_2^{-1}\left(1 - 2 \cdot \frac{k - \log T}{n}\right)$.

2. *It holds that* $\frac{\sum_{i=0}^{j^\star} i \cdot \Delta_i}{\sum_{i=0}^{j^\star} \Delta_i} \geq (1 - 2\gamma) j^\star$.

3. *It holds that* $\frac{\Delta_{j^\star+1}}{\sum_{i=0}^{j^\star} \Delta_i} \leq 2^{\sqrt{n}}$.

We defer the proof to the end of the section. Our main result for this section is the following.

Theorem 6.2. *Fix any sufficiently small* $\varepsilon > 0$. *For any integers* $k, q \in \mathbb{N}$, *so that* $q = 2^{k_0}$ *is a power of* 2. *Let* $n_0 = k_0/\varepsilon$ *and* $n = k/\varepsilon$. *Let* $\mathcal{C}_{\text{out}} \subseteq \mathbb{F}_q^n$ *be an* $\mathbb{F}_2$*-linear code of rate* ε, *and let* $\mathcal{C}_{\text{in}} \subseteq \mathbb{F}_2^{n_0}$ *be a random linear code of rate* ε.

Assume further that there exist constants $\bar{c}_\gamma, \bar{c}_\eta$ *such that for every nonzero* $c \in \mathcal{C}_{\text{out}}$,

$$H_\infty^{\bar{c}_\eta \varepsilon}(c) \geq (1 - \bar{c}_\gamma \varepsilon) \log q,$$

and that $n_0 = \Omega_{\bar{c}_\gamma}\left(\frac{\log(1/\varepsilon)}{\varepsilon^2}\right)$. *Then, with probability at least*

$$1 - 2^{-\Omega_{\bar{c}_\gamma}(\varepsilon^2 n_0 + \sqrt{n_0})}$$

over the choice of $\mathcal{C}_{\text{in}}$, *the concatenated code* $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ *has relative distance* $\frac{1}{2} - O_{\bar{c}_\gamma, \bar{c}_\eta}(\varepsilon)$.

Remark 9 (Do there exist good $\mathcal{C}_{\text{out}}$?)**.** *It is not clear (to us) whether a random linear code satisfies the min-entropy condition of Theorem 6.2 with high probability (if it did, it would give an alternate proof of Theorem 4.2). However, as a proof of concept we note that a random linear code does satisfy the property that the symbols of every nonzero codeword are* not *contained in a set of size smaller than* $q^{1-\gamma}$. *Note that this is necessary from any* c *with* $H_\infty(c) \geq (1 - \gamma) \log_2(q)$.

Indeed, taking a random linear code $\mathcal{C}_{\text{out}}$ *of dimension* $k = \varepsilon n$, *the probability that a given nonzero codeword violates that constraint is* $\sum_{i=1}^{q^{1-\gamma}} \binom{q}{i} (i/q)^n \approx q^{-\gamma(n - q^{1-\gamma})}$. *Taking the union bound over all* q^k *codewords in* $\mathcal{C}_{\text{out}}$, *we get that no nonzero codewords violates the constraint with high probability, say for* $\gamma = \frac{1}{2}\varepsilon$, *assuming* $q = O(n)$.

Proof of Theorem 6.2: Throughout the proof, let

$$\gamma = \bar{c}_\gamma \varepsilon \text{ and } \eta = \bar{c}_\eta \varepsilon,$$

where $\bar{c}_\gamma$ and $\bar{c}_\eta$ are the constants from the theorem statement.

As $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is a linear code, to lower bound the distance it suffices to lower bound the minimum-weight codeword. To that end, fix any nonzero codeword $c \in \mathcal{C}_{\text{out}}$, and consider the word

$$w = (\mathcal{C}_{\text{in}}(c_1), \mathcal{C}_{\text{in}}(c_2), \ldots, \mathcal{C}_{\text{in}}(c_n)).$$

Order the elements of $\mathbb{F}_q$ as $\sigma_0, \sigma_1, \ldots, \sigma_{q-1}$ so that

$$\Pr_{\mathcal{D}_c}[\sigma_0] \geq \Pr_{\mathcal{D}_c}[\sigma_1] \geq \cdots \geq \Pr_{\mathcal{D}_c}[\sigma_{q-1}],$$

and define

$$p_t(c) = \Pr_{\mathcal{D}_c}[\sigma_t]$$

for $t \in \{0, 1, \ldots, q-1\}$.

To bound the weight of w, it suffices to do so under a worst-case (not necessarily linear) mapping from $\mathbb{F}_q$ to $\mathcal{C}_{\text{in}}$. That is, this worst-case mapping will map the most frequent symbols in w to the lowest-weight codewords in $\mathcal{C}_{\text{in}}$. Concretely for $j \in \{0, 1, \ldots, n_0\}$, let

$$\Delta_j = \{x \in \mathcal{C}_{\text{in}} \,:\, \mathsf{weight}(x) = j\}$$

be the number of codewords of $\mathcal{C}_{\text{in}}$ of weight j, and let

$$\ell_j = \Delta_0 + \Delta_1 + \cdots + \Delta_j$$

be the number of codewords of $\mathcal{C}_{\text{in}}$ of weight at most j, with the convention that $\ell_{-1} = 0$. Then we consider a worst-case encoding of $\mathbb{F}_q$ to $\mathcal{C}_{\text{in}}$ so that the set of symbols

$$\left\{\sigma_{\ell_{(j-1)}}, \sigma_{\ell_{(j-1)}+1}, \ldots, \sigma_{\ell_j - 1}\right\}$$

maps to the set of Δ_j codewords of $\mathcal{C}_{\text{in}}$ of weight j. Note that some of these sets may be empty. For example, $\mathcal{C}_{\text{in}}$ is not expected have any small-weight nonzero codewords, so, e.g., Δ_1 is very likely zero.

With this notation, we can bound the weight of $w = c \circ \mathcal{C}_{\text{in}}$ by $\mathsf{weight}(w) \geq d(c)$, where

$$d(c) \triangleq n \cdot \sum_{j=0}^{n_0} \left(\left(\sum_{t=\ell_{j-1}}^{\ell_j - 1} p_t(c) \right) \cdot j \right) = n \cdot \sum_{j=d_{\text{in}}}^{n_0} \left(\left(\sum_{t=\ell_{j-1}}^{\ell_j - 1} p_t(c) \right) \cdot j \right), \tag{16}$$

where d_{in} denotes the minimum distance of $\mathcal{C}_{\text{in}}$.

Next, we lower bound $d(w)$ by considering a worst-case empirical distribution $\mathcal{D}_c$ for the symbols in c. That is, we will find the values for p_t that minimize Equation (16) while still corresponding to a distribution $\boldsymbol{p}$ that obeys our assumption that $H_\infty^\eta(\boldsymbol{p}) \geq (1-\gamma)\log_2(q)$. For convenience, let

$$b = (1-\gamma)\log q$$

be our bound on the smoothed min-entropy.

Formally, we have that for all non-zero $c \in \mathcal{C}_{\text{out}}$,

$$d(c) \geq d \triangleq \min_{\boldsymbol{p}: H_\infty^\eta(\boldsymbol{p}) \leq b} n \cdot \sum_{j=d_{\text{in}}}^{n_0} \left(\left(\sum_{t=\ell_{(j-1)}}^{\ell_j - 1} p_t \right) \cdot j \right), \tag{17}$$

where the minimum is over all distributions $\boldsymbol{p} = (p_0, p_1, \ldots, p_{q-1})$ so that $p_0 \geq p_1 \geq \cdots \geq p_{q-1} \geq 0$ and $H_\infty^\eta(\boldsymbol{p}) \leq b$.

For intuition, consider what this worst-case distribution $\boldsymbol{p}$ would be if our requirement were only that the *non-smoothed* entropy $H_\infty(\boldsymbol{p}) \geq b$. Then the worst-case distribution

$\boldsymbol{p}$ would be the distribution that is uniform on the set of 2^b symbols σ that map to the lowest-weight codewords; in particular, we'd have $p_t = 2^{-b}$ for $t = 0, \ldots, 2^b - 1$. Given the η-smoothing allowance, then, the worst-case distribution would simply shift η of this mass to the symbol that is mapped to the zero codeword. This means that the worst-case distribution $\boldsymbol{p}$ is the one given by values $p_t(b)$ so that:

$$p_t(b) = \begin{cases} \eta + 2^{-b} & t = 0 \\ 2^{-b} & t = 1, \ldots, (1-\eta)2^b - 1 \\ 0 & \text{else} \end{cases}$$

For a reason that will be apparent soon, we will in fact work with a slightly smaller entropy bound $b' < b$. Towards defining b', set

$$T \triangleq (1-\eta)2^b,$$

and let

$$j^\star = \min\{j \leq n_0 : \ell_{j^\star} \geq T\}.$$

We then set b' so that

$$\ell_{j^\star - 1} = (1-\eta)2^{b'} \triangleq T'.$$

Clearly, the worst-case $p_t(b')$ is even *worse* than $p_t(b)$. Also note that $\frac{T}{T'} \leq 1 + \frac{\Delta_{j^\star}}{\ell_{j^\star - 1}}$.

Now that we know what the worst-case $\boldsymbol{p}$ looks like, and we still need to bound the value d in Equation (17). Given our expression for $\boldsymbol{p}$, we see that

$$d \geq n \sum_{j=d_{\text{in}}}^{j^\star - 1} \left(\left(\sum_{t=\ell_{(j-1)}}^{\ell_j - 1} 2^{-b} \right) \cdot j \right) = \frac{n}{2^{b'}} \sum_{j=d_{\text{in}}}^{j^\star - 1} (\Delta_j \cdot j), \tag{18}$$

Next, we will apply Lemma 6.1 with our choice of γ, and T'. We conclude that with probability at least

$$1 - 2^{-\Omega(\gamma k_0 + \gamma^2 n_0 + \sqrt{n_0})} = 1 - 2^{-\Omega_{\bar{c}_\gamma}(\varepsilon^2 n_0 + \sqrt{n_0})}$$

over the choice of $\mathcal{C}_{\text{in}}$, the favorable case holds. (Note that we could indeed apply Lemma 6.1, since $\gamma \geq \frac{8}{\sqrt{n_0}}$ and $k_0 \geq \frac{24 \log n_0}{\gamma}$ by our lower bound on n_0.)

The first conclusion of Lemma 6.1 implies that $j^\star - 1 = \alpha n_0$ for some $\alpha \in [0, 1]$ that satisfies

$$\begin{aligned} \alpha &\geq h_2^{-1}\left(1 - 2 \cdot \frac{\log q - \log T'}{n_0}\right) = h_2^{-1}\left(1 - 2 \cdot \frac{\log q - \log T + (\log T - \log T')}{n_0}\right) \\ &\geq h_2^{-1}\left(1 - 2 \cdot \frac{\log q - \log T + \sqrt{n_0} + 1}{n_0}\right) \geq h_2^{-1}\left(1 - 2 \cdot \frac{\log q - \log T}{n_0} - \frac{2}{\sqrt{n_0}}\right), \end{aligned}$$

where the bound on $\log T - \log T'$ follows from the third conclusion of Lemma 6.1. Further,

$$\begin{aligned}\alpha &\geq h_2^{-1}\left(1 - 2\varepsilon \cdot \left(\gamma + \frac{\log\left(\frac{1}{1-\eta}\right)}{\log q}\right) - \frac{2}{\sqrt{n_0}}\right) \\ &\geq h_2^{-1}\left(1 - 2\varepsilon\left(\gamma + \frac{2\eta}{\log q}\right) - \varepsilon\right) \geq h_2^{-1}(1 - 4\varepsilon\gamma) \geq \frac{1}{2} - \sqrt{\gamma\varepsilon \ln 2}\end{aligned}$$

where we have used the fact that $h_2^{-1}(1-x^2) \leq \frac{1}{2} - \frac{\sqrt{\ln 2}}{2}x$ (see, e.g., [GRS, Lemma B.2.4]), and that $n_0 \geq 4\varepsilon^{-2}$. We also used the fact that we may assume that $\varepsilon \leq \frac{1}{2}$ and q is larger than some constant depending on $\bar{c}_\gamma, \bar{c}_\eta$.

Next, the second conclusion of Lemma 6.1 implies that

$$\sum_{j=0}^{j^\star-1}(\Delta_j)\cdot j \geq (1-2\gamma)\cdot(j^\star-1)\cdot\left(\sum_{j=0}^{j^\star-1}\Delta_j\right) \geq (1-2\gamma)\cdot\alpha n_0 \cdot T',$$

Thus, returning to Equation (18), we have

$$\begin{aligned}\frac{d}{n_0\cdot n} &\geq \frac{1}{n_0\cdot 2^{b'}}\sum_{j=0}^{j^\star-1}(\Delta_j\cdot j) \\ &\geq (1-2\gamma)\cdot\alpha\cdot\frac{T'}{2^{b'}} = (1-2\gamma)\cdot\alpha\cdot(1-\eta),\end{aligned}$$

using the definition of T' in the final line. Since $N = n_0 n$ is the length of the code $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$, altogether we have that the relative distance of $\mathcal{C} = \mathcal{C}_{\text{out}} \circ \mathcal{C}_{\text{in}}$ is at least

$$\begin{aligned}\frac{d}{N} &\geq (1-2\gamma)(1-\eta)\alpha \\ &\geq (1-2\gamma)(1-\eta)\left(\frac{1}{2} - \sqrt{\gamma\varepsilon\ln 2}\right) \\ &\geq \frac{1}{2} - \left(\sqrt{\bar{c}_\gamma \ln 2} + \bar{c}_\gamma + \frac{1}{2}\bar{c}_\eta\right)\varepsilon = \frac{1}{2} - O_{\bar{c}_\gamma,\bar{c}_\eta}(\varepsilon),\end{aligned}$$

as desired.

■

We are left with proving Lemma 6.1.

Proof of Lemma 6.1: Similarly to what we did in the proof of Theorem 4.2, we will work with codes whose weight distribution deviates only slightly from that of a random code. While Theorem 4.2 required only an upper bound on the weights, here we also need a lower bound.

Write $p = 2^{-(n-k)}$ and $p' = 2^{-(n-k)} \cdot \left(1 - 2^{-k}\right)$ and fix $1 \leq i \leq n$. Observe that

$$\binom{n}{i} \cdot p' \leq \mathop{\mathbb{E}}_{\mathcal{C}}[\Delta_i] = \binom{n}{i} \cdot \frac{2^k - 1}{2^n - 1} \leq \binom{n}{i} \cdot p$$

and

$$\mathop{\mathrm{Var}}_{\mathcal{C}}[\Delta_i] \leq \binom{n}{i} \cdot p,$$

where the latter follows from Lemma 2.4. Fix a $\tau > 0$ to be determined soon. Markov's inequality now yields

$$\Pr_{\mathcal{C}}\left[\Delta_i \geq 2^{\tau n} \cdot \binom{n}{i} \cdot p\right] \leq 2^{-\tau n}$$

and Chebyshev's inequality yields

$$\Pr\left[\Delta_i \leq \frac{\binom{n}{i} \cdot p}{2^{\tau n}}\right] \leq \frac{1}{(1 - 2^{-\tau n} - 2^{-k})^2 \cdot \binom{n}{i} \cdot p}.$$

Fixing any $1 \leq \ell \leq \frac{n}{2}$ and taking a union bound, it holds with probability at least $1 - \frac{n}{2^{\tau n}} - \frac{n}{(1-2^{-\tau n}-2^{-k})^2 \cdot \binom{n}{\ell} \cdot p}$ that

$$\Delta_i \leq 2^{\tau n} \cdot \binom{n}{i} \cdot p \quad \text{for all } 1 \leq i \leq n \tag{19}$$

and

$$\Delta_i \geq \frac{\binom{n}{i} \cdot p}{2^{\tau n}} \quad \text{for all } \ell \leq i \leq \frac{n}{2}. \tag{20}$$

Note that Equation (19) means that $\mathcal{C}^\perp$ is τ-nice, in the sense of Definition 2.2. Take

$$\tau = \min\left\{\frac{1}{8}\gamma^2 \cdot \left(h_2^{-1}\left(1 - \frac{k}{n}\right)\right)^2, \frac{k \cdot \gamma}{n} - \frac{\log n}{n}\right\}, \ \ell = \left\lfloor h_2^{-1}\left(1 - \frac{2}{3} \cdot \frac{k}{n}\right)n \right\rfloor.$$

We bound the probability that Equations (19) and (20) then hold simultaneously, according to which of the two terms above minimize τ. If it is the first one, then the probability is at least

$$1 - \frac{n}{2^{\frac{1}{8}\gamma^2 \cdot \left(h_2^{-1}\left(1-\frac{k}{n}\right)\right)^2 \cdot n}} - \frac{n}{\left(1 - 2^{-\frac{1}{8}\gamma^2 \cdot \left(h_2^{-1}\left(1-\frac{k}{n}\right)\right)^2 \cdot n} - 2^{-k}\right)^2 \cdot \binom{n}{\ell} \cdot p} \triangleq 1 - \delta_1 - \delta_2.$$

To bound δ_1, we use the fact that $h_2^{-1}(1-x) \geq \frac{1}{2} - 5x^2$ whenever $x \leq \frac{1}{4}$ [GRS, Lemma B.2.4], and get $n\delta_1 \leq 2^{-(\gamma^2/128)n}$, since $h_2^{-1}(1 - \frac{k}{n}) \geq \frac{1}{4}$ follows from our upper bound on k. To bound δ_2, first note that $1 - 2^{-\gamma^2 \cdot \left(h_2^{-1}\left(1-\frac{k}{n}\right)\right)^2 \cdot n} - 2^{-k} \geq \frac{1}{2}$ since γ is large enough. Next, $\binom{n}{\ell} \geq \frac{1}{\sqrt{2n}} 2^{(1-\frac{2k}{3n})n}$, so $\binom{n}{\ell} p \geq \frac{1}{\sqrt{2n}} 2^{k/3} \geq 2^{k/4}$ (using our lower bound on k). We then have

$\delta_2 \le n \cdot 2^{-k/4}$, and overall, $\delta_1 + \delta_2 \le 2^{-(\gamma^2/256)n} + 2^{-k/6}$, again using our lower bound on k and the fact that γ is large enough.

Next, we bound the probability that Equations (19) and (20) in the case where $\tau = \frac{k\cdot\gamma}{n} - \frac{\log n}{n}$, namely,

$$1 - \frac{n}{2^{\left(\frac{k\cdot\gamma}{n} - \frac{\log n}{n}\right)n}} - \frac{n}{\left(1 - 2^{-\left(\frac{k\cdot\gamma}{n} - \frac{\log n}{n}\right)n} - 2^{-k}\right)^2 \cdot \binom{n}{\ell} \cdot p} \triangleq 1 - \delta_1' - \delta_2'$$

In this case, δ_1' can be upper bounded by $2^{-(\gamma/4)k}$ since $\frac{\log n}{n} \le \frac{k\cdot\gamma}{n}$ and by our lower bound on k. For δ_2', we again use the fact that $\binom{n}{\ell}p \ge 2^{k/4}$, and we also have that $1 - 2^{-\tau\cdot n} - 2^{-k} \ge \frac{1}{2}$, again from our lower bound on k. Thus, $\delta_1' + \delta_2' \le 2^{-\Omega(\gamma k)}$.

Assuming Equations (19) and (20) hold, we show that they imply Items 1 and 2, and begin with Item 1. By Equation (19), in order to bound $j^\star$, it suffices to solve the following equation for j:

$$\sum_{i=0}^{j} \binom{n}{i} 2^{-(n-k-\tau n)} \ge T.$$

Writing $j = \alpha n$ and using standard bounds on the sum of binomial coefficients, we need to find the smallest α for which αn is an integer, and

$$2^{h_2(\alpha)n - \frac{1}{2}\log n - 1} \cdot 2^{-n+k+\tau n} \ge T.$$

Thus, the α for which $j^\star = \alpha n$ satisfies

$$\alpha \ge h_2^{-1}\left(1 - \tau - \frac{k - \log T + \log n}{n}\right) \ge h_2^{-1}\left(1 - 2\cdot\frac{k - \log T}{n}\right),$$

since $\tau \le \frac{k\cdot\gamma}{n} - \frac{\log n}{n} \le \frac{k - \log T - \log n}{n}$.

We now prove Item 2. Let $j' = \lceil \alpha \cdot (1-\gamma) \cdot n \rceil$. By Equation (19),

$$A \triangleq \sum_{i=0}^{j'-1} \Delta_i \le \sum_{i=0}^{j'-1} \binom{n}{i} \cdot 2^{-n+k+\tau n} \le 2^{n\left(h_2(\alpha(1-\gamma))+\tau+\frac{k}{n}-1\right)}.$$

By Item 1 and our assumption that $T \ge 2^{\frac{2k}{3}}$,

$$j^\star = \alpha \cdot n \ge h_2^{-1}\left(1 - 2\cdot\frac{k - \log T}{n}\right)\cdot n \ge h_2^{-1}\left(1 - \frac{2}{3}\cdot\frac{k}{n}\right)\cdot n \ge \ell.$$

Hence, Equation (20) yields $\Delta_{j^\star} \ge \binom{n}{i} \cdot p \cdot 2^{-\tau n}$, and so,

$$B \triangleq \sum_{i=j'+1}^{j^*} \Delta_i \ge \Delta_{j^\star} \ge \binom{n}{j^\star} \cdot p \cdot 2^{-\tau n} \ge 2^{n\cdot\left(h_2(\alpha) - \frac{\log(2n)}{2n} - 1 + \frac{k}{n} - \tau\right)}.$$

Therefore,

$$\frac{\sum_{i=0}^{j^\star} i\cdot\Delta_i}{\sum_{i=0}^{j^\star}\Delta_i} \geq \frac{j'\cdot\sum_{i=j'}^{j^\star}\Delta_i}{\sum_{i=0}^{j^\star}\Delta_i} = j'\cdot\frac{B}{A+B} \geq (1-\gamma)\cdot j^\star\cdot\frac{1}{1+\frac{A}{B}}.$$

Now,

$$\begin{aligned}
\frac{\log\left(\frac{A}{B}\right)}{n} &\leq h_2(\alpha(1-\gamma)) - h_2(\alpha) + 2\tau + \frac{\log(2n)}{2n} \\
&\leq h_2(\alpha(1-\gamma)) - h_2(\alpha) + 4\tau && \text{by the lower bound on } \gamma \\
&\leq -(\alpha\gamma)^2 + 4\tau && \text{since } h_2'(x)\geq 0 \text{ and } h_2''(x)\leq -2 \\
&\leq -(\alpha\gamma)^2 + \frac{1}{2}\gamma^2\cdot\left(h_2^{-1}\left(1-\frac{k}{n}\right)\right)^2 \\
&\leq -\frac{1}{2}\gamma^2\cdot\left(h_2^{-1}\left(1-\frac{k}{n}\right)\right)^2 \triangleq -\theta. && \text{by Item 1 and the assumption } T\geq 2^{\frac{2k}{3}}.
\end{aligned}$$

We conclude that

$$\frac{\sum_{i=0}^{j^\star} i\cdot\Delta_i}{\sum_{i=0}^{j^\star}\Delta_i} \geq \frac{(1-\gamma)j^\star}{1+2^{-\theta n}}.$$

All that is left is to show that $2^{-\theta n}\leq\gamma$. Indeed, this easily follows from our lower bound on k.

Finally, let us prove Item 3. Following the same reasoning as above, for some $\tau>0$, we have that $\Delta_{j^\star+1}\leq 2^{\tau n}\cdot\binom{n}{j^\star+1}\cdot p$ and $\Delta_i\geq 2^{-\tau n}\cdot\binom{n}{j^\star}\cdot p$ with probability at least

$$1-\frac{1}{2^{\tau n}}-\frac{1}{(1-2^{-\tau n}-2^{-k})^2\cdot\binom{n}{j^\star}\cdot p}, \tag{21}$$

where $p=2^{-n+k}$. Now, we readily get

$$\frac{\Delta_{j^\star+1}}{\sum_{i=0}^{j^\star+1}\Delta_i}\leq\frac{\Delta_{j^\star+1}}{\Delta_{j^\star}}\leq\frac{2^{\tau n}\cdot\binom{n}{j^\star+1}\cdot p}{2^{-\tau n}\cdot\binom{n}{j^\star}\cdot p}\leq 2^{2\tau n}\cdot\frac{n-j^\star}{j^\star+1}\leq 2^{2\tau n+2},$$

using the fact that $j^\star\geq\frac{1}{4}n$. Set $\tau=\frac{1}{4\sqrt{n}}$. The above bound is thus at most $2^{\sqrt{n}}$, and the success probability, Equation (21), is at least

$$1-2^{-\frac{1}{4}\sqrt{n}}-\frac{1}{4}\cdot 2^{-k/4},$$

where we used $1-2^{-\tau n}-2^{-k}\geq\frac{1}{2}$, and

$$\binom{n}{j^\star}2^{-n+k}\geq 2^{\left(h_2(\alpha)-1+\frac{k}{n}-\frac{2\log n}{n}\right)n}\geq 2^{\left(\frac{k}{n}-2\cdot\frac{k-\log T}{n}-\frac{2\log n}{n}\right)n}\geq 2^{\left(\frac{k}{3n}-\frac{2\log n}{n}\right)n}\geq 2^{-k/4}.$$

■

Acknowledgements

We thank Amnon Ta-Shma for helpful and interesting discussions, and collaboration at the beginning of this work. This work was done partly while the authors were visiting the Simons Institute for the Theory of Computing.

References

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *Information Theory, IEEE Transactions on*, 38(2):509–516, 1992.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AJQ+20] Vedat Levi Alev, Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. List decoding of direct sum codes. In *Proceedings of the 31st Symposium on Discrete Algorithms (SODA 2020)*, pages 1412–1425. ACM-SIAM, 2020.

[BD22] Guy Blanc and Dean Doron. New near-linear time decodable codes closer to the GV bound. In *Proceedings of the 37th Computational Complexity Conference (CCC 2022)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2022.

[BT13] Avraham Ben-Aroya and Amnon Ta-Shma. Constructing small-bias sets from algebraic-geometric codes. *Theory of Computing*, 9(5):253–272, 2013.

[For65] G. David Forney. Concatenated codes. Technical Report 440, Research Laboratory of Electronics, MIT, 1965.

[GI04] Venkatesan Guruswami and Piotr Indyk. Efficiently decodable codes meeting gilbert-varshamov bound for low rates. In *Proceedings of the 15th Symposium on Discrete Algorithms (SODA 2004)*, pages 756–757. ACM-SIAM, 2004.

[GI05] Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005.

[Gil52] Edgar N. Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.

[GM22] Venkatesan Guruswami and Jonathan Mosheiff. Punctured Low-Bias Codes Behave Like Random Linear Codes. In *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS 2022)*, pages 36–45. IEEE, 2022.

[GR08] Venkatesan Guruswami and Atri Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Transactions on Information Theory*, 54(1):135–150, 2008.

[GR10] Venkatesan Guruswami and Atri Rudra. The existence of concatenated codes list-decodable up to the hamming bound. *IEEE Transactions on information theory*, 56(10):5195–5206, 2010.

[GRS] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*.

[GW13] Venkatesan Guruswami and Carol Wang. Linear-algebraic list decoding for variants of Reed-Solomon codes. *IEEE Transactions on Information Theory*, 59(6):3257–3268, 2013.

[HRZW19] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes and applications. *SIAM Journal on Computing*, pages FOCS17–157, 2019.

[JMV90] Dieter Jungnickel, Alfred J. Menezes, and Scott A. Vanstone. On the number of self-dual bases of GF(q^m) over GF(q). *Proceedings of the American Mathematical Society*, 109(1):23–29, 1990.

[JQST20] Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. Unique decoding of explicit ε-balanced codes near the Gilbert–Varshamov bound. In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pages 434–445. IEEE, 2020.

[JST21] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of Ta-Shma's codes via splittable regularity. In *Proceedings of the 53rdth Annual Symposium on Theory of Computing (STOC 2021)*, pages 1527–1536. ACM, 2021.

[JST23] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. List decoding of tanner and expander amplified codes from distance certificates. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS 2023)*, pages 1682–1693. IEEE, 2023.

[Jus72] Jørn Justesen. Class of constructive asymptotically good algebraic codes. *IEEE Transactions on Information Theory*, 18(5):652–656, 1972.

[Kop15] Swastik Kopparty. List-decoding multiplicity codes. *Theory of Computing*, 11(1):149–182, 2015.

[KRRZ+20] Swastik Kopparty, Nicolas Resch, Noga Ron-Zewi, Shubhangi Saraf, and Shashwat Silas. On list recovery of high-rate tensor codes. *IEEE Transactions on Information Theory*, 67(1):296–316, 2020.

[KV03] Ralf Koetter and Alexander Vardy. Algebraic soft-decision decoding of reed-solomon codes. *IEEE Transactions on Information Theory*, 49(11):2809–2825, 2003.

[Mas63] James L. Massey. Threshold decoding. Technical Report 410, Research Laboratory of Electronics, MIT, 1963.

[MRSY24] Jonathan Mosheiff, Nicolas Resch, Kuo Shang, and Chen Yuan. Randomness-efficient constructions of capacity-achieving list-decodable codes. *arXiv preprint*, 2024.

[MU17] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[PP24] Aaron (Louie) Putterman and Edward Pyne. Pseudorandom Linear Codes Are List-Decodable to Capacity. In *Proceedings of the 15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pages 90:1–90:21. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.

[RR23] Silas Richelson and Sourya Roy. Gilbert and Varshamov meet Johnson: List-decoding explicit nearly-optimal binary codes. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS 2023)*, pages 194–205. IEEE, 2023.

[Rud07] Atri Rudra. *List decoding and property testing of error-correcting codes*. University of Washington, 2007.

[SL80] Gadiel Seroussi and Abraham Lempel. Factorization of symmetric matrices and trace-orthogonal bases in finite fields. *SIAM Journal on Computing*, 9(4):758–767, 1980.

[Ta-17] Amnon Ta-Shma. Explicit, almost optimal, ε-balanced codes. In *Proceedings of the 49th Annual Symposium on Theory of Computing (STOC 2017)*, pages 238–251. ACM, 2017.

[Tho83] Christian Thommesen. The existence of binary linear concatenated codes with Reed–Solomon outer codes which asymptotically meet the Gilbert–Varshamov bound. *IEEE Transactions on Information Theory*, 29(6):850–853, 1983.

[Var57] Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

[Zya71] Victor Vasilievich Zyablov. An estimate of the complexity of constructing binary linear cascade codes. *Problemy Peredachi Informatsii*, 7(1):5–13, 1971.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il