Electronic Colloquium on Computational Complexity, Report No. 122 (2019)

LDPC Codes Achieve List Decoding Capacity*

Jonathan Mosheiff[1], Nicolas Resch[2], Noga Ron-Zewi[3], Shashwat Silas[4], and Mary Wootters[4]

[1] *Weizmann Institute*
[3] *University of Haifa*
[2] *Carnegie Mellon University*
[4] *Stanford University*

September 13, 2019

Abstract

We show that Gallager's ensemble of Low-Density Parity Check (LDPC) codes achieve list-decoding capacity. These are the first graph-based codes shown to have this property. Previously, the only codes known to achieve list-decoding capacity were completely random codes, random linear codes, and codes constructed by algebraic (rather than combinatorial) techniques. This result opens up a potential avenue towards truly linear-time list-decodable codes which achieve list-decoding capacity.

Our result on list decoding follows from a much more general result: any *local* property satisfied with high probability by a random linear code is also satisfied with high probability by a random LDPC code from Gallager's distribution. Local properties are properties characterized by the exclusion of small sets of codewords, and include list-decoding, list-recovery and average-radius list-decoding. Along the way, we give a characterization of sets of codewords that are likely to appear in a random linear code, which may be of independent interest.

*JM is partially supported by Irit Dinur's ERC-CoG grant 772839. NRe is partially supported by NSERC grant CGSD2-502898, NSF grants CCF-1422045, CCF-1814603, CCF-1527110, CCF-1618280, CCF-1910588, NSF CAREER award CCF-1750808 and a Sloan Research Fellowship. NRo Is partially supported by BSF grant 2014359. SS and MW are partially supported by NSF grants CCF-1844628, CCF-1814629, and a Sloan Research Fellowship. SS is partially supported by a Google Graduate Fellowship.

ISSN 1433-8092

1 Introduction

In this paper, we study sets $C \subset \Sigma^n$ of strings of length n, with the combinatorial property that not too many elements of C are contained in any small enough Hamming ball. In the language of coding theory, such a C is a *list-decodable code.* List-decoding is an important primitive in coding theory, with applications ranging from communication to complexity theory. However, as discussed below, most constructions of *capacity-achieving* (aka, optimal) list-decodable codes are fundamentally algebraic, despite a rich history of combinatorial—and in particular, graph-based—constructions of error correcting codes.

We show that a random ensemble of *Low-Density Parity-Check (LDPC) codes* achieves list-decoding capacity with high probability. LDPC codes are the prototypical example of graph-based codes, and are popular both in theory and in practice because of their extremely efficient algorithms. One of the motivations for this work is that we do not currently know any linear-time algorithms for list-decoding any code up to capacity; since graph-based codes offer linear-time algorithms for a variety of other coding-theoretic tasks, our result opens up the possibility of using these constructions for linear-time list-decoding algorithms.

List Decoding. Formally, a code $C \subset \Sigma^n$ is (α, L)-list-decodable if for all $z \in \Sigma^n$,

$$|\{c \in C \,:\, \mathrm{dist}(c, z) \leq \alpha\}| \leq L.$$

Above, $\mathrm{dist}(c, z)$ is relative Hamming distance,

$$\mathrm{dist}(c, z) = \frac{1}{n}|\{i \,:\, c_i \neq z_i\}|.$$

Elements $c \in C$ are called codewords, Σ is called the alphabet, and n is called the length of the code.

The fundamental trade-off in list-decoding is between the parameter α and the size $|C|$ of the code, given that the list size L is reasonably small. We would like both α and $|C|$ to be large, but these requirements are at odds: the larger the code C is, the closer together the codewords have to be, which means that α cannot be as large before some Hamming ball of radius α has many codewords in it. The size of a code C is traditionally quantified by the rate R of C, which is defined as

$$R = \frac{\log_{|\Sigma|}(|C|)}{n}.$$

The rate of C is a number between 0 and 1, and larger rates are better.

List decoding has been studied since the work of Elias and Wozencraft in the 1950's [Eli57, Woz58], and by now we have a good understanding of what is possible and what is not. The classical *list-decoding capacity theorem* states that there exist codes over alphabets of size $|\Sigma| = q$ and of rate $R \geq 1 - h_q(\alpha) - \varepsilon$ which are $(\alpha, 1/\varepsilon)$-list-decodable, where

$$h_q(x) := x \log_q(q-1) - x \log_q(x) - (1-x)\log_q(1-x) \tag{1}$$

is the q-ary entropy function. Conversely, any such code with rate $R \geq 1 - h_q(\alpha) + \varepsilon$ must have exponential list sizes, in the sense that there is some $z \in \Sigma^n$ so that $|\{c \in C \,:\, \mathrm{dist}(c, z) \leq \alpha\}| = \exp_{\varepsilon,\alpha}(n)$.[1]

[1]Here and throughout the paper, $\exp(n)$ denotes $2^{\Theta(n)}$, and subscripts indicate that we are suppressing the dependence on those parameters.

A code of rate $R \geq 1 - h_q(\alpha) - \varepsilon$ that is (α, L)-list decodable for $L = O_{\varepsilon,\alpha}(1)$ is said to achieve list-decoding capacity, and a major question in list-decoding is which codes have this property. By now we have three classes of examples. First, it is not hard to see that completely random codes achieve list-decoding capacity with high probability. Second, a long line of work (discussed more below) has established that *random linear codes* do as well: we say that a code over the alphabet $\Sigma = \mathbb{F}_q$ is linear if it is a linear subspace of $\mathbb{F}_q^n$,[2] and a random linear code is a random subspace. Third, there are several explicit constructions of codes which achieve list-decoding capacity; as discussed below, these constructions all rely importantly on algebraic techniques.

LDPC Codes. Graph-based codes, such as LDPC codes, are a class of codes which is notably absent from the list of capacity-achieving codes above. Originally introduced by Gallager in the 1960's [Gal62], codes defined from graphs have become a class of central importance in the past 30 years.

Here is one way to define a code using a graph. Suppose that $G = (V, W, E)$ is a bipartite graph with $|V| = n$ and $|W| = m$ for $m \leq n$. Then G naturally defines a linear code $C \subset \mathbb{F}_q^n$ of rate at least $1 - m/n$ as follows:

$$C = \left\{ c \in \mathbb{F}_q^n \,:\, \forall j \in W, \sum_{i \in \Gamma(j)} \alpha_{i,j} c_i = 0 \right\},$$

where $\Gamma(i)$ denotes the neighbors of i in G and $\alpha_{i,j} \in \mathbb{F}_q$ are fixed coefficients. (See Figure 1). That is, each vertex in W serves as a parity check, and the code is defined as all possible labelings of vertices in V which obey all of the parity checks. When the right-degree of G is small, the resulting code is called a Low-Density Parity Check (LDPC) code.

LDPC codes and related constructions (in particular, Tanner codes [Tan81] and expander codes [SS94, Zém01]) are notable for their efficient algorithms for unique decoding; in fact, the only linear-time encoding/decoding algorithms we have for unique decoding (that is, list decoding with $L = 1$) are based on such codes.

Motivating question and our results. We currently do not know of any purely combinatorial constructions of capacity-achieving list-decodable codes. We also currently do not know of any linear-time algorithms to list-decode any code to capacity. Since graph-based codes and LDPC codes in particular are notable for their linear-time algorithms, this state of affairs motivates the following question:

Question 1.1. *Are there (families) of LDPC codes that achieve list-decoding capacity?*

In this work we show that the answer to Question 1.1 is "yes." In fact, we show a stronger result: random LDPC codes (the same ensemble studied by Gallager in his seminal work nearly 60 years ago [Gal62]), satisfy with high probability any sufficiently nice ("local") property that random linear codes satisfy with high probability. It turns out that list-decodability is such a property, and this answers Question 1.1. Along the way, we develop a characterization of sets of codewords contained in a random linear code, which may be of independent interest. We describe our main results in more detail below.

[2] Here and throughout the paper, $\mathbb{F}_q$ denotes the finite field with q elements.

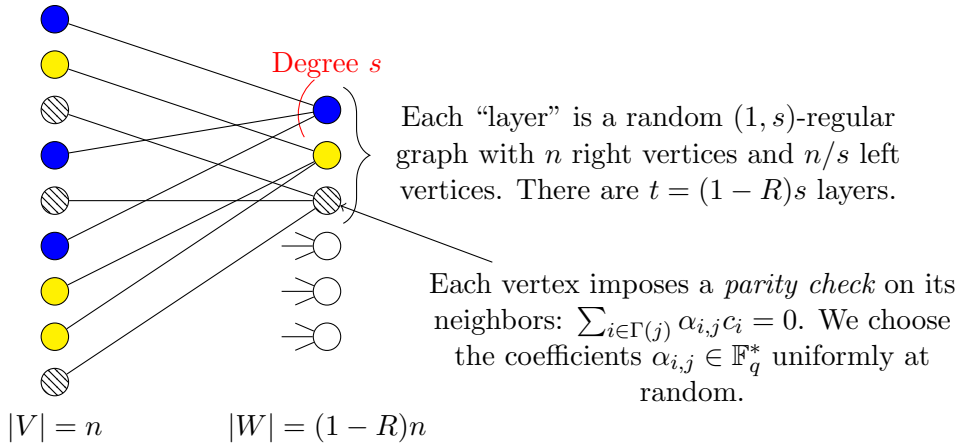

Figure 1: A random (t, s)-regular bipartite graph that gives rise to a random s-LDPC code of rate R. Here, we set $t := s(1 - R)$.

1.1 Main Results

Random LDPC codes achieve list-decoding capacity. We study the so-called "Gallager ensemble" of binary LDPC codes introduced by Gallager in the 1960's [Gal62], as well as its natural generalization to larger alphabets.[3]

Fix a rate $R \in (0, 1)$ and a sparsity parameter s, and let $t = (1 - R)s$. We assume that t is an integer. To define the ensemble of random s-LDPC codes of rate R, we need to specify a distribution on the underlying bipartite graphs and a distribution on the coefficients $\alpha_{i,j}$. We define the distribution on graphs as follows. Let $G_i = (V, W_i, E_i)$ for $i = 1, \ldots, t$ be independent uniformly random $(1, s)$-regular bipartite graphs with a shared left vertex set V of size n and disjoint right vertex sets W_i, each of size n/s. Then let $G = (V, W, E)$ be the union of these graphs, where $W = \bigcup_{i=1}^{t} W_i$. Finally, we choose the coefficients $\alpha_{i,j}$ for $(i, j) \in E$ to be uniformly random in $\mathbb{F}_q^*$. The ensemble of s-random LDPC codes of rate R is illustrated in Figure 1.

Our main theorem about the list-decodability of random LDPC codes is a reduction from the list-decodability of random linear codes:

Theorem 1.2. *For any $R \in (0, 1)$, $\varepsilon > 0$, prime power q, and $L \geq 1$ there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$. Suppose that a random linear code of rate R over $\mathbb{F}_q$ is (α, L)-list decodable with high probability. Then a random s-LDPC code of rate $R - \varepsilon$ over $\mathbb{F}_q$ is (α, L)-list decodable with high probability.*

Remark 1.3 (The parity of s)**.** *All of our results hold for even s as well as odd s. However, the proof is slightly simpler for odd s, so for clarity we state and prove the theorem in this case.*

Instantiating this with a result of [GHK11] on list decoding of random linear codes, we get the following corollary.

[3] For binary codes, our definition coincides with Gallager's. For larger alphabets our definition is somewhat different: Gallager's ensemble chooses the coefficients $\alpha_{i,j}$ to be all ones, while we choose them to be random elements of $\mathbb{F}_q^*$.

Corollary 1.4. *For any prime power q, $\alpha \in (0, 1 - 1/q)$, and $\varepsilon \in (0, 1 - h_q(\alpha))$ there exists $L = O_\alpha(1/\varepsilon)$ and $s \geq 1$ so that a random s-LDPC code of rate $1 - h_q(\alpha) - \varepsilon$ over $\mathbb{F}_q$ is (α, L)-list-decodable with high probability.*

Remark 1.5 (Other parameter regimes)**.** *We state Corollary 1.4 as one example of what can be obtained by combining Theorem 1.2 with one result on random linear codes. The result of [GHK11] degrades as $\alpha \to 1-1/q$, and so Corollary 1.4 degrades as well. However, there has been a great deal of work on the list-decodability of random linear codes as $\alpha \to 1 - 1/q$ (summarized in Section 1.2 below), and Theorem 1.2 implies that these results carry over to random LDPC codes as well.*

Random LDPC codes achieve any local property that random linear codes achieve. Theorem 1.2 follows as a corollary of a much more general theorem. We show that any "local" property which is satisfied by random linear codes with high probability is also satisfied by random LDPC codes with high probability.

Informally, a local property is a property which can be defined by the exclusion of certain bad sets. For example, a code C is (α, L)-list-decodable if it does *not* contain any sets $B \subset \Sigma^n$ of size larger than L so that B is contained in a Hamming ball of radius α. Along with list-decodability, local properties include many related notions like *list recovery, average-radius list decoding,* and *erasure list decoding.* A long line of work (discussed more in Section 1.2) has established that these properties hold for random linear codes with high probability, so our reduction immediately implies that they hold with high probability for LDPC codes as well.

Formally, we define a local property as follows. Let $\pi : [n] \to [n]$ be a permutation on $[n]$. For a string $x \in \Sigma^n$, we let $\pi(x) \in \Sigma^n$ denote the string obtained by permuting the coordinates of x according to π, and for a subset $B \subseteq \Sigma^n$, we let $\pi(B) := \{\pi(x) \mid x \in B\}$. We say that a collection $\mathcal{B}$ of subsets of Σ^n is permutation invariant if for any $B \in \mathcal{B}$ we also have that $\pi(B) \in \mathcal{B}$.

Definition 1.6 (Local property)**.** *Let $\mathcal{P} = \{P_n\}_{n \in \mathbb{N}}$, where each P_n is a property of length n codes over Σ. We say that $\mathcal{P}$ is a b-local property if for any $n \in \mathbb{N}$ there exists a permutation-invariant collection $\mathcal{B}_n$ of subsets of Σ^n, where $|B| \leq b$ for all $B \in \mathcal{B}_n$, such that*

$$C \subseteq \Sigma^n \text{ satisfies } P_n \iff B \nsubseteq C \text{ for all } B \in \mathcal{B}_n.$$

We say that a random ensemble of codes $\mathcal{C} = \{C_n\}_{n \in \mathbb{N}}$ satisfies $\mathcal{P}$ with high probability if $\lim_{n \to \infty} \Pr[C_n \text{ satisfies } P_n] = 1$.

A random linear code of rate R over $\mathbb{F}_q$ is defined[4] as the kernel of a uniformly random matrix $H \in \mathbb{F}_q^{(1-R)n \times n}$. Notice that such a code has rate R with high probability.

Our main theorem is that any local property which is satisfied with high probability by a random linear code is also satisfied with high probability by a random LDPC code:

Theorem 1.7 (Main)**.** *For any $R \in (0, 1)$, $\varepsilon > 0$, prime power q, and $b \geq 1$ there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$. Let $\mathcal{P}$ be a b-local property, and suppose that a random linear code of rate R over $\mathbb{F}_q$ satisfies $\mathcal{P}$ with high probability. Then a random s-LDPC code of rate $R - \varepsilon$ over $\mathbb{F}_q$ satisfies $\mathcal{P}$ with high probability.*

[4]There are a few natural ways to define a random linear code: for example we could also define it as a uniformly random subspace of dimension Rn, or we could define it as the image of a uniformly random $n \times Rn$ matrix, or we could define it as we do here, as the kernel of a uniformly random $(1 - R)n \times n$ matrix. It can be shown that these distributions are quite close to each other, and in particular, any property that holds for one with high probability holds for the others.

Remark 1.8 (The dependence on ε, R, q, b)**.** *The parameter s_0 depends on ε, R, q and b. An inspection of the proof (c.f. Remarks 4.3 and 5.3) shows that we may take*

$$s_0 = O\left(\frac{b\log(q) + \log(q/\varepsilon)}{h_q^{-1}(1-R-\varepsilon)}\right).$$

Theorem 1.2 follows immediately from Theorem 1.7. Indeed, (α, L)-list-decodability is a local property: for any $n \in \mathbb{N}$, let $\mathcal{B}_n$ be the collection of all sets of $L+1$ vectors in $\mathbb{F}_q^n$ contained in some Hamming ball of radius α.

The existence of a reduction like the one in Theorem 1.7 is surprising, at least to the authors. There is a lot more structure in a random LDPC code that in a random linear code. For example, we know of linear-time unique decoding algorithms for random LDPC codes,[5] but it is unlikely that any efficient unique decoding algorithm exists for random linear codes.[6] Thus it is unexpected that this much more structured ensemble would share many properties—in a black-box way—with random linear codes.

We give a high-level overview of the proof of Theorem 1.7 in Section 2 below after a discussion of related work in Section 1.2.

1.2 Related Work

List-decodability of random ensembles of codes. As mentioned above, it is not hard to see that a completely random code $C \subset \Sigma^n$ achieves list-decoding capacity. There has also been work studying more structured random ensembles of codes, notably random linear codes. Zyablov and Pinsker [ZP81] showed that random linear codes of rate $1 - h_q(\alpha) - \varepsilon$ are (α, L)-list-decodable with high probability, where L is independent of n but depends exponentially on $1/\varepsilon$. Two decades later, [GHSZ02] showed that there exist binary linear codes with list-size $O(1/\varepsilon)$, and their techniques were recently extended to hold with high probability in [LW18]. In the meantime, [GHK11] showed that random linear codes over any alphabet achieve capacity with $L = O(1/\varepsilon)$ when α is bounded away from $1 - 1/q$; [CGV13, Woo13, RW14, RW18] extended these results to get list sizes nearly as good even for large α, although the problem is still open in some parameter regimes.

Several variants of list-decoding have been studied for random linear codes, including *list-recovery* [RW18], *average-radius list-decoding* [Woo13, RW14, RW18], and list-recovery from erasures [Gur03].[7] All of these properties are local, and so our main theorem implies that LDPC codes satisfy them with high probability.

List-decodability of explicit codes. Obtaining explicit constructions of codes which achieve list-decoding capacity was a major open problem until it was solved about a decade ago. The first

[5] This follows, for example, from [SS94] because the underlying random graph is with high probability a good expander.

[6] Unique decoding of random linear codes is related to the problem of Learning Noisy Parities (LNP) and Learning With Errors (LWE), which are thought to be hard.

[7] List-recovery is a generalization of list-decoding where the input is a list of sets $Z_1, \ldots, Z_n$ of size at most ℓ (instead of a received word $z \in \Sigma^n$, which can be seen as the $\ell = 1$ case), and goal is to find all of the codewords $c \in C$ so that $c_i \in Z_i$ for at least a $1-\alpha$ fraction of the $i \in [n]$. Average-radius list-decoding is strengthening of list-decoding where instead of requiring that no set of $L+1$ codewords has *all* codewords close to some z, we require that no set of $L+1$ codewords is on-average close to z. List-decoding from erasures is a weaker notion than list-decoding, where $z \in (\Sigma \cup \{\perp\})^n$ has some *erased* symbols, and the goal is to recover all $c \in C$ which agree with z on all the observed places.

explicit codes to provably achieve capacity were the *Folded Reed-Solomon Codes* of Guruswami and Rudra [GR08]. These codes are variants on the classic *Reed-Solomon codes* and are based on polynomials over finite fields. Since then, there have been several constructions of such codes, also based on algebraic techniques, including *Univariate Multiplicity Codes* [GW13, Kop15, KRSW18], variants of Algebraic-Geometry Codes [GX12, GX13], and manipulations of these codes [DL12, GK16, HRW17, KRRZ+19]. However, the state-of-the-art for explicit constructions still requires quite large (but constant) alphabet and list sizes. These codes can be efficiently list-decoded in polynomial time; the fastest algorithm is that of [HRW17, KRRZ+19], which runs in nearly-linear time $O(n^{1+o(1)})$.

While graph-based techniques have been used to modify the underlying algebraic constructions (for example the expander-based distance-amplification technique of [AEL95] is used in [HRW17, KRRZ+19] to obtain near-linear-time list-decoding), to the best of our knowledge there are no results establishing list-decodability up to capacity for purely graph-based codes such as LDPC codes or expander codes.[8]

LDPC Codes Achieve Capacity on the Binary Symmetric Channel. LDPC Codes have been studied extensively in the context of unique decoding, especially in a model of random errors. Informally, a code is said to achieve capacity on the Binary Symmetric Channel (BSC) if there is some algorithm which can, with high probability, uniquely decode a code of rate $R = 1 - h_2(\alpha) - \varepsilon$ from an α-fraction of *random* errors. It is known that Gallager's LDPC codes nearly achieve capacity on the BSC as n gets large, under maximum-likelihood decoding [Gal62, Gur06], and recently it was shown that certain LDPC codes achieve capacity for smaller block lengths under efficient decoding algorithms as well [KRU13]. Achieving capacity on the BSC is related to achieving list-decoding capacity (in particular, the capacities are the same, $R = 1 - h_q(\alpha)$). However, there is no formal connection along these lines, and to the best of our knowledge these results about the BSC do not imply anything about the list-decodability of LDPC codes.

1.3 Discussion and open questions

In this work, we answer Question 1.1 with a very strong "yes." There are LDPC codes that achieve list-decoding capacity, and moreover there are many of them, and moreover these codes also likely satisfy any local property—that is, any property which can be defined by ruling out small bad sets of codewords—which is likely satisfied by a random linear code. However, we feel that our results are just the tip of the iceberg. They raise several interesting questions:

1. **What other properties are local?** We have shown that random LDPC codes satisfy with high probability any local property that random linear codes satisfy with high probability. There are several natural examples of local properties, including distance, list-decoding and list-recovery. What other examples are there?

2. **Derandomization?** Our results hold for a random ensemble of LDPC codes. It is natural to ask whether (or to what extent) this construction can be derandomized. In particular, it does not seem as though the underlying graph being an expander would be sufficient.

[8] We note that [HW18] give capacity-achieving graph-based codes for zero-error list-recovery (with erasures), where the input is lists $Z_1, \ldots, Z_n$ so that most lists have small size, and the goal is to return all codewords $c \in C$ that satisfy $c_i \in Z_i$ for all i. It does not seem easy to adapt these techniques for general list-recovery and hence for list-decoding.

3. **Algorithms?** Our results are combinatorial, but one of our main motivations is algorithmic. At the moment we do not know of any truly linear-time list-decoding algorithms for any capacity-achieving list-decodable codes. Since essentially all known linear-time algorithms in coding theory arise from graph-based codes, such codes are a natural candidate for linear-time list-decoding. Now that we know that random LDPC codes achieve list-decoding capacity combinatorially, can we list-decode them efficiently?

1.4 Organization

In Section 2, we give a high-level overview of the proof of our main theorem, Theorem 1.7. This proof relies on three building blocks:

- First, Theorem 2.2 characterizes the sorts of sets $B \subset \Sigma^n$ that are contained in a random linear code. We prove this theorem in Section 3. As discussed in Section 2.2, this result is reminiscent of results about local properties of random graphs.
- Second, Lemma 2.5 shows that sets B that meet the characterization above are contained in an s-random LDPC code with roughly the same probability that they are contained in a random linear code, provided that the s-random LDPC code has good distance: the distance of a code is the minimum distance between any two codewords. We prove this lemma in Section 4 using Fourier analysis.
- Third, Theorem 2.6 shows that s-random LDPC codes do indeed have good distance. This was already shown by Gallager in the binary case; we give an alternative proof of this fact that also extends to large alphabets. We prove Theorem 2.6 in Section 5 using techniques from exponential families.

2 High-level idea: proof of Theorem 1.7

In this section we prove our main theorem (Theorem 1.7) using the building blocks outlined in Section 1.4. We will establish these building blocks in later sections. The purpose of this section is to give a high-level idea of the structure of the proof, deferring the technical parts to later sections.

2.1 Notation and definitions

Because we are studying local properties, we need some notation around sets $B \subseteq \mathbb{F}_q^n$. For such a set B of size ℓ, it will be convienient to view B as a matrix $M \in \mathbb{F}_q^{n \times \ell}$ with the elements of B as the columns. (The ordering of the columns will not matter). We say that M is contained in a code $C \subseteq \mathbb{F}_q^n$ (written "$M \subset C$") if all of the columns of M belong to C.

The notion of permutation-invariant properties leads us to think about permutations of the *rows* of such a matrix $M \in \mathbb{F}_q^{n \times \ell}$. Motivated by this, we define the row distribution of M to be τ_M: that is, for any $v \in \mathbb{F}_q^\ell$ we let

$$\tau_M(v) := \frac{\text{number of appearances of } v \text{ as a row in } M}{n}.$$

Let $\mathcal{D}_{n,\ell}$ denote the collection of possible row distributions of matrices in $\mathbb{F}_q^{n\times\ell}$, *i.e.,* distributions τ over $\mathbb{F}_q^\ell$ where $\tau(v)\cdot n \in \mathbb{N}$ for any $v \in \mathrm{Supp}(\tau)$.[9] The number of possible row distributions of matrices in $\mathbb{F}_q^{n\times\ell}$ is just the number of ways to partition n things into at most q^ℓ groups, so

$$|\mathcal{D}_{n,\ell}| \le \binom{n+q^\ell-1}{q^\ell-1}. \tag{2}$$

For a distribution $\tau \in \mathcal{D}_{n,\ell}$, let $\mathcal{M}_{n,\tau}$ denote the collection of matrices $M \in \mathbb{F}_q^{n\times\ell}$ with row distribution τ.

Finally, let $H(\tau)$ and $H_q(\tau)$ denote the entropy and base-q-entropy of a random variable distributed according to τ:

$$H(\tau) := -\sum_{x\in\mathrm{Supp}(\tau)} \tau(x)\log(\tau(x)) \qquad \text{and} \qquad H_q(\tau) := \frac{H(\tau)}{\log q}.$$

Let

$$d(\tau) := \dim(\mathrm{span}(\mathrm{Supp}(\tau))).$$

We next define an *implicitly rare* distribution, which will be key in our characterization of what row distributions appear in a random linear code (Theorem 2.2 below).

Definition 2.1 (Implicitly rare distribution)**.** *We say that a distribution τ over $\mathbb{F}_q^\ell$ is γ-rare if*

$$H_q(\tau) < \gamma\cdot d(\tau).$$

We say that τ is γ-implicitly rare if there exists a matrix $A \in \mathbb{F}_q^{m\times\ell}$ such that the distribution Av for $v\sim\tau$ is γ-rare.

Example 1 (Implicitly rare distribution)**.** *Consider the distribution τ on $\mathbb{F}_q^\ell$ which is uniform on vectors whose first coordinate is in the set $S\subset\mathbb{F}_q$ for some set $S\neq\{0\}$ of size $|S|<q$. Then τ is γ-implicitly rare for $\gamma > \log_q(|S|)$. Indeed, let $A = [1\,0\,\ldots\,0] \in \mathbb{F}_q^{1\times n}$ be the matrix which projects onto the first coordinate. and consider the distribution τ' on $\mathbb{F}_q$ given by $A\cdot v$, $v\sim\tau$. Then τ' is a distribution on $\mathbb{F}_q^1$ with $H_q(\tau') = \log_q(|S|)$ and $d(\tau') = 1$.*

We will work with the parity-check matrix view of a random s-LDPC code C. Let $H \in \mathbb{F}_q^{(1-R)n\times n}$ be the adjacency matrix of the graph G in Figure 1 where the nonzero entries are given by the coefficients $\alpha_{i,j}$ of the parity checks. Then we can define a random s-LDPC code C as

$$C = \left\{x\in\mathbb{F}_q^n \,:\, H\cdot x = 0\right\}.$$

We introduce some notation to talk about the structure of H, which we will use throughout the paper. This is illustrated in Figure 2.

Let $F \in \{0,1\}^{(n/s)\times n}$ be the matrix $F = (F_1 \mid F_2 \mid \ldots \mid F_{n/s})$, where each $F_i \in \{0,1\}^{(n/s)\times s}$ has all-ones i-th row, and the rest of the rows are all-zeros. Let $\Pi \in \{0,1\}^{n\times n}$ be a random permutation matrix, and let $D \in \mathbb{F}_q^{n\times n}$ be a diagonal matrix with diagonal entries that are uniform in $\mathbb{F}_q^*$. Let $H_1,\ldots,H_{(1-R)\cdot s}$ be sampled independently according to the distribution $F\cdot\Pi\cdot D$. Then let $H \in \mathbb{F}_q^{(1-R)n\times n}$ be the matrix obtained by stacking $H_1,\ldots,H_{(1-R)\cdot s}$ on top of each other (see Figure 2). Then H is the parity-check matrix for a random s-LDPC code of rate R. We will refer to each H_i as a "layer" of H.

[9] Notice that $\mathcal{D}_{n,\ell}$ depends on q as well, but we suppress this dependence in the notation for readability.

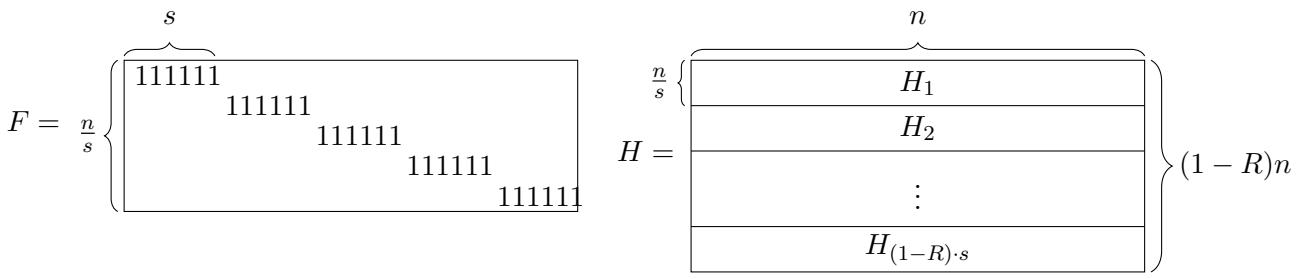

Figure 2: The matrices F and H. Each layer H_i of H is drawn independently according to the distribution $F \cdot \Pi \cdot D$, where $\Pi \in \{0,1\}^{n \times n}$ is a random permutation and $D \in \mathbb{F}_q^{n \times n}$ is a diagonal matrix with diagonal entries that are uniform in $\mathbb{F}_q^*$.

2.2 Characterization of matrices contained in a random linear code

The first building block is Theorem 2.2, which provides a characterization of matrices that are contained in a random linear code with high probability in terms of their row distributions. We prove this theorem in Section 3.

Theorem 2.2 (Characterization of matrices contained in random linear code)**.** *Suppose that $\tau \in \mathcal{D}_{n,\ell}$ is a $(1 - R - \varepsilon)$-implicitly rare distribution. Let $C \subseteq \mathbb{F}_q^n$ be a random linear code of rate R. Then*

$$\Pr\left[\exists M \in \mathcal{M}_{n,\tau}, M \subset C\right] \leq q^{-\varepsilon n}.$$

Conversely, suppose that $\tau \in \mathcal{D}_{n,\ell}$ is not $(1 - R + \varepsilon)$-implicitly rare. Let $C \subseteq \mathbb{F}_q^n$ be a random linear code of rate R. Then

$$\Pr\left[\exists M \in \mathcal{M}_{n,\tau}, M \subset C\right] \geq 1 - \binom{n + q^{2\ell} - 1}{q^{2\ell} - 1}^3 \cdot q^{-\varepsilon n}.$$

Theorem 2.2 can be seen as an analog to known results for local properties of random graphs. A local graph property is a property of a graph that is characterized by the graph containing, or not containing, a certain constant-sized subgraph. This notion is comparable to a linear code containing, or not containing, a constant number of codewords having a certain row distribution when viewed as matrix columns. In the case of graphs, the property should be permutation-invariant with regard to vertex labeling; in the linear code case, it should be permutation-invariant with regard to coordinate labeling.

Intuitively, the probability that a certain graph appears as a subgraph of $G(n,p)$ grows as p approaches 1. It is well-known (see for example [Bol01, Sec. 4.2]) that for any constant-sized graph H there is some threshold p_0^H such that the expected number of appearances of H as a subgraph of $G(n,p)$ either tends to 0 or ∞ depending on whether p is smaller or larger than p_0^H.

Similarly, in the linear code setting, the probability that some matrix with a certain row distribution is contained in a random linear code of rate R grows as R approaches 1. Moreover, it is not hard to show that the expected number of appearances of a matrix M of row distribution τ either tends to 0 or ∞ depending on whether R is smaller or larger than $1 - \frac{H_q(\tau)}{d(\tau)}$, or equivalently on whether $H_q(\tau)$ is smaller or larger than $(1 - R) \cdot d(\tau)$. This corresponds to our notion of rare distribution.

Note that if H appears as a subgraph of G, then all subgraphs of H also appear as subgraphs of G. Thus $G(n,p)$ contains an H-subgraph with high probability only if any subgraph H' of H is also contained in $G(n,p)$ with high probability. It turns out that this condition is in fact both necessary and sufficient, and $G(n,p)$ contains an H-subgraph with probability tending either to 0 or 1 depending on whether p is smaller or larger than the minimum of $p_0^{H'}$ over all subgraphs H' of H.

In our setting, if a linear code C contains a matrix M, then it also contains any linear image of M, which corresponds to our notion of implicitly rare distribution. Theorem 2.2 then shows that a random linear code of rate R contains a matrix M with certain row distribution τ with probability tending to either 0 or 1 depending on whether R is smaller or larger than the minimum of $1 - \frac{H_q(\tau')}{d(\tau')}$ over all linear images τ' of τ. Interestingly, Theorem 2.2 is proven using similar methods to those used to prove the analogous facts about random graphs.

2.3 Matrices contained in a random s-LDPC code

The second building block shows that any implicitly rare distribution that also happens to be δ-*smooth* (defined below) is unlikely to show up in a random LDPC code.

Definition 2.3 (Smooth distribution)**.** *We say that a distribution τ over $\mathbb{F}_q^\ell$ is δ-smooth if for any $u \in \mathbb{F}_q^\ell \setminus \{0\}$, $\Pr_{v\sim\tau}[\langle u, v\rangle \neq 0] \geq \delta$.*

Remark 2.4 (Relationship to distance)**.** *In coding-theoretic terms, a matrix M is δ-smooth if and only if the code $\left\{Mu : u \in \mathbb{F}_q^\ell\right\}$ has relative distance at least δ. Indeed, the relative weight of any codeword Mu in this code is*

$$\frac{1}{n}\sum_{i\in[n]} \mathbf{1}_{\langle u, e_i^T M\rangle \neq 0} = \Pr_{v\sim\tau}[\langle u, v\rangle \neq 0].$$

The following lemma bounds the probability that a matrix with smooth and implicitly rare row distribution is contained in a random LDPC code with sufficiently large sparsity parameter. We prove this lemma in Section 4.

Lemma 2.5 (Matrices contained in a random s-LDPC code)**.** *For any $\delta, \varepsilon > 0$, prime power q, and $\ell \geq 1$, there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$, and sufficiently large n. Suppose that $\tau \in \mathcal{D}_{n,\ell}$ is a δ-smooth and $(1 - R - \varepsilon)$-implicitly rare distribution. Let $C \subseteq \mathbb{F}_q^n$ be a random s-LDPC code. Then*

$$\Pr\left[\exists M \in \mathcal{M}_{n,\tau}, M \subset C\right] \leq q^{-\varepsilon n/2}.$$

If we ignore the requirement of smoothness, the conclusion of Lemma 2.5 for random LDPC codes is the same as the conclusion of Theorem 2.2. In particular, the two together imply that any smooth class $\mathcal{M}_{n,\tau}$ which is unlikely to appear in a random linear code is also unlikely to appear in a random LDPC code.

The proof of Lemma 2.5 proceeds by Fourier analysis. The basic idea is as follows: since C is a random s-LDPC code, each parity-check corresponds (essentially) to an independent and uniformly random set of s coordinates in $[n]$.[10] Thus, the probability that a matrix $M \in \mathcal{M}_{n,\tau}$ is in C can be

[10]This is not exactly true because the parity checks that belong to the same layer are not independent; however, we show that this does not make a large difference to the argument.

derived from the probability that s random vectors $v_1, \ldots, v_s \sim \tau$ sum to zero. This probability is given by a convolution $\tau^{*s}(0) = \tau * \tau * \cdots * \tau(0)$ of τ with itself s times. The convolution is in turn controlled by s'th powers of the Fourier coefficients $\hat{\tau}(w)$ of τ. As we will see, the condition that τ be δ-smooth implies that the nonzero Fourier coefficients $\hat{\tau}(w)$ are bounded away from 1, and this means that if s is large enough, the contributions $\hat{\tau}(w)^s$ of the nonzero coefficients to $\tau^{*s}(0)$ will become small. Then we can take a union bound over all matrices $M \in \mathcal{M}_{n,\tau}$.

2.4 Distance of random s-LDPC codes

As noted above, the first two building blocks show that any δ-smooth class $\mathcal{M}_{n,\tau}$ behaves the same for random linear matrices as it does for random LDPC matrices. The third and final building block shows that we may restrict our attention to δ-smooth matrices.

As noted in Remark 2.4, the condition that M be δ-smooth is the same as the condition that the code generated by M has relative distance at least δ. Thus, if $C \subset \mathbb{F}_q^n$ has relative distance δ, it only contains δ-smooth matrices. Fortunately, it is well-known that *binary* random s-LDPC codes have good distance, and that in fact the distance approaches the *Gilbert-Varshamov* (GV) bound with high probability.[11] Theorem 2.6 generalizes this result to s-LDPC codes over any alphabet. Below we recall that $h_q(x)$ (as in (1)) is the q-ary entropy function.

Theorem 2.6 (Random LDPC codes achieve the GV bound)**.** *For any $\delta \in (0, 1-1/q)$, $\varepsilon > 0$, and prime power q there exists $s_0 \geq 1$ such that the following holds for any $s \geq s_0$. Let $R \leq 1 - h_q(\delta) - \varepsilon$. Then a random s-LDPC code of rate R over $\mathbb{F}_q$ has relative distance at least δ with high probability.*

Remark 2.7 (Comparison to Gallager's proof)**.** *Gallager's proof for binary random s-LDPC codes in [Gal62] uses generating functions. We give an alternative proof using ideas from exponential families, which follows the approach of recent work by Linial and the first author [LMar]. Our proof extends to random s-LDPC codes over any alphabet. We note that Gallager left it as an open problem in [Gal62] to obtain a result like this for larger alphabets, but his definition was slightly different than ours: the coefficients $\alpha_{i,j}$ in his parity checks were all ones, while ours are taken randomly from $\mathbb{F}_q^*$.*

Despite having different frameworks, our proof and that of [Gal62] turn out to yield similar equations. In particular our proof of Lemma 5.2 is very similar to the corresponding proof in [Gal62] at a technical level. We highlight where the proofs diverge in Remark 5.9.

2.5 Proof of Theorem 1.7 from Theorem 2.2, Lemma 2.5 and Theorem 2.6

Our main theorem (Theorem 1.7) now follows as an immediate consequence of the building blocks above. We restate Theorem 1.7 here:

Theorem (Main theorem (Theorem 1.7), restated)**.** *For any $R \in (0,1)$, $\varepsilon > 0$, prime power q, and $b \geq 1$ there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$. Let $\mathcal{P}$ be a b-local property, and suppose that a random linear code of rate R over $\mathbb{F}_q$ satisfies $\mathcal{P}$ with high probability. Then a random s-LDPC code of rate $R - \varepsilon$ over $\mathbb{F}_q$ satisfies $\mathcal{P}$ with high probability.*

Proof. Let $C := \{C_n\}_{n \in \mathbb{N}}$ be a random s-LDPC code of rate $R - \varepsilon$ over $\mathbb{F}_q$, where s is a sufficiently large odd integer (depending on R, ε, q, and b). Let $\mathcal{P} = \{P_n\}_{n \in \mathbb{N}}$, and for $n \in \mathbb{N}$ let $\mathcal{B}_n$ be the

[11] The GV bound refers to the rate-distance trade-off $R = 1 - h_q(\delta)$, which is approached by a random linear code.

collection of subsets of $\mathbb{F}_q^n$ defining P_n. By the linearity of C, we may assume that each set $B \in \mathcal{B}_n$ contains linearly independent elements. Let

$$\delta := h_q^{-1}(1-R) > 0.$$

For $n \in \mathbb{N}$, let $\mathcal{M}_n$ be the collection of matrices corresponding to sets $B \in \mathcal{B}_n$, and let $\mathcal{D}_n$ be the collection of row distributions of matrices $M \in \mathcal{M}_n$ that are δ-smooth, and let $\mathcal{E}_n$ be the collection of row distributions of matrices $M \in \mathcal{M}_n$ that are not δ-smooth. In this notation, our goal is to show that with probability tending to 1, the code C_n does not contain any matrix $M \in \mathcal{M}_{n,\tau}$ for $\tau \in \mathcal{D}_n \cup \mathcal{E}_n$.

Since the rate of C is $R-\varepsilon$, we have $\delta = h_q^{-1}(1-R(C)-\varepsilon)$. Thus, Theorem 2.6 and Remark 2.4 imply that with probability $1-o(1)$, C does not contain any matrices M that are not δ-smooth, and in particular does not contain $M \in \mathcal{M}_{n,\tau}$ for any $\tau \in \mathcal{E}_n$.

Next, fix $\tau \in \mathcal{D}_n$. Assuming that n is sufficiently large, Theorem 2.2, together with our assumption that a random linear code of rate R over $\mathbb{F}_q$ satisfies $\mathcal{P}$ with high probability, imply that τ is γ-implicitly rare for $\gamma = 1-R+\varepsilon/2 = 1-(R-\varepsilon)-\varepsilon/2$. Lemma 2.5 implies in turn that C_n contains some matrix $M \in \mathcal{M}_{n,\tau}$ with probability at most $q^{-\varepsilon n/4}$. Finally, noting that $|\mathcal{D}_n| \le |\bigcup_{\ell=1}^{b} \mathcal{D}_{n,\ell}| \le b \cdot \binom{n+q^b-1}{q^b-1}$ by (2), we conclude that C_n contains some matrix $M \in \mathcal{M}_{n,\tau}$ for $\tau \in \mathcal{D}_n$ with probability at most

$$|\mathcal{D}_n| \cdot q^{-\varepsilon n/4} \le b \cdot \binom{n+q^b-1}{q^b-1} \cdot q^{-\varepsilon n/4},$$

which goes to 0 as $n \to \infty$. Thus, by a union bound the probability that c_n contains any matrix $M \in \mathcal{M}_{n,\tau}$ for $\tau \in \mathcal{D}_n \cup \mathcal{E}_n$ also tends to 0 as $n \to \infty$, and this completes the proof. □

3 Characterization of matrices contained in a random linear code: proof of Theorem 2.2

In this section we prove Theorem 2.2, which we restate below.

Theorem (Characterization of matrices contained in random linear code (Theorem 2.2, restated))**.** *Suppose that $\tau \in \mathcal{D}_{n,\ell}$ is a $(1-R-\varepsilon)$-implicitly rare distribution. Let $C \subseteq \mathbb{F}_q^n$ be a random linear code of rate R. Then*

$$\Pr\left[\exists M \in \mathcal{M}_{n\tau}, M \subset C\right] \le q^{-\varepsilon n}.$$

Conversely, suppose that $\tau \in \mathcal{D}_{n,\ell}$ is not $(1-R+\varepsilon)$-implicitly rare. Let $C \subseteq \mathbb{F}_q^n$ be a random linear code of rate R. Then

$$\Pr\left[\exists M \in \mathcal{M}_{n\tau}, M \subset C\right] \ge 1 - \binom{n+q^{2\ell}-1}{q^{2\ell}-1}^3 \cdot q^{-\varepsilon n}.$$

The proof uses the following standard facts.

Fact 3.1. *A matrix $M \in \mathbb{F}_q^{n\times\ell}$ is contained in a random linear code $C \subseteq \mathbb{F}_q^n$ of rate R with probability $q^{-(1-R)\cdot\mathrm{rank}(M)\cdot n}$.*

We include the proof of Fact 3.1 for completeness.

Proof. Let $v_1, \ldots, v_{\mathrm{rank}(M)}$ be columns of M that form a basis for the column span of M. Then for each v_i, $\Pr[v_i \in C] = q^{-(1-R)n}$. Since the v_i are linearly independent, the events that they are contained in a random linear code C are stochastically independent, and so the probability that all $\mathrm{rank}(M)$ of these vectors are contained in C is $q^{-(1-R)\cdot\mathrm{rank}(M)\cdot n}$. □

Fact 3.2 ([CS+04], Lemma 2.2)**.** *For any distribution $\tau \in \mathcal{D}_{n,\ell}$,*

$$2^{H(\tau)\cdot n} \cdot \binom{n+q^\ell-1}{q^\ell-1}^{-1} \leq |\mathcal{M}_{n,\tau}| \leq 2^{H(\tau)\cdot n}.$$

3.1 Implicitly rare distributions are unlikely to appear in a random linear code

We begin by proving the first part of the theorem, that $\Pr[\exists M \in \mathcal{M}_{n\tau}, M \subset C] \leq q^{-\varepsilon n}$ when τ is $(1-R-\varepsilon)$-implicitly rare.

Let $\tau \in \mathcal{D}_{n,\ell}$ be $(1-R-\varepsilon)$-implicitly rare. Let $A \in \mathbb{F}_q^{m\times\ell}$ be such that the distribution τ' given by Av for $v \sim \tau$ is $(1-R-\varepsilon)$-rare. By Fact 3.1, a matrix $M' \in \mathcal{M}_{n,\tau'}$ is contained in C with probability $q^{-(1-R)\cdot\mathrm{rank}(M')\cdot n} = q^{-(1-R)\cdot d(\tau')\cdot n}$, and so

$$\Pr\left[\exists M \in \mathcal{M}_{n,\tau'}, M \subset C\right] \leq |\mathcal{M}_{n,\tau'}| \cdot q^{-(1-R)\cdot d(\tau')\cdot n} \leq q^{(H_q(\tau')-(1-R)\cdot d(\tau'))\cdot n} \leq q^{-\varepsilon n},$$

where the first inequality follows by Fact 3.2, and the second inequality follows since τ' is $(1-R-\varepsilon)$-rare. Finally, note that if C contains some matrix $M \in \mathcal{M}_{n,\tau}$, then by linearity, $M' := MA^T \in \mathcal{M}_{n,\tau'}$ is also contained in C. So we conclude

$$\Pr[\exists M \in \mathcal{M}_{n,\tau}, M \subset C] \leq q^{-\varepsilon n}.$$

3.2 Distributions that are not implicitly rare are likely to appear in a random linear code

We now proceed to the second part of the theorem, which is more involved. Suppose that $\tau \in \mathcal{D}_{n,\ell}$ is not $1-R+\varepsilon$ implicitly rare.

First, by the linearity of C, we may assume that $d(\tau) = \ell$. Indeed, by the definition of $d(\tau)$, there is some matrix $B \in \mathbb{F}_q^{d(\tau)\times\ell}$ so that the distribution τ' given by $Bv, v \sim \tau$ has $d(\tau') = d(\tau)$. If τ is not $(1-R+\varepsilon)$-implicitly rare, then neither is τ' (because if the distribution given by $Aw, w \sim \tau'$ was rare for some $A \in \mathbb{F}_q^{m\times d(\tau)}$, then the distribution given by $ABv, v \sim \tau$ is also rare). Moreover, by linearity of C we have that $M \subseteq C$ if and only if $M \cdot B^T \subseteq C$. Thus, noting that $\binom{n+q^\ell-1}{q^\ell-1}$ is increasing in ℓ, we may as well work with the distribution τ' on $\mathbb{F}_q^{d(\tau')}$ and set $\ell = d(\tau')$. Thus, in the following we will assume that $d(\tau) = \ell$.

For a matrix $M \in \mathbb{F}_q^{n\times\ell}$, let X_M be the indicator variable for the event that $M \subseteq C$, and let $X = \sum_{M\in\mathcal{M}_{n,\tau}} X_M$. Our goal then is to show that $X > 0$ with high probability, and we do so by showing that $\mathrm{Var}(X) = o(\mathbb{E}^2(x))$.

We first show a lower bound on $\mathbb{E}[X]$. By Facts 3.1 and 3.2,

$$\mathbb{E}[X] = |\mathcal{M}_{n,\tau}| \cdot q^{-(1-R)\cdot\ell\cdot n} \geq q^{(H_q(\tau)-(1-R)\cdot\ell)\cdot n} \cdot \binom{n+q^\ell-1}{q^\ell-1}^{-1}. \tag{3}$$

Next we show an upper bound on $\mathrm{Var}\,(X)$. Given a pair of matrices $M, M' \in \mathcal{M}_{n,\tau}$, we let $(M|M')$ denote the $(n \times (2\ell))$-matrix consisting of a left $n \times \ell$ block equals to M, and a right $n \times \ell$ block equals to M'. Then in this notation we have

$$\begin{aligned}
\mathrm{Var}\,(X) &= \sum_{M,M' \in \mathcal{M}_{n,\tau}} \Big(\mathbb{E}[X_M \cdot X_{M'}] - \mathbb{E}[X_M] \cdot \mathbb{E}[X_{M'}] \Big) \\
&= \sum_{M,M' \in \mathcal{M}_{n,\tau}} \Big(\Pr\left[(M|M') \subseteq C\right] - \Pr[M \subseteq C] \cdot \Pr[M' \subseteq C] \Big) \\
&= \sum_{M,M' \in \mathcal{M}_{n,\tau}} \Big(q^{-(1-R)\cdot \mathrm{rank}(M|M')\cdot n} - q^{-2\cdot(1-R)\cdot \ell \cdot n} \Big).
\end{aligned}$$

Notice that in the above sum, terms for which $\mathrm{rank}(M|M') = 2\ell$ vanish. Let

$$\mathcal{M} := \Big\{ (M|M') \mid M, M' \in \mathcal{M}_{n,\tau} \text{ and } \mathrm{rank}(M|M') < 2\ell \Big\},$$

and

$$\mathcal{D} := \{\tau_M \mid M \in \mathcal{M}\}. \tag{4}$$

Then we have

$$\begin{aligned}
\mathrm{Var}\,(X) &\leq \sum_{M \in \mathcal{M}} q^{-(1-R)\cdot \mathrm{rank}(M)\cdot n} \\
&= \sum_{\tau' \in \mathcal{D}} \sum_{M \in \mathcal{M}_{n,\tau'}} q^{-(1-R)\cdot \mathrm{rank}(M)\cdot n} \\
&= \sum_{\tau' \in \mathcal{D}} |\mathcal{M}_{n,\tau'}| \cdot q^{-(1-R)\cdot d(\tau')\cdot n}. \\
&\leq \sum_{\tau' \in \mathcal{D}} q^{(H_q(\tau') - (1-R)\cdot d(\tau'))\cdot n}
\end{aligned}$$

where the last inequality follows by Fact 3.2. Finally, Claim 3.3 below shows that for any $\tau' \in \mathcal{D}$,

$$H_q(\tau') - (1-R)\cdot d(\tau') \leq 2(H_q(\tau) - (1-R)\cdot \ell) - \varepsilon,$$

which implies in turn that

$$\mathrm{Var}\,(X) \leq |\mathcal{D}| \cdot q^{2(H_q(\tau) - (1-R)\cdot \ell)\cdot n} \cdot q^{-\varepsilon n} \leq \binom{n + q^{2\ell} - 1}{q^{2\ell} - 1} \cdot q^{2(H_q(\tau) - (1-R)\cdot \ell)\cdot n} \cdot q^{-\varepsilon n}. \tag{5}$$

Above, we used the fact that $\mathcal{D} \subseteq \mathcal{D}_{n,2\ell}$ and applied (2). Combining (3) and (5), by Chebyshev's inequality we conclude that

$$\Pr[X = 0] \leq \frac{\mathrm{Var}\,(X)}{\mathbb{E}^2(X)} \leq \binom{n + q^{2\ell} - 1}{q^{2\ell} - 1}^3 \cdot q^{-\varepsilon n}.$$

To complete the proof, we prove Claim 3.3 which we used above.

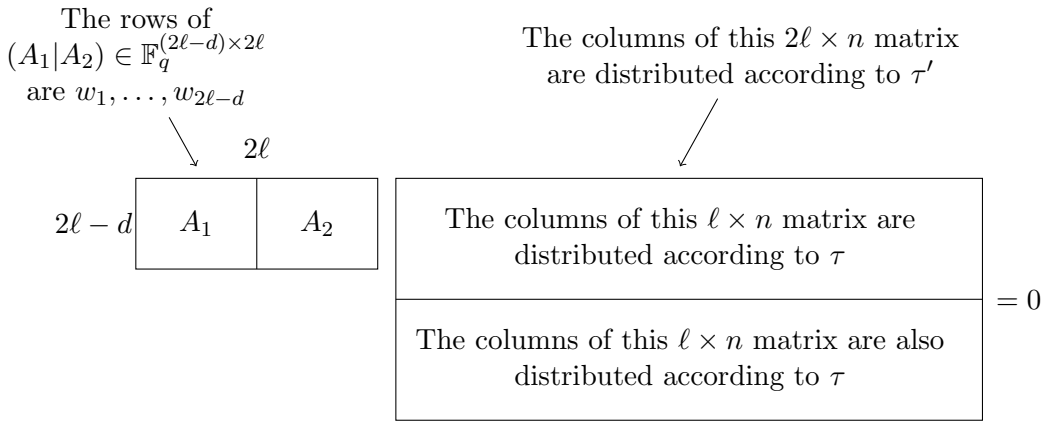

Figure 3: Notation in the proof of Claim 3.3

Claim 3.3. *Let $\mathcal{D}$ be as in (4). For any $\tau' \in \mathcal{D}$,*

$$H_q(\tau') - (1-R)\cdot d(\tau') \leq 2(H_q(\tau) - (1-R)\cdot \ell) - \varepsilon.$$

Proof. In what follows, let $d := d(\tau')$, and $V := \mathrm{span}(\mathrm{Supp}(\tau')) \subseteq \mathbb{F}_q^{2\ell}$. Let $w_1, \ldots, w_{2\ell-d} \in \mathbb{F}_q^{2\ell}$ be a basis for $V^\perp$. Let $\pi_1 : \mathbb{F}_q^{2\ell} \to \mathbb{F}_q^{\ell}$ (π_2, respectively) denote the projection of a vector $w \in \mathbb{F}_q^{2\ell}$ to the first (last, respectively) ℓ coordinates. Finally, let A be the matrix whose rows are $w_1, \ldots, w_{2\ell-d}$, and let $A_1 \in \mathbb{F}_q^{(2\ell-d)\times \ell}$ (A_2, respectively) denote the matrix whose rows are $\pi_1(w_1), \ldots, \pi_1(w_{2\ell-d})$ ($\pi_2(w_1), \ldots, \pi_2(w_{2\ell-d})$, respectively). See Figure 3 for a diagram of this notation.

We claim that all rows of A_1 are linearly independent, and so $\mathrm{rank}(A_1) = 2\ell - d$. To see this suppose in contradiction that $\pi_1(w_1), \ldots, \pi_1(w_{2\ell-d})$ are linearly dependent. Then there exists a non-trivial linear combination of $w_1, \ldots, w_{2\ell-d}$ that sums to a non-zero vector of the form $(0, w)$. But this means that $\pi_2(\mathrm{Supp}(\tau')) = \mathrm{Supp}(\tau)$ is orthogonal to w, in contradiction to our assumption that $\mathrm{Supp}(\tau) = \mathbb{F}_q^{\ell}$. Consequently, recalling that $d(\tau) = \ell$, the distribution τ'' given by $A_1 w$ for $w \sim \tau$ has $d(\tau'') = 2\ell - d$.

Let $I_q(X;Y) = I(X;Y)/\log(q)$ denote the base-q mutual information of X and Y. Now for $v \sim \tau'$ we have,

$$\begin{aligned}
H_q(\tau') &= H_q(v) \\
&= H_q(\pi_1(v)) + H_q(\pi_2(v)) - I_q(\pi_1(v); \pi_2(v)) \\
&= 2H_q(\tau) - I_q(\pi_1(v); \pi_2(v)) \\
&\leq 2H_q(\tau) - I_q(A_1\pi_1(v); -A_2\pi_2(v)) \\
&= 2H_q(\tau) - H_q(A_1\pi_1(v)) \\
&\leq 2H_q(\tau) - (1 - R + \varepsilon)\cdot d(\tau'') \\
&= 2H_q(\tau) - (1 - R + \varepsilon)\cdot (2\ell - d),
\end{aligned}$$

where the first inequality follows since $A_1\pi_1(v) + A_2\pi_2(v) = Av = 0$, and the last inequality follows by assumption that τ is not $(1 - R + \varepsilon)$-implicitly rare. Rearranging, and recalling the assumption that $2\ell > d$, gives the desired conclusion. □

4 Matrices contained in a random LDPC code: proof of Lemma 2.5

In this section we prove our second building block, Lemma 2.5, which we re-state below.

Lemma (Matrices contained in a random s-LDPC code (Lemma 2.5, restated))**.** *For any $\delta, \varepsilon > 0$, prime power q, and $\ell \geq 1$, there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$, and sufficiently large n. Suppose that $\tau \in \mathcal{D}_{n,\ell}$ is a δ-smooth and $(1-R-\varepsilon)$-implicitly rare distribution. Let $C \subseteq \mathbb{F}_q^n$ be a random s-LDPC code. Then*

$$\Pr\left[\exists M \in \mathcal{M}_{n,\tau}, M \subset C\right] \leq q^{-\varepsilon n/2}.$$

Remark 4.1 (The parity of s, again)**.** *Lemma 2.5 holds for even s as well as odd s, but the proof is slightly simpler for odd s, so we state and prove it in this case for clarity. This is the only place in the proof of Theorem 1.7 where we use the parity of s, and so this remark implies Remark 1.3.*

We begin with some definitions from Fourier analysis which we will need.

4.1 Fourier-analytic facts

We give here some basic definitions and facts from Fourier analysis of functions over $\mathbb{F}_q$. We refer the reader to, for example, [LN94, O'D14] for more details and proofs of these facts. In what follows assume that $q = p^h$ for a prime p. The trace map of $\mathbb{F}_q$ over $\mathbb{F}_p$ is the function $\mathrm{tr} : \mathbb{F}_q \to \mathbb{F}_p$ given by

$$\mathrm{tr}(\alpha) = \alpha + \alpha^p + \alpha^{p^2} + \cdots + \alpha^{p^{h-1}}.$$

For a function $f : \mathbb{F}_q^n \to \mathbb{C}$, we define the Fourier transform $\hat{f} : \mathbb{F}_q^n \to \mathbb{C}$ of f by

$$\hat{f}(y) = \mathbb{E}_{x \in \mathbb{F}_q^n}\left[f(x) \cdot \chi_x(y)\right],$$

where $\chi_x(y) = \omega_p^{\mathrm{tr}(\langle x,y\rangle)}$, and $\omega_p = e^{2\pi i/p}$. Then we have the decomposition

$$f(x) = \sum_{y \in \mathbb{F}_q^n} \hat{f}(y) \cdot \chi_y(x).$$

We define an inner product on the space of $\mathbb{C}$-valued functions on $\mathbb{F}_q^n$ by

$$\langle f, g\rangle = \mathbb{E}_{x \in \mathbb{F}_q^n}\left[f(x) \cdot \overline{g(x)}\right].$$

Plancherel's identity then asserts that

$$\langle f, g\rangle = \sum_{x \in \mathbb{F}_q^n} \hat{f}(x) \cdot \overline{\hat{g}(x)}.$$

An important special case is Parseval's identity:

$$\langle f, f\rangle = \sum_{x \in \mathbb{F}_q^n} |\hat{f}(x)|^2.$$

The convolution of a pair of functions $f, g : \mathbb{F}_q^n \to \mathbb{C}$ is given by

$$(f * g)(x) = \mathbb{E}_{y \in \mathbb{F}_q^n}\left[f(y) \cdot g(x-y)\right].$$

Convolution interacts nicely with the Fourier transform:

$$\widehat{f * g}(x) = \hat{f}(x) \cdot \hat{g}(x).$$

Finally, we define inductively $f^{*1} := f$, and $f^{*s} = f^{*(s-1)} * f$ for an integer $s \geq 2$.

4.2 Proof of Lemma 2.5

The main technical lemma in the proof of Lemma 2.5 shows that the probability that a matrix with smooth row distribution is contained in a random LDPC with sufficiently large sparsity parameter is roughly the same as in a random linear code (cf., Fact 3.1).

Lemma 4.2. *For any $\delta, \varepsilon > 0$, prime power q, and $\ell \geq 1$ there exists $s_0 \geq 1$ such that the following holds for any odd $s \geq s_0$, and sufficiently large n. Let $M \in \mathbb{F}_q^{n \times \ell}$ be a matrix, and suppose that τ_M is δ-smooth. Then M is contained in a random s-LDPC $C \subseteq \mathbb{F}_q^n$ of rate R with probability at most $q^{-(1-\varepsilon)\cdot(1-R)\cdot\ell\cdot n}$.*

We prove the above lemma in Section 4.3. Next we show how it implies Lemma 2.5.

Proof of Lemma 2.5. Let $A \in \mathbb{F}_q^{m \times \ell}$ be such that the distribution τ' given by Av for $v \sim \tau$ is $(1 - R - \varepsilon)$-rare. Without loss of generality we may assume that $\operatorname{rank}(A) = m$. Otherwise, we can replace A with a submatrix $A' \in \mathbb{F}_q^{\operatorname{rank}(A) \times \ell}$ whose rows are a basis for the row-span of A, noting that the distribution τ'' given by $A'v$ for $v \sim \tau$ satisfies that $H_q(\tau'') = H_q(\tau')$ and $d(\tau'') = d(\tau')$.

Next observe that the distribution τ' is δ-smooth: indeed, for $u \in \mathbb{F}_q^m \setminus \{0\}$ we have that

$$\Pr_{v \sim \tau}[\langle u, Av \rangle \neq 0] = \Pr_{v \sim \tau}[\langle A^T u, v \rangle \neq 0] \geq \delta,$$

where the last inequality follows since $A^T u \neq 0$ by assumption that $\operatorname{rank}(A) = m$.

By Lemma 4.2, a matrix $M' \in \mathcal{M}_{n,\tau'}$ is contained in C with probability at most $q^{-(1-\varepsilon/2)\cdot(1-R)\cdot d(\tau')\cdot n}$, and so C contains some matrix $M' \in \mathcal{M}_{n,\tau'}$ with probability at most

$$|\mathcal{M}_{n,\tau'}| \cdot q^{-(1-\varepsilon/2)\cdot(1-R)\cdot d(\tau')\cdot n} \leq q^{(H_q(\tau') - (1-\varepsilon/2)\cdot(1-R)\cdot d(\tau'))\cdot n} \leq q^{((1-R-\varepsilon) - (1-\varepsilon/2)\cdot(1-R))\cdot d(\tau')\cdot n} \leq q^{-\varepsilon n/2},$$

where the first inequality follows by Fact 3.2, and the second inequality follows since τ' is $(1-R-\varepsilon)$-rare.

Finally, note that if C contains some matrix $M \in \mathcal{M}_{n,\tau}$, then by linearity, $M' := MA^T \in \mathcal{M}_{n,\tau'}$ is also contained in C. So we conclude that C contains some matrix $M \in \mathcal{M}_{n,\tau}$ with probability at most $q^{-\varepsilon n/2}$. □

4.3 Proof of Lemma 4.2

Let $H \in \mathbb{F}_q^{((1-R)\cdot n) \times n}$ be the parity-check matrix of C with layers $H_1, H_2, \ldots, H_{(1-R)\cdot s}$, as in Section 2.1. Recall that each layer H_i is an independent sample from $FD\Pi$, where F is as in Figure 2, $\Pi \in \{0,1\}^{n \times n}$ is a random permutation matrix, and $D \in \mathbb{F}_q^{n \times n}$ is a diagonal matrix with diagonal entries that are independent and uniformly random in $\mathbb{F}_q^*$. Let Λ be a random matrix sampled according to the distribution $\mathcal{D}\Pi M$. Then by independence of the layers,

$$\begin{aligned}
\Pr[M \subseteq C] &= \Pr[HM = 0] \\
&= \big(\Pr[H_1 M = 0]\big)^{(1-R)\cdot s} \\
&= \big(\Pr[FD\Pi M = 0]\big)^{(1-R)\cdot s} \\
&= \big(\Pr[F\Lambda = 0]\big)^{(1-R)\cdot s}. \qquad (6)
\end{aligned}$$

So it suffices to bound the probability that $F\Lambda = 0$.

Next observe that each row in Λ has the same marginal distribution P given by λv for $v \sim \tau_M$ and uniform random $\lambda \in \mathbb{F}_q^*$. Let $\Lambda' \in \mathbb{F}_q^{n \times \ell}$ be a random matrix in which each row is independently sampled according to P. We claim that

$$\Pr[F\Lambda = 0] \leq O\left(n^{\frac{q^\ell - 1}{2}}\right) \cdot \Pr[F\Lambda' = 0]. \tag{7}$$

Indeed,

$$\begin{aligned}\Pr[F\Lambda = 0] &= \Pr\left[F\Lambda' = 0 \mid \tau_{\Lambda'} = \tau_M\right] \\ &= \frac{\Pr\left[F\Lambda' = 0 \wedge \tau_{\Lambda'} = \tau_M\right]}{\Pr\left[\tau_{\Lambda'} = \tau_M\right]} \\ &\leq \frac{\Pr\left[F\Lambda' = 0\right]}{\Pr\left[\tau_{\Lambda'} = \tau_M\right]}.\end{aligned}$$

Now we have

$$\Pr\left[\tau_{\Lambda'} = \tau_M\right] = \binom{n}{n \cdot \tau_M(v_1), \ldots, n \cdot \tau_M(v_{q^\ell})} \cdot \prod_{v \in \mathbb{F}_q^\ell} \tau_M(v)^{n \cdot \tau_M(v)}$$

where $v_1, \ldots, v_{q^\ell}$ are the elements of $\mathbb{F}_q^\ell$, and then (7) follows from Stirling's approximation.

Thus, it is enough to bound the probability that $F\Lambda' = 0$. We can express this probability as

$$\Pr\left[F\Lambda' = 0\right] = \left(\Pr_{u_1, \ldots, u_s \sim P}\left[\sum_{i=1}^{s} u_i = 0\right]\right)^{n/s} = \left(q^{\ell \cdot (s-1)} \cdot P^{*s}(0)\right)^{n/s}. \tag{8}$$

Next we bound $P^{*s}(0)$. In terms of Fourier transform, we can write

$$P^{*s}(0) = \sum_{y \in \mathbb{F}_q^\ell} \widehat{P^{*s}}(y) \cdot \chi_y(0) = \sum_{y \in \mathbb{F}_q^\ell} \left(\hat{P}(y)\right)^s.$$

Claim 4.4 below shows that $\hat{P}(y) \leq q^{-\ell} \cdot \left(1 - \frac{q}{q-1} \cdot \delta\right)$ for any $y \in \mathbb{F}_q^\ell \setminus \{0\}$, and by the assumption that s is odd this implies in turn that

$$P^{*s}(0) = \left(\hat{P}(0)\right)^s + \sum_{y \in \mathbb{F}_q^\ell \setminus \{0\}} \left(\hat{P}(y)\right)^s \leq q^{-\ell \cdot s} + q^{-\ell \cdot (s-1)} \cdot \left(1 - \frac{q}{q-1} \cdot \delta\right)^s. \tag{9}$$

Finally, combining Equations (6), (7), (8), and (9) we conclude that

$$\Pr[M \subseteq C] \leq O\left(n^{\frac{q^\ell - 1}{2} \cdot (1-R) \cdot s}\right) \cdot \left(q^{-\ell} + \left(1 - \frac{q}{q-1} \cdot \delta\right)^s\right)^{(1-R) \cdot n} \leq q^{-(1-\varepsilon) \cdot (1-R) \cdot \ell \cdot n},$$

where the last inequality holds for large enough s depending on $\delta, \varepsilon, q, \ell$, and sufficiently large n.

Remark 4.3 (The choice of s)**.** *An inspection of the proof shows that we may take*

$$s_0 = O\left(\frac{\ell}{\log_q\left(\frac{1}{1 - \delta/(1-1/q)}\right)}\right).$$

In particular, noting that $\ell \le b$ and that

$$\log_q\left(\frac{1}{1-\delta/(1-1/q)}\right) = \frac{1}{\ln(q)}\sum_{i=1}^{\infty}\frac{1}{i}\left(\frac{\delta}{1-1/q}\right)^i,$$

this part of the proof requires us to take

$$s_0 \ge C_0 \cdot \frac{b\log(q)}{\delta}$$

for some constant $C_0 > 0$. There is one other place in the proof of Theorem 1.7 that requires s_0 to be sufficiently large; we comment on this in Remark 5.3.

Now, all that remains is to prove Claim 4.4 which we used above.

Claim 4.4. *For any $y \in \mathbb{F}_q^\ell \setminus \{0\}$,*

$$\hat{P}(y) \le q^{-\ell}\cdot\left(1 - \frac{q}{q-1}\cdot\delta\right).$$

Proof of Claim 4.4. We have

$$\begin{aligned}
\hat{P}(y) &= q^{-\ell}\cdot\sum_{x\in\mathbb{F}_q^\ell} P(x)\cdot\omega_p^{\mathrm{tr}(\langle y,x\rangle)}\\
&= q^{-\ell}\cdot\mathbb{E}_{v\sim\tau_M}\mathbb{E}_{\lambda\in\mathbb{F}_q^*}\left[\omega_p^{\mathrm{tr}(\langle y,\lambda v\rangle)}\right]\\
&= q^{-\ell}\cdot\left(\Pr_{v\sim\tau_M}[\langle v,y\rangle\neq 0]\cdot\mathbb{E}_{\xi\in\mathbb{F}_q^*}\left[\omega_p^{\mathrm{tr}(\xi)}\right] + \Pr_{v\sim\tau_M}[\langle v,y\rangle = 0]\cdot\mathbb{E}_{\lambda\in\mathbb{F}_q^*}\left[\omega_p^{\mathrm{tr}(0)}\right]\right)\\
&= q^{-\ell}\cdot\left(\Pr_{v\sim\tau_M}[\langle v,y\rangle\neq 0]\cdot\frac{-1}{q-1} + \Pr_{v\sim\tau_M}[\langle v,y\rangle = 0]\cdot 1\right)\\
&\le q^{-\ell}\cdot\left(\frac{-\delta}{q-1} + (1-\delta)\right) = q^{-\ell}\cdot\left(1-\frac{q}{q-1}\cdot\delta\right),
\end{aligned}$$

where the last inequality follows by assumption that τ_M is δ-smooth. □

This completes the proof of Lemma 4.2.

5 Random LDPC codes achieve the GV bound: proof of Theorem 2.6

In this section we prove Theorem 2.6, which shows that an LDPC code over any alphabet approaches the Gilbert-Varshamov bound with high probability. We restate the theorem below.

Theorem (Random LDPC codes achieve the GV bound (Theorem 2.6, restated))**.** *For any $\delta \in (0, 1-1/q)$, $\varepsilon > 0$, and prime power q there exists $s_0 \ge 1$ such that the following holds for any $s \ge s_0$. A random s-LDPC code of rate $R \ge 1 - h_q(\delta) - \varepsilon$ over $\mathbb{F}_q$ has relative distance at least δ with high probability.*

5.1 Proof of Theorem 2.6, given a lemma

In this section we give an outline of the proof of Theorem 2.6 and prove the theorem based on Lemma 5.2 that we state below and prove in subsequent subsections.

Our goal is to show that a random s-LDPC code C has good distance, or equivalently that there are no low-weight codewords in C with high probability. To that end, we introduce the following notation.

Definition 5.1. *For $\lambda \in (0,1)$ such that λn is an integer, let $P_\lambda = \Pr[u \in C]$, where $u \in \mathbb{F}_q^n$ has weight λ. Note that this probability is the same for every u of weight λ, so P_λ is well-defined.*

Our main challenge is to find sufficiently tight upper bounds on these terms P_λ for $0 < \lambda \leq \delta$. The proof proceeds by giving a bound on P_λ in terms of a certain function $\varphi : (0, \frac{q-1}{q}] \to \mathbb{R}_{\leq 0}$. We will prove the following lemma below in Sections 5.2 and 5.3. We will define φ below in Section 5.2, but for now we introduce its important properties in the following lemma (which we also prove below).

Lemma 5.2. *There is a function $\varphi : \left(0, \frac{q-1}{q}\right] \to \mathbb{R}_{\leq 0}$ which has the following properties.*

1. *For every $\lambda \in \left(0, \frac{q-1}{q}\right]$,*
$$\log_q P_\lambda \leq \varphi(\lambda)(1-R)n.$$
2. *The function φ satisfies*
$$\varphi(\lambda) \leq \log_q \left(1 + (q-1)\left(1 - \frac{q}{q-1}\lambda\right)^s\right) - 1$$
for all $\lambda \in (0, \frac{q-1}{q}]$.
3. *The function $\frac{\varphi(\lambda)}{h_q(\lambda)}$ is strictly increasing in the range $0 < \lambda \leq \frac{q-1}{q}$.*

Before we prove Lemma 5.2, we show how it implies Theorem 2.6.

Proof of Theorem 2.6. Our goal is to show that if C is a random s-LDPC code as in the statement of Theorem 2.6, then with high probability there are no codewords in C of relative weight less than δ. In the following, we assume without loss of generality that δn is an integer. Now

$$\begin{aligned}
\Pr[C \text{ has relative distance less than } \delta] &\leq \sum_{i=1}^{\delta n} P_{\frac{i}{n}} \left| \left\{ u \in \mathbb{F}_q^n \mid \mathrm{wt}(u) = \frac{i}{n} \right\} \right| && (10)\\
&\leq \sum_{i=1}^{\delta n} P_{\frac{i}{n}} q^{nh_q(\frac{i}{n})} \\
&\leq \sum_{i=1}^{\delta n} q^{(\varphi(\frac{i}{n})(1-R)+h_q(\frac{i}{n}))n} && (11)\\
&= \sum_{i=1}^{\delta n} q^{nh_q(\frac{i}{n})\left(\frac{(1-R)\varphi(\frac{i}{n})}{h_q(\frac{i}{n})}+1\right)} && (12)\\
&\leq \sum_{i=1}^{\delta n} q^{nh_q(\frac{i}{n})\left(\frac{(1-R)\varphi(\delta)}{h_q(\delta)}+1\right)}. && (13)
\end{aligned}$$

Above, (10) follows from the union bound, (11) from Item 1 of Lemma 5.2, and (13) from Item 3 of Lemma 5.2. By Item 2 of Lemma 5.2,

$$\frac{(1-R)\varphi(\delta)}{h_q(\delta)}+1=\frac{(1-R)\cdot\left(\log_q\left(1+(q-1)\left(1-\frac{q}{q-1}\delta\right)^s\right)-1\right)}{h_q(\delta)}+1.$$

Recall our hypothesis that the rate of the code satisfies $R\le 1-h_q(\delta)-\varepsilon$, and so $1-R\ge h_q(\delta)+\varepsilon$. Noting that $\log_q\left(1+(q-1)\left(1-\frac{q}{q-1}\delta\right)^s\right)-1\le 0$, we may thus bound the right hand side from above by

$$\begin{aligned}
&\frac{(h_q(\delta)+\varepsilon)\cdot\left(\log_q\left(1+(q-1)\left(1-\frac{q}{q-1}\delta\right)^s\right)-1\right)}{h_q(\delta)}+1\\
&=\left(1+\frac{\varepsilon}{h_q(\delta)}\right)\cdot\left(\log_q\left(1+(q-1)\left(1-\frac{q}{q-1}\delta\right)^s\right)-1\right)+1\\
&=\left(1+\frac{\varepsilon}{h_q(\delta)}\right)\cdot\log_q\left(1+(q-1)\left(1-\frac{q}{q-1}\delta\right)^s\right)-\frac{\varepsilon}{h_q(\delta)}\\
&\le\left(1+\frac{\varepsilon}{h_q(\delta)}\right)\frac{(q-1)}{\ln(q)}\left(1-\frac{q\delta}{q-1}\right)^s-\frac{\varepsilon}{h_q(\delta)}.
\end{aligned}$$

Thus, as long as s is sufficiently large in terms of δ,ε and q, we conclude that

$$\frac{(1-R)\varphi(\delta)}{h_q(\delta)}+1\le-\frac{\varepsilon}{2h_q(\delta)}\le-\frac{\varepsilon}{2}.$$

Hence, the right-hand side of (13) is upper bounded by

$$\sum_{i=1}^{\delta n}q^{-\frac{nh_q(\frac{i}{n})\varepsilon}{2}}.$$

This sum is dominated by its first term, so it is at most $O(n^{-\Omega(1)})$.

□

Remark 5.3 (The choice of s)**.** *An inspection of the proof above shows that it suffices to take $s\gtrsim\ln(q/\varepsilon)/\delta$. Thus, this part of the proof requires that $s_0\gtrsim\ln(q/\varepsilon)/\delta$.*

Remark 5.4 (Polynomially small failure probability)**.** *In the proof, we see that the failure probability, while $o(1)$, is only polynomially small in n. In fact, this is tight: it is not hard to see that an s-random LDPC code C (for $s=O(1)$) contains a codeword of weight 2 with probability $n^{-O(1)}$.*

5.2 The function φ and proof of Lemma 5.2, Items 1 and 2

Let $\lambda\in\left(0,\frac{q-1}{q}\right]$ such that λn is an integer, and let $u\in\mathbb{F}_q^n$ have weight λn. Let $H_1,\dots,H_t$ be the layers of the the parity-check matrix H of C, as in Figure 2. Note that the matrices $H_1,\dots,H_t$ are identically and independently distributed. In particular, the events $\Pr(H_iu=0)$ are independent. Hence,

$$P_\lambda=\Pr[u\in C]=\Pr[Hu=0]=\Pr[H_1u=0]^t. \tag{14}$$

Since the distribution of H_1 is invariant to permutation of coordinates, this last probability does not depend on the vector u as long as it is of relative weight λ. Hence,

$$\Pr[H_1 u = 0] = \Pr[H_1 \bar{u} = 0] = \Pr[F\bar{u} = 0],$$

where $\bar{u}$ is uniformly sampled from the set of all vectors of weight λ in $\mathbb{F}_q^n$. Therefore,

$$P_\lambda = \Pr[F\bar{u} = 0]^t,$$

where F is as in Figure 2.

We turn to bound this expression. Let $\beta \in \left(0, \frac{q-1}{q}\right]$. Denote by $\mu_q(\beta)$ the distribution on $\mathbb{F}_q$ which is 0 with probability $1-\beta$ and uniform on $\mathbb{F}_q^*$ with probability β. When β is clear from context, we shorthand $\mu_q = \mu_q(\beta)$. Let $v \in \mathbb{F}_q^n$ be a random vector whose entries are i.i.d. random variables sampled according to μ_q, which we denote by $v \sim \mu_q^n$. Observe that the distribution of v, conditioned on $\mathrm{wt}(v) = \lambda$, is identical to the distribution of $\bar{u}$. Hence, by Bayes' rule,

$$\Pr[F\bar{u} = 0] = \Pr[Fv = 0 \mid \mathrm{wt}(v) = \lambda] = \Pr[\mathrm{wt}(v) = \lambda \mid Fv = 0] \cdot \frac{\Pr[Fv = 0]}{\Pr[\mathrm{wt}(v) = \lambda]} \leq \frac{\Pr[Fv = 0]}{\Pr[\mathrm{wt}(v) = \lambda]} \tag{15}$$

where the probabilities are over the choice of $v \sim \mu_q(\beta)^n$.

We proceed to bound the right-hand side of (15). For the denominator, note that

$$\Pr[\mathrm{wt}(v) = \lambda] = \binom{n}{\lambda n} \beta^{\lambda n} (1-\beta)^{(1-\lambda)n} \geq q^{-D_{\mathrm{KL}_q}(\lambda \| \beta) n} \tag{16}$$

where above $D_{\mathrm{KL}_q}(x \parallel y)$ denotes the KL Divergence,

$$D_{\mathrm{KL}_q}(x \parallel y) = -x \log_q \frac{y}{x} - (1-x) \log_q \frac{1-y}{1-x} \text{ for } x \in [0,1] \text{ and } y \in (0,1).$$

We next focus on the numerator. The following notation will be useful:

Definition 5.5. *For $k \in \mathbb{N}$, let*

$$\mathbb{V}_q^k = \left\{ w \in \mathbb{F}_q^k : \sum_{i=1}^k w_i = 0 \right\}.$$

Let $f_1, \ldots, f_{\frac{n}{s}}$ denote the rows of the matrix F. Note that the vectors $f_1, \ldots f_{\frac{n}{s}}$ have disjoint supports, so the products $f_i v$ are independently and identically distributed. Hence, $\Pr[Fv = 0] = \Pr[f_1 v = 0]^{\frac{n}{s}}$. Observe that the distribution of v is symmetric to multiplication of each entry by a nonzero element of $\mathbb{F}_q$. Consequently,

$$\Pr_{v \sim \mu_q^n}[Fv = 0] = \Pr_{v \sim \mu_q^n}[f_1 v = 0]^{\frac{n}{s}} = \Pr_{v \sim \mu_q^n}\left[\sum_{i=1}^s v_i = 0\right]^{\frac{n}{s}} = \left(\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s]\right)^{n/s}. \tag{17}$$

The following lemma gives a closed form for this last expression.

Lemma 5.6.

$$\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s] = \frac{1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^s}{q}.$$

Proof. We proceed by induction. The base case ($s = 0$) is immediate. Now suppose that the statement holds for $s-1$ and let $\pi : \mathbb{F}_q^s \to \mathbb{F}_q^{s-1}$ denote the projection onto the first $s-1$ coordinates. Then

$$\begin{aligned}
\Pr_{w\sim\mu_q^s}[w \in \mathbb{V}_q^s] &= \Pr_{w\sim\mu_q^s}\left[\pi(w) \in \mathbb{V}_q^{s-1}\right] \cdot \Pr_{w\sim\mu_q^s}[w_s = 0] + \Pr_{w\sim\mu_q^s}\left[\pi(w) \notin \mathbb{V}_q^{s-1}\right] \cdot \Pr_{w\sim\mu_q^s}\left[w_s = -\sum_{i=1}^{s-1} w_i \mid \pi(w) \notin \mathbb{V}_q^{s-1}\right] \\
&= \frac{1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^{s-1}}{q} \cdot (1-\beta) + \left(1 - \frac{1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^{s-1}}{q}\right) \cdot \frac{\beta}{q-1} \\
&= \frac{1}{q} + \left(1 - \frac{q\beta}{q-1}\right)^s \left(\frac{q-1}{q}\right),
\end{aligned}$$

which establishes the inductive hypothesis for s. □

Motivated by the computations above, we can define the following useful shorthands:

Definition 5.7. *For $\lambda, \beta \in (0, \frac{q-1}{q}]$, define*

$$Z(\beta) = \Pr_{w\sim\mu_q^s}\left[w \in \mathbb{V}_q^s\right] = \frac{1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^s}{q}, \tag{18}$$

$$\psi(\lambda, \beta) = sD_{\mathrm{KL}q}\left(\lambda \parallel \beta\right) + \log_q Z(\beta)$$

From Equations (14), (15), (16) and (17), we conclude that

$$\begin{aligned}
\log_q P_\lambda = t \log_q \Pr[F\bar{u} = 0] &\le tn\left(D_{\mathrm{KL}q}\left(\lambda \parallel \beta\right) + \frac{\log_q\left(1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^s\right) - 1}{s}\right) \\
&= (1-R)n\left(sD_{\mathrm{KL}q}\left(\lambda \parallel \beta\right) + \log_q\left(1 + (q-1)\left(1 - \frac{q\beta}{q-1}\right)^s\right) - 1\right) \\
&= (1-R)n\psi(\lambda, \beta)
\end{aligned} \tag{19}$$

for every $\beta \in \left(0, \frac{q-1}{q}\right]$. Above, we have used the choice $t = (1-R)s$.

This motivates the following definition:

Definition 5.8. *Let Z and ψ be as in Definition 5.7. Define:*

$$\varphi(\lambda) = \inf_{\beta \in (0, \frac{q-1}{q}]} \psi(\lambda, \beta).$$

Definition 5.8, along with (19), implies that $\log_q P_\lambda \le \varphi(\lambda)$, which establishes Item 2 of Lemma 5.2. Next we establish Item 1 of Lemma 5.2. This follows from from Def 5.8, since

$$\varphi(\lambda) \le \psi(\lambda, \lambda) = \log_q\left(1 + (q-1)\left(1 - \frac{q\lambda}{q-1}\right)^s\right) - 1,$$

using the fact that $D_{\mathrm{KL}q}\left(\lambda \parallel \lambda\right) = 0$.

This almost completes the proof of Lemma 5.2, except for Item 3, which we establish in the next section using calculus.

5.3 Proof of Item 3 of Lemma 5.2

In this section we prove Item 3, which will establish Lemma 5.2 and hence Theorem 2.6.

Remark 5.9 (Difference between [Gal62] and this proof)**.** *This is the part of the proof where the technical similarity between our proof and Gallager's breaks down. The part of [Gal62] which corresponds to our Item 3 consists of an intricate analytic argument which does not seem (to us) to generalize to larger alphabets. Thus, our proof has to rely on a different, more general, argument, which we give below.*

Before proving Item 3 of Lemma 5.2, we need to better understand the relation between a given $\lambda \in (0, \frac{q-1}{q}]$, and the β which minimizes the expression $\psi(\lambda, \beta)$.

Lemma 5.10. *Let $\lambda \in (0, \frac{q-1}{q}]$. Then, $\psi(\lambda, \beta)$ is minimized by a unique $\beta \in (0, \frac{q-1}{q}]$. This β is the only solution for*

$$\mathbb{E}_{w \sim \mu_q(\beta)} \left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s \right] = \lambda.$$

Proof. We compute the derivative.

$$\begin{aligned}
\frac{d \log_e Z(\beta)}{d\beta} &= \frac{1}{\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s]} \cdot \frac{d\left(\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s]\right)}{d\beta} \\
&= \frac{1}{\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s]} \cdot \sum_{w \in \mathbb{V}_q^s} \frac{d\left(\frac{\beta}{q-1}\right)^{s \cdot \mathrm{wt}(w)} (1-\beta)^{s \cdot (1 - \mathrm{wt}(w))}}{d\beta} \\
&= \frac{\sum_{w \in \mathbb{V}_q^s} \left(\left(\frac{\beta}{q-1}\right)^{s \cdot \mathrm{wt}(w)} (1-\beta)^{s \cdot (1-\mathrm{wt}(w))} \cdot s \cdot \left(\frac{\mathrm{wt}(w)}{\beta} - \frac{1 - \mathrm{wt}(w)}{1-\beta} \right) \right)}{\Pr_{w \sim \mu_q^s}[w \in \mathbb{V}_q^s]} \\
&= s \cdot \left(\frac{\mathbb{E}_{w \sim \mu_q^s} \left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s \right]}{\beta} - \frac{1 - \mathbb{E}_{w \sim \mu_q^s} \left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s \right]}{1-\beta} \right). \qquad (20)
\end{aligned}$$

Also, it is not hard to see that

$$\frac{\partial D_{\mathrm{KL}_q}(\lambda \parallel \beta)}{\partial \beta} = \log_q e \cdot \left(\frac{1-\lambda}{1-\beta} - \frac{\lambda}{\beta} \right).$$

Consequently,

$$\begin{aligned}
\frac{\partial \psi(\lambda, \beta)}{\partial \beta} &= s \frac{\partial D_{\mathrm{KL}_q}(\lambda \parallel \beta)}{\partial \beta} + \frac{d \log_q Z(\beta)}{d\beta} \\
&= \log_q e \cdot \left(\frac{s(1-\lambda)}{1-\beta} - \frac{s\lambda}{\beta} + \frac{d \log_e Z(\beta)}{d\beta} \right) \\
&= s \cdot \log_q e \cdot \left(\mathbb{E}_{w \sim \mu_q^s} \left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s \right] - \lambda \right) \left(\frac{1}{1-\beta} + \frac{1}{\beta} \right).
\end{aligned}$$

We conclude that $\frac{\partial \psi(\lambda, \beta)}{\partial \beta}$ has the same sign as $\mathbb{E}_{w \sim \mu_q^s} \left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s \right] - \lambda s$. The lemma now follows from the following claim:

Claim 5.11. *As β increases in the range $(0, \frac{q-1}{q}]$ the function $\mathbb{E}_{w\sim\mu_q^s}\left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s\right]$ strictly increases from 0 to $\frac{q-1}{q}$.*

Proof. Due to (18) and (20),

$$\begin{aligned}
\mathbb{E}_{w\sim\mu_q^s}\left[\mathrm{wt}(w) \mid w \in \mathbb{V}_q^s\right] &= \left(\frac{d\log_e Z(\beta)}{s\cdot d\beta} + \frac{1}{1-\beta}\right)\beta(1-\beta) \\
&= \left(\frac{\frac{dZ(\beta)}{d\beta}}{s\cdot Z(\beta)} + \frac{1}{1-\beta}\right)\beta(1-\beta) \\
&= \left(\frac{-q\left(1-\frac{q\beta}{q-1}\right)^{s-1}}{1+(q-1)\left(1-\frac{q\beta}{q-1}\right)^{s}} + \frac{1}{1-\beta}\right)\beta(1-\beta) \\
&= \beta\frac{1-\left(1-\frac{q\beta}{q-1}\right)^{s-1}\cdot(1+q\beta)}{1+(q-1)\left(1-\frac{q\beta}{q-1}\right)^{s}}, \qquad (21)
\end{aligned}$$

and the claim readily follows. □

The proof of the lemma is thus concluded. □

Lemma 5.10 and Claim 5.11 justify the following definition:

Definition 5.12. *For $\lambda \in (0, \frac{q-1}{q}]$, denote the $\beta \in (0, \frac{q-1}{q}]$ which minimizes $\psi(\lambda, \beta)$ by $\beta(\lambda)$. The inverse of this function is denoted $\lambda(\beta)$.*

By Lemma 5.10 and Equation (21),

$$\lambda(\beta) = \beta\frac{1-\left(1-\frac{q\beta}{q-1}\right)^{s-1}}{1+(q-1)\left(1-\frac{q\beta}{q-1}\right)^{s}}. \qquad (22)$$

Remark 5.13. *Unfortunately, there are good reasons to suspect that the function $\beta(\lambda)$ has no closed-form expression (see, e.g., the discussion about backward mapping in [WJ08, Sec. 3.4.2]), so we prefer to work with its inverse.*

It is convenient to extend the definition of these functions to the closed interval $[0, \frac{q-1}{q}]$ by taking limits, namely, $\lambda(0) = \beta(0) = 0$, and

$$\begin{aligned}
\varphi(0) &= \lim_{\lambda\to 0}\varphi(\lambda) = \lim_{\lambda\to 0}\psi(\lambda,\beta(\lambda)) \lim_{\beta\to 0}\psi(\lambda(\beta),\beta) = \lim_{\beta\to 0} D_{\mathrm{KL}_q}\left(\lambda(\beta)\parallel\beta\right) + \log_q Z(\beta) \\
&= \lim_{\beta\to 0} D_{\mathrm{KL}_q}\left(\lambda(\beta)\parallel\beta\right) = \lim_{\beta\to 0} -\lambda(\beta)\log_q\beta = 0.
\end{aligned}$$

We are now able to prove Item 3 of Lemma 5.2.

Proof of Lemma 5.2, Item 3. Let $\alpha(\lambda) = \frac{\varphi(\lambda)}{h_q(\lambda)}$. The claim follows immediately from the four following claims:

Claim 5.14. $\alpha(\frac{q-1}{q}) = -1$.

Claim 5.15. $\alpha(\lambda) < -1$ *for some* $\lambda \in (0, \frac{q-1}{q})$.

Claim 5.16. *There exists* $\varepsilon > 0$ *such that* $\alpha(\lambda) > -\frac{s}{2}$ *for all* $\lambda \in (0, \varepsilon)$.

Claim 5.17. *For each* $y \in (-\frac{s}{2}, -1]$, *the equation* $\alpha(\lambda) = y$ *has at most one solution* $\lambda \in (0, \frac{q-1}{q}]$.

Indeed, Claims 5.14 and 5.17 show that $\alpha(\lambda) \neq 1$ for $\lambda < \frac{q-1}{q}$. Since α is continuous, it is either upper bounded or lower bounded by -1 in the whole range $(0, \frac{q-1}{q}]$. Claim 5.15 implies the former. By Claim 5.17, if $-\frac{s}{2} < \alpha(\lambda_0) < -1$ for some $\lambda_0 \in (0, \frac{q-1}{q})$, then α must be strictly increasing in the range $[\lambda_0, \frac{q-1}{q}]$. The lemma now follows from Claim 5.16. We proceed to prove these claims.

Proof of Claim 5.14. Note that $\alpha(\frac{q-1}{q}) = \varphi(\frac{q-1}{q})$. Due to Item 2,

$$\varphi\left(\frac{q-1}{q}\right) \leq -1.$$

In the reverse direction,

$$\begin{aligned}\varphi(\lambda) = \min_{\beta} \psi(\lambda, \beta) &= \min_{\beta} \left(s \cdot D_{\mathrm{KL}_q}(\lambda \parallel \beta) + \log_q Z(\beta)\right) \\ &\geq \min_{\beta} \left(s \cdot D_{\mathrm{KL}_q}(\lambda \parallel \beta)\right) - 1 \geq -1\end{aligned}$$

for all λ. The first inequality above holds since $Z(\beta) \geq \frac{1}{q}$, due to (18) . □

Proof of Claim 5.15. By Item 1,

$$\alpha(\lambda) \leq \frac{\log_q\left(1 + (q-1)\left(1 - \frac{q}{q-1}\lambda\right)^s\right) - 1}{h_q(\lambda)}. \tag{23}$$

Let $\lambda = \frac{q-1}{q} - \varepsilon$. As ε tends from above to 0, the numerator of (23)'s right-hand side is $-1 + \Theta(\varepsilon^s)$, while the denominator is $1 - \Theta(\varepsilon^2)$. Thus, for ε small enough, (23) yields $\alpha(\lambda) < -1$. □

Proof of Claim 5.16. Let

$$\bar{Z}(\beta) = \Pr_{w \sim \mathcal{B}_q^s}\left(w \in \mathbb{V} \wedge \mathrm{wt}(w) \leq \frac{2}{s}\right) = (1-\beta)^s + \binom{s}{2}(1-\beta)^{s-2}\beta^2$$

and

$$\bar{\psi}(\beta, \lambda) = sD_{\mathrm{KL}_q}(\lambda \parallel \beta) + \log_q \bar{Z}(\beta).$$

Clearly, $\bar{\psi}(\beta, \lambda)$ is a lower bound on $\psi(\beta, \lambda)$, so

$$\varphi(\lambda) \geq \min_{\beta \in (0, \frac{q-1}{q}]} \bar{\psi}(\lambda, \beta).$$

Note that

$$\frac{\partial \bar{\psi}(\lambda, \beta)}{\partial \beta} = \frac{s}{\beta(1-\beta)}\left(\frac{2(s-1)}{\left(\frac{1-\beta}{\beta}\right)^2 + \binom{s}{2}} - \lambda\right),$$

Hence, for $\lambda < \frac{2}{s}$, the minimum of $\bar{\psi}(\lambda, \beta)$ is attained at $\beta_0 = \frac{y}{1+y}$, where

$$y = \left(\frac{\lambda}{2(s-1) - \binom{s}{2}\lambda} \right)^{\frac{1}{2}}.$$

Therefore,

$$\alpha(\lambda) = \frac{\varphi(\lambda)}{h_q(\lambda)} \geq \frac{\bar{\psi}(\lambda, \beta_0)}{h_q(\lambda)} = \frac{s}{2} \left(-1 + \frac{\lambda \left(\log_q \left(2(s-1) - \binom{s}{2}\lambda \right) - \log_q(1 - \lambda s) \right) + (1-\lambda)\log_q(1-\lambda)}{h_q(\lambda)} \right).$$

For λ small enough, the right-hand side is clearly larger than $-\frac{s}{2}$. □

Proof of Claim 5.17. Denote $\beta^* = \beta(\lambda)$. Let $y \in (-\frac{s}{2}, -1]$, and define the function $\varphi_y(\lambda) = \varphi(\lambda) - y h_q(\lambda)$. We seek to show that $\varphi_y(\lambda)$ has at most one root in the range $(0, \frac{q-1}{q}]$. This is a consequence of the following three statements, proven below:

1. $\frac{d\varphi_y(\lambda)}{d\lambda}$ has at most one extremal point in the open interval $(0, \frac{q-1}{q})$.
2. $\frac{d\varphi_y(\lambda)}{d\lambda}(\frac{q-1}{q}) = 0$.
3. $\varphi_y(0) = 0$.

Indeed, the first statement implies that $\frac{d\varphi_y(\lambda)}{d\lambda}$ has at most two roots in the interval $(0, \frac{q-1}{q}]$. The second statement says that one of these roots is at $\frac{q-1}{q}$, so $\frac{d\varphi_y(\lambda)}{d\lambda}$ has at most one root in $(0, \frac{q-1}{q})$. Consequently $\varphi_y(\lambda)$ has at most one extremal point and two roots in $[0, \frac{q-1}{q}]$. Due to the third statement, one of these roots is 0, so there can only be one root in $(0, \frac{q-1}{q}]$. We turn to prove these statements.

Statement 3 is trivial. For Statement 2, note that in the derivative

$$\frac{d\varphi(\lambda)}{d\lambda} = \frac{\partial\psi(\lambda,\beta)}{\partial\beta}_{\beta=\beta^*} \cdot \frac{d\beta^*}{d\lambda} + \frac{\partial\psi(\lambda,\beta)}{\partial\lambda}_{\beta=\beta^*},$$

the first term vanishes since ψ has a minimum at (λ, β^*). Hence,

$$\frac{d\varphi(\lambda)}{d\lambda} = \frac{\partial\psi(\lambda,\beta)}{\partial\lambda}_{\beta=\beta^*} = s\frac{\partial D_{\mathrm{KL}q}(\lambda \parallel \beta)}{\partial\lambda}_{\beta=\beta^*} = s\log_q \frac{\lambda(1-\beta^*)}{(1-\lambda)\beta^*}.$$

In particular, $\beta(\frac{q-1}{q}) = \frac{q-1}{q}$, so

$$\frac{d\varphi_y(\lambda)}{d\lambda}_{\lambda=\frac{q-1}{q}} = \frac{d\varphi(\lambda)}{d\lambda}_{\lambda=\frac{q-1}{q}} - y\frac{dh_q(\lambda)}{d\lambda}_{\lambda=\frac{q-1}{q}} = 0,$$

since, in the last transition, the two terms vanish.

We turn to Statement 1. Define the new variable $x = 1 - \frac{q\beta^*}{q-1}$. Note the following useful relations, the second of which follows from Equation (22):

$$\beta^* = \frac{q-1}{q}(1-x) \tag{24}$$

and

$$\frac{\lambda}{1-\lambda} = \frac{\beta^*}{1-\beta^*} \cdot \frac{1-x^{s-1}}{1+(q-1)x^{s-1}}. \tag{25}$$

By (24) and (25),

$$\begin{aligned}
\frac{d\varphi_y(\lambda)}{d\lambda} &= s\frac{\partial D_{\mathrm{KL}_q}(\lambda \parallel \beta)}{\partial \lambda}_{|\beta=\beta^*} - y\frac{dh_q(\lambda)}{d\lambda} \\
&= s\log_q \frac{\lambda(1-\beta^*)}{(1-\lambda)\beta^*} + y\log_q \frac{\lambda}{1-\lambda} \\
&= s\log_q \frac{1-\beta^*}{\beta^*} + (s+y)\log_q \frac{\lambda}{1-\lambda} \\
&= -y\log_q \frac{1+(q-1)x}{(q-1)(1-x)} + (s+y)\log_q \frac{1-x^{s-1}}{1+(q-1)x^{s-1}}.
\end{aligned}$$

Now,

$$\frac{d^2\varphi_y(\lambda)}{dxd\lambda} \cdot \ln q = \frac{-yq}{(1+(q-1)x)(1-x)} - \frac{(s+y)(s-1)qx^{s-2}}{\left(1-x^{s-1}\right)\left(1+(q-1)x^{s-1}\right)}.$$

This second derivative vanishes when

$$\frac{-(s+y)}{y} = \frac{\left(1-x^{s-1}\right)\left(1+(q-1)x^{s-1}\right)}{(s-1)(1+(q-1)x)(1-x)x^{s-2}}.$$

Equivalently,

$$\frac{-(s+y)}{y} = \frac{1}{s-1}\sum_{i=0}^{s-2} \frac{x^{-i}+(q-1)x^{i+1}}{1+(q-1)x}. \tag{26}$$

By examining each term of this sum separately, it is straightforward to verify that the right-hand side of (26) is a convex function of x, which tends to ∞ (resp. 1) as $x \to 0$ (resp. $x \to 1$). Since $y > -\frac{s}{2}$, the left-hand side of (26) is larger than 1, so there is a unique $x \in (0,1)$ which solves (26). Statement 1 follows. □

This establishes Item 3 of Lemma 5.2. □

Having completed the proof of Lemma 5.2, we have finished the proof of Theorem 2.6.

Acknowledgements

The first author would like to thank Yael Hacohen and Nati Linial for useful conversations. The second author would like to thank Venkat Guruswami for helpful feedback on a draft of this work.

References

[AEL95] Noga Alon, Jeff Edmonds, and Michael Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519. IEEE, 1995.

[Bol01] Béla Bollobás. *Random Graphs, Second Edition*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2001.

[CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 432–442, 2013.

[CS+04] Imre Csiszár, Paul C Shields, et al. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528, 2004.

[DL12] Zeev Dvir and Shachar Lovett. Subspace evasive sets. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 351–358. ACM, 2012.

[Eli57] Peter Elias. List decoding for noisy channels. *Wescon Convention Record, Part 2*, pages 94–104, 1957.

[Gal62] Robert G. Gallager. Low-density parity-check codes. *IRE Trans. Information Theory*, 8(1):21–28, 1962.

[GHK11] Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. *IEEE Trans. Information Theory*, 57(2):718–725, 2011.

[GHSZ02] Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Trans. Information Theory*, 48(5):1021–1034, 2002.

[GK16] Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016.

[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.

[Gur03] Venkatesan Guruswami. List decoding from erasures: Bounds and code constructions. *IEEE Transactions on Information Theory*, 49(11):2826–2833, 2003.

[Gur06] Venkatesan Guruswami. Iterative decoding of low-density parity check codes (a survey). *arXiv preprint cs/0610022*, 2006.

[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.

[GX12] Venkatesan Guruswami and Chaoping Xing. Folded codes from function field towers and improved optimal rate list decoding. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 339–350. ACM, 2012.

[GX13] Venkatesan Guruswami and Chaoping Xing. List decoding Reed-Solomon, Algebraic-Geometric, and Gabidulin subcodes up to the Singleton bound. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 843–852. ACM, 2013.

[HRW17] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes & applications. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 204–215. IEEE, 2017.

[HW18] Brett Hemenway and Mary Wootters. Linear-time list recovery of high-rate expander codes. *Information and Computation*, 261:202–218, 2018.

[Kop15] Swastik Kopparty. List-decoding multiplicity codes. *Theory of Computing*, 11(5):149–182, 2015.

[KRRZ+19] Swastik Kopparty, Nicolas Resch, Noga Ron-Zewi, Shubhangi Saraf, and Shashwat Silas. On list recovery of high-rate tensor codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 2019.

[KRSW18] Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved decoding of folded reed-solomon and multiplicity codes. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 212–223. IEEE, 2018.

[KRU13] Shrinivas Kudekar, Tom Richardson, and Rüdiger L Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Transactions on Information Theory*, 59(12):7761–7813, 2013.

[LMar] Nati Linial and Jonathan Mosheiff. On the weight distribution of random binary linear codes. *Random Structures & Algorithms*, to appear.

[LN94] Rudolf Lidl and Harald Niederreiter. *Introduction to finite fields and their applications*. Cambridge university press, 1994.

[LW18] Ray Li and Mary Wootters. Improved list-decodability of random linear binary codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[RW14] Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 764–773. ACM, 2014.

[RW18] Atri Rudra and Mary Wootters. Average-radius list-recovery of random linear codes. In *Proceedings of the 2018 ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2018.

[SS94] Michael Sipser and Daniel A Spielman. Expander codes. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 566–576. IEEE, 1994.

[Tan81] R Tanner. A recursive approach to low complexity codes. *IEEE Transactions on information theory*, 27(5):533–547, 1981.

[WJ08] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 853–860, 2013.

[Woz58] Jack Wozencraft. List decoding. *Quarter Progress Report*, 48:90–95, 1958.

[Zém01] Gillés Zémor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001.

[ZP81] Victor Vasilievich Zyablov and Mark Semenovich Pinsker. List concatenated decoding. *Problemy Peredachi Informatsii*, 17(4):29–33, 1981.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il