LDPC Codes Achieve List Decoding Capacity

Jonathan Mosheiff
Computer Science Department
Carnegie Mellon University
Pittsburgh, USA
jmosheif@cs.cmu.edu

Nicolas Resch
Computer Science Department
Carnegie Mellon University
Pittsburgh, USA
nresch@cs.cmu.edu

Noga Ron-Zewi
Department of Computer Science
University of Haifa
Haifa, Israel
noga@cs.haifa.ac.il

Shashwat Silas
Department of Computer Science
Stanford University
Stanford, USA
silas@stanford.edu

Mary Wootters
Department of Computer Science
Stanford University
Stanford, USA
marykw@stanford.edu

2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS) | 978-1-7281-9621-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/FOCS46700.2020.00050

***Abstract*—We show that Gallager's ensemble of Low-Density Parity Check (LDPC) codes achieves list-decoding capacity with high probability. These are the first graph-based codes shown to have this property. This result opens up a potential avenue towards truly linear-time list-decodable codes that 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.

In order to prove our results on LDPC codes, we establish sharp thresholds for when local properties are satisfied by a random linear code. More precisely, we show that for any local property $\mathcal{P}$, there is some R^* so that random linear codes of rate slightly less than R^* satisfy $\mathcal{P}$ with high probability, while random linear codes of rate slightly more than R^* with high probability do not. We also give a characterization of the threshold rate R^*.

This is an extended abstract. The full version is available at https://arxiv.org/abs/1909.06430.

I. 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 $z \in \Sigma^n$,

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

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

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

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

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00050

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]

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, most of these constructions 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. 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: We 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 I.1. *Are there (families) of LDPC codes that achieve list-decoding capacity?*

A. Contributions

Motivated by Question I.1, our contributions are as follows.

(1) We show that the answer to Question I.1 is "yes." More precisely, we show that random LDPC codes (the same ensemble studied by Gallager in his seminal work nearly 60 years ago [Gal62]), achieve list-decoding capacity with high probability.
(2) In fact, we show a stronger result: random LDPC codes satisfy with high probability any *local* property that random linear codes satisfy with high probability. We define local properties precisely below; informally, a local property is one defined by the exclusion of certain bad sets. List-decodability is a local property—it can be defined by the exclusion of any big set of vectors that are too close together—and this answers Question I.1.
(3) Along the way, we develop a characterization of the local properties that are satisfied with high probability by a random linear code. We show that for any local property $\mathcal{P}$, there is a threshold R^* so that random linear codes of rate slightly less than R^* satisfy $\mathcal{P}$ with high probability, while random linear codes of rate slightly greater than R^*

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

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

with high probability do not. Moreover, we give a characterization of the threshold R^*.

In [GLM+20], the above characterization is used to compute lower bounds on the list-decoding and list-recovery parameters of random linear codes. This additional application does not directly relate to LDPC codes.

We describe each of these contributions in more detail below.

(1) 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 $\mathbb{F}_q^*$.

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

Theorem I.2. *For any $R \in (0,1)$, $\varepsilon > 0$, prime power q, $\alpha \in (0, 1-1/q)$ 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 I.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.

Corollary I.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 I.5 (Other parameter regimes)**.** *We state Corollary I.4 as one example of what can be obtained by combining Theorem I.2 with one result on random linear codes. The result of [GHK11] degrades as $\alpha \to 1-1/q$, and so Corollary I.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 I-B below), and Theorem I.2 implies that these results carry over to random LDPC codes as well.*

(2) Random LDPC codes achieve any local property that random linear codes achieve: Theorem I.2 follows as a corollary of a much more general theorem. We show that any "local" property that 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 I-B) 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 $B \in \mathcal{B}$ and permutation $\pi : [n] \to [n]$, we also have that $\pi(B) \in \mathcal{B}$.

Definition I.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*

$$C \subseteq \Sigma^n \text{ satisfies } P_n \iff B \not\subseteq C \text{ for all } B \in \mathcal{B}_n.$$

We say that a family of random codes $C = \{C_{n_i}\}_{i \in \mathbb{N}}$ (where $\{n_i\}$ is an increasing sequence) satisfies $\mathcal{P}$ with high probability if $\lim_{i \to \infty} \Pr[C_{n_i} \text{ satisfies } P_{n_i}] = 1$. Similarly, we say that C almost surely does not satisfy $\mathcal{P}$ if $\lim_{i \to \infty} \Pr[C_{n_i} \text{ satisfies } P_{n_i}] = 0$.

[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 $\mathbb{F}_q^*$.

A code property is monotone decreasing if given a code C satisfying P, it holds that every code $C' \subseteq C$ also satisfies P. Note that every local property is monotone decreasing.

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.

For any $n \in \mathbb{N}$ and $R \in [0,1]$ such that $R \cdot n \in \mathbb{N}$, we denote a random linear length n code of rate R by $C^n_{\mathrm{RLC}}(R)$. Likewise, given s, n and R such that $s \mid n$ and $R \cdot s \in \mathbb{N}$, we denote a random s-LDPC code of length n and rate R by $C^n_{s\mathrm{LDPC}}(R)$. Whenever we use these notations, it is implicitly assumed that the relevant divisibility conditions are satisfied.

Let $\mathcal{P} = \{P_n\}_{n\in\mathbb{N}}$ be a monotone decreasing property of linear codes. We define

$$R^n_{\mathrm{RLC}}(\mathcal{P}) := \sup\{R \in [0,1] : \Pr[C^n_{\mathrm{RLC}}(R) \text{ satisfies } P_n] \geq 1/2\} \tag{2}$$

if such an R exists. Otherwise we define $R^n_{\mathrm{RLC}}(\mathcal{P}) = 0$.

Remark I.7. *If $\mathcal{P}$ is a monotone decreasing property then the function $\Pr[C^n_{\mathrm{RLC}}(R)$ satisfies $P_n]$ is monotone decreasing in R. This can be proved by a standard coupling argument, akin to [Bol01, Thm. 2.1].*

With the notation out of the way, we are ready to state our more general theorem about random LDPC codes. Essentially, this theorem says that every local property that holds with high probability for a random linear code also holds with high probability for a random s-LDPC code of approximately the same rate. This approximation improves as s grows.

Theorem I.8 (Main)**.** *Let $\mathcal{P} = (P_n)_{n\in\mathbb{N}}$ be a b-local property with $\bar{R} := \limsup_{n\to\infty} R^n_{\mathrm{RLC}}(\mathcal{P}) < 1$. For any $\varepsilon > 0$ and prime power q, there exists $s_0 = s_0(\varepsilon, \bar{R}, q, b) \geq 1$ such that for any odd $s \geq s_0$ and any sequence $\{R_n\}_{n\in\mathbb{N}}$, if $R_n \leq R^n_{\mathrm{RLC}}(\mathcal{P}) - \varepsilon$ for all n, then the code ensemble $C^n_{s\mathrm{LDPC}}(R_n)$ satisfies $\mathcal{P}$ with high probability.*

Remark I.9 (The dependence on $\varepsilon, \bar{R}, q, b$)**.** *An inspection of the proof shows that we may take*

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

[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.

The existence of a reduction like the one in Theorem I.8 is surprising, at least to the authors. There is a lot more structure in a random LDPC code than 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.

Remark I.10 (A converse to Theorem I.8?)**.** *One may be tempted to conjecture that the converse of Theorem I.8 holds as well. Namely, in the setting of Theorem I.8, if $R_{n_i} \geq R^{n_i}_{\mathrm{RLC}}(\mathcal{P}) + \varepsilon$ for all i, then the code ensemble $C_{s\mathrm{LDPC}}(R_n)$ almost surely does not satisfy $\mathcal{P}$. However, this turns out to be false, due to the following example. Assume that $q = 2$ and consider the 1-local property $\mathcal{P} := (P_n)_{n\in\mathbb{N}}$, where P_n is the set of all length n linear codes that only contain even weight codewords. It is not hard to see (e.g., using Theorem II.8) that $R^n_{\mathrm{RLC}}(\mathcal{P})$ tends to 0 as $n \to \infty$. On the other hand, if $\frac{n}{s}$ is even, then every s-LDPC code (including, say, a code of rate $\frac{1}{2}$) satisfies $\mathcal{P}$, contradicting this conjecture.*

However, the above counter-example relies on a technicality involving divisibility criteria. It is an interesting question whether a natural converse of Theorem I.8 holds if we additionally assume that $\mathcal{P}$ belongs to some natural class of "nicely behaved" properties that precludes counter-examples of this sort.

(3) A characterization of local properties satisfied by random linear codes: In order to prove Theorems I.2 and I.8, we develop a new characterization of the local properties satisfied by a random linear code. Our formal theorem is given as Theorem II.8. Informally, this theorem implies that for any monotone decreasing property $\mathcal{P}$, there is a sharp threshold R^* so that random linear codes of rate slightly less than R^* with high probability satisfy $\mathcal{P}$, while random linear codes of rate slightly larger than R^* with high probability do not. Moreover, we give a characterization of R^*.

Formally, we have the following definition, recalling the definition of $R^n_{\mathrm{RLC}}(R_n)$ from (2).

Definition I.11 (Sharpness for random linear codes)**.** *We say that the property $\mathcal{P}$ is* sharp for random linear codes *if for every $\varepsilon > 0$ there holds:*

[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.

- *If $R_n \leq R^n_{\mathrm{RLC}}(\mathcal{P}) - \varepsilon$ for large enough n, then the code ensemble $C^n_{\mathrm{RLC}}(R_n)$ ($n \in \mathbb{N}$) satisfies $\mathcal{P}$ with high probability.*
- *If $R_n \geq R^n_{\mathrm{RLC}}(\mathcal{P}) + \varepsilon$ for large enough n, then the code ensemble $C^n_{\mathrm{RLC}}(R_n)$ ($n \in \mathbb{N}$) almost surely does not satisfy $\mathcal{P}$.*

If a property $\mathcal{P}$ is sharp, we sometimes refer to $R^n_{\mathrm{RLC}}(\mathcal{P})$ as the *threshold* for $\mathcal{P}$.

Theorem II.8 has two corollaries. The first is that local properties are sharp for random linear codes:

Corollary I.12. *Every local property is sharp for random linear codes.*

The second corollary of Theorem II.8 is a characterization of $R^n_{\mathrm{RLC}}(\mathcal{P})$. This characterization requires some definitions to state formally, so we defer the formal statement to Theorem II.8. However, it has an intuitive interpretation, which we sketch here.

Recall that a local property is defined by a permutation-invariant collection $\mathcal{B}_n$ of excluded sets. For simplicity of exposition, suppose that all of the sets $B \in \mathcal{B}_n$ have size exactly b, and moreover that they all have dimension exactly b. (This assumption is helpful for exposition but not necessary for our analysis). In this case, it is easy to compute the probability that each individual set $B \in \mathcal{B}_n$ is contained in $C_{\mathrm{RLC}}(R)$ (see Fact II.2):

$$\Pr\left[B \subseteq C_{\mathrm{RLC}}(R)\right] = q^{-(1-R)nb}.$$

Thus, we have

$$\mathbb{E}\left|\{B \in \mathcal{B}_n \,:\, B \subseteq C_{\mathrm{RLC}}(R)\}\right| = |\mathcal{B}_n| \cdot q^{-(1-R)nb}.$$

Thus, as long as

$$R < R^{\mathbb{E}}_{\mathrm{RLC}}(\mathcal{B}_n) := 1 - \frac{\log|\mathcal{B}_n|}{nb},$$

we are guaranteed by Markov's inequality that with high probability, no elements of $\mathcal{B}_n$ appear in $C_{\mathrm{RLC}}(R)$. However, what if $R > R^{\mathbb{E}}_{\mathrm{RLC}}(\mathcal{B}_n)$? It turns out that the statement above is not tight: in some cases it is likely that no elements of $\mathcal{B}_n$ appear in $C_{\mathrm{RLC}}(R)$ even if the rate R is significantly larger than $R^{\mathbb{E}}_{\mathrm{RLC}}(\mathcal{B}_n)$. We give an example in Example II.5 of when this can occur.

Our result in Theorem II.8 pins down exactly when this can occur. Informally, it happens only because some projection $\mathcal{B}'_n$ of the collection $\mathcal{B}_n$ is more favorable than one might expect, in the sense that $R^{\mathbb{E}}_{\mathrm{RLC}}(\mathcal{B}'_n)$ is larger than one might expect. In this case, the "correct" threshold is precisely $R^{\mathbb{E}}_{\mathrm{RLC}}(\mathcal{B}'_n)$.

Thus, Theorem II.8 also provides a characterization of which sorts of "bad" lists B (up to a permutation of the coordinates) are contained in a random linear code of a particular rate. We hope that this characterization will be useful in the study of random linear codes themselves, in addition to random LDPC codes.

The full power of Theorem II.8 (including the characterization of $R^n_{\mathrm{RLC}}(\mathcal{P})$ described above) is used to prove Theorem I.8. However, given Theorem I.8, Theorem I.2 readily follows from Corollary I.12 itself:

Proof of Theorem I.2: Let $\mathcal{P}$ denote the property of being (α, L)-list-decodable. Note that $\mathcal{P}$ is a local property: for any $n \in \mathbb{N}$, take $\mathcal{B}_n$ to be the collection of all sets of $L+1$ vectors in $\mathbb{F}_q^n$ contained in some Hamming ball of radius α. Now, fix some $R \in (0,1)$ and assume that a random linear code of rate R satisfies $\mathcal{P}$ with high probability. Corollary I.12 implies that $R^n_{\mathrm{RLC}}(\mathcal{P}) \leq R + o_{n\to\infty}(1)$.

Next, it is not hard to verify that $\limsup_{n\to\infty} R^n_{\mathrm{RLC}}(\mathcal{P}) \leq 1 - h_q(\alpha) < 1$. Indeed, it follows from the list-decoding capacity theorem (e.g. [LW18, Thm 1.1]) that for large enough n there are no (α, L)-list-decodable codes of rate $1 - h_q(\alpha) + \varepsilon$. In particular, this means that a random linear code of rate $1 - h_q(\alpha) + \varepsilon$ almost surely does not satisfy $\mathcal{P}$.

Theorem I.8 now immediately yields Theorem I.2. ■

We give a high-level overview of the proof of Theorem I.8 in Section II below after a discussion of related work in Section I-B.

B. 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 constant-sized 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 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]

Finally, we note that recent work [DHK$^+$19] has given an algorithm to list-decode codes based on high-dimensional expanders, but these results are far from list-decoding capacity.

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.

Relationship to graph properties: Finally, we note that our results providing sharp thresholds of local properties for random linear codes are reminiscent of classic results about local properties of random graphs. We discuss this connection more in Remark II.10.

C. Discussion and open questions

In this work, we answer Question I.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. Our results 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) **What other applications of Theorem II.8 are there?** In subsequent work [GLM$^+$20], the characterization of a sharp threshold for local properties of random linear codes (Theorem II.8) was already demonstrated to be useful beyond our work on LDPC codes. We hope to see additional applications of this result. For example, Remark II.9 implies that to prove that $C_{\mathrm{RLC}}(R - \varepsilon)$ satisfies a local property $\mathcal{P}$ with probability $1 - 2^{-\Omega(n)}$, it suffices to show that $C_{\mathrm{RLC}}(R)$ satisfies $\mathcal{P}$ with some tiny probability (at least $2^{-o(n)}$). Are there situations where this could be useful?

[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 a strengthening of list-decoding where instead of requiring that no set of $L+1$ codewords are *all* close to some z, we require that no set of $L+1$ codewords has small *average* distance 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 the observed coordinates.

[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) **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.
4) **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?

D. Organization and main building blocks

In Section II, we give a high-level overview of the proof of Theorem I.8. This proof relies on three building blocks:

- First, Lemma II.7 establishes sharp thresholds for certain local properties, and effectively characterizes the sorts of sets $B \subseteq \mathbb{F}_q^n$ that are contained in a random linear code. We prove this lemma in Section 3 of the full version. Using Lemma II.7 we prove Theorem II.8, which pins down a sharp threshold for any local property of a random linear code.
- Second, Lemma II.13 shows that for a set B with a certain property called δ-*smoothness*, the probability that B appears in a random s-LDPC code is not much larger than the probability that it appears in a random linear code of the same rate. We prove this Lemma II.13 in Section 4 of the full version using Fourier analysis.
  Together with Lemma II.7, Lemma II.13 implies that any property satisfied with high probability by a random linear code is also satisfied with high probability by a random s-LDPC code of similar rate, provided that we can restrict our attention to δ-smooth sets B. It turns out that for any code with good distance,[9] we may indeed restrict our attention to such sets, so it remains to show that random s-LDPC codes have good distance.
- Third, Theorem II.14 shows that random s-LDPC codes do indeed have good distance with high probability. 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 II.14 in Section 5 of the full version using techniques from exponential families.

Together, these three building blocks can be used to establish Theorem I.8, as we show next in Section II.

II. High-level idea: proof of Theorem I.8

In this section we prove our main theorem (Theorem I.8) using the building blocks outlined in Section I-D. 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. However, we will need a few technical definitions, outlined in Section II-A.

A. 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 τ_M, the row distribution of M, as follows: for any $v \in \mathbb{F}_q^\ell$,

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

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 $v \in \mathrm{supp}(\tau)$.[10] 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

$$|\mathcal{D}_{n,\ell}| \leq \binom{n + q^\ell - 1}{q^\ell - 1}. \tag{3}$$

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. We say that a code C *contains* τ to mean that $M \subset C$ for some matrix $M \in \mathcal{M}_{n,\tau}$. Let

$$\mathcal{L}_\tau = \{n \in \mathbb{N} \mid \tau(u) \cdot n \text{ is an integer for all } u \in \mathbb{F}_q^\ell\}.$$

Note that for C to contain τ, a trivial necessary condition is that the length of C belongs to $\mathcal{L}_\tau$. Let $\mathcal{P}_\tau$ denote the ℓ-local property of not containing any matrix from

[9] The distance of a code is the minimum distance between any two codewords.

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

the set $\mathcal{M}_{n,\tau}$. Properties of the form $\mathcal{P}_\tau$ are particularly useful to us due to the following observation:

Observation II.1 (Local property decomposition)**.** *Let $\mathcal{P} = (P_n)_{n\in\mathbb{N}}$ be an ℓ-local property for some $\ell \in \mathbb{N}$. Then, for every $n \in \mathbb{N}$ there exists $T_n \subseteq \mathcal{D}_{n,\ell}$ such that*

$$C \subseteq \mathbb{F}_q^n \text{ satisfies } P_n \iff C \text{ satisfies } P_\tau \text{ for all } \tau \in T_n.$$

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

$$H(\tau) := -\sum_{x\in\mathrm{supp}(\tau)} \tau(x)\log(\tau(x))$$

and

$$H_q(\tau) := \frac{H(\tau)}{\log q}.$$

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

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 from the definition of a random s-LDPC code in Section I-A. 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.

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. 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.

We will also require the following standard facts:

Fact II.2. *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 $q^{-(1-R)\cdot\mathrm{rank}(M)\cdot n}$.*

Fact II.3 ([CS+04], Lemma 2.2)**.** *For any distribution $\tau \in \mathcal{D}_{n,\ell}$,*

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

B. Sharp thresholds for local properties for random linear codes

The first building block is Lemma II.7 below, which shows that for every distribution $\tau \in \mathcal{D}_{n,\ell}$, the property $\mathcal{P}_\tau$ is sharp for random linear codes. Moreover we give a simple characterization of $R_{\mathrm{RLC}}(\mathcal{P}_\tau)$. As an easy corollary, we get Theorem II.8, which generalizes Lemma II.7 to any local property, not necessarily of the form $\mathcal{P}_\tau$.

Before stating Lemma II.7 we give some intuition. Fix some distribution τ over $\mathbb{F}_q^\ell$. Let C be a random linear code of length $n \in \mathcal{L}_\tau$ and rate R. We seek a threshold rate, above which C is likely to contain τ. It is natural to attempt a first-moment approach to this problem and ask what is the expected number of matrices from $\mathcal{M}_{n,\tau}$ which are contained in C. Note that $|\mathcal{M}_{n,\tau}| = q^{n\cdot H_q(\tau)}\cdot\mathrm{poly}(n)$. Indeed, if $u_1,\ldots,u_{q^\ell}$ are an enumeration of $\mathbb{F}_q^\ell$, then $\mathcal{M}_{n,\tau}$ is in one-to-one correspondence with partitions on $[n]$ into q^ℓ subsets of sizes $n\tau(u_1),\ldots,n\tau(u_{q^\ell})$. That is, $|\mathcal{M}_{n,\tau}| = \binom{n}{n\tau(u_1),\ldots,n\tau(u_{q^\ell})} = q^{nH_q(\tau)}\cdot\mathrm{poly}(n)$, where the last estimate follows from Fact II.3, and relies on our assumption that $n \in \mathcal{L}_\tau$.

Given $M \in \mathcal{M}_{n,\tau}$, the code C contains M with probability $q^{-n\cdot(1-R)\cdot d(\tau)}$ (see Fact II.2). Hence, in expectation, C contains roughly $q^{n\cdot(H_q(\tau)-(1-R)\cdot d(\tau))}$ matrices from $\mathcal{M}_{n,\tau}$. In particular, this expectation grows (resp. decays) exponentially in n, when R is larger (resp. smaller) than $1-\frac{H_q(\tau)}{d(\tau)}$. This motivates the following definition.

Definition II.4 (Expectation threshold)**.** *Given a distribution τ over $\mathbb{F}_q^\ell$, define the* expectation-threshold

$$R^{\mathbb{E}}_{\mathrm{RLC}}(\tau) := 1-\frac{H_q(\tau)}{d(\tau)}.$$

It follows immediately from a first-moment argument that if $R < R^{\mathbb{E}}_{\mathrm{RLC}}(\tau)$ then C satisfies $\mathcal{P}_\tau$ with probability $1-e^{-\Omega(n)}$. In particular, as n grows we get the lower bound

$$R^n_{\mathrm{RLC}}(\mathcal{P}_\tau) \ge R^{\mathbb{E}}_{\mathrm{RLC}}(\tau) - o(1). \tag{4}$$

However, as the following example shows, this bound is not tight.

Example II.5. *Let $q=2$, $\ell=3$ and consider the distribution τ over $\mathbb{F}_2^3$, given by the following table:*

u	$\tau(u)$
$(1,0,0)$	$1/4$
$(0,1,0)$	$1/4$
$(1,0,1)$	$1/4$
$(0,1,1)$	$1/4$
Every other vector	0

It is straightforward to compute $R^{\mathbb{E}}_{\mathrm{RLC}}(\tau) = 1 - \frac{H_2(\tau)}{d(\tau)} = 1 - \frac{2}{3} = \frac{1}{3}$.

We claim that $R^n_{\mathrm{RLC}}(\mathcal{P}_\tau)$ *is bounded away from* $R^{\mathbb{E}}_{\mathrm{RLC}}(\tau)$. *Let* $A := \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right) \in \mathbb{F}_2^{2\times 3}$ *represent the linear map which projects a vector onto its first two coordinates. Let* τ' *denote the distribution of* Au, *where* u *is a random vector sampled from* τ. *Thus,* τ' *is distributed as follows:*

u	$\tau'(u)$
$(1,0)$	$1/2$
$(0,1)$	$1/2$
Every other vector	0

Note that a code C *which contains a matrix* M *from* $\mathcal{M}_{n,\tau}$ *must contain the first two columns of* M: *that is, the matrix* MA^T. *Consequently, every code which satisfies* $\mathcal{P}_{\tau'}$ *also satisfies* $\mathcal{P}_\tau$, *and so* $R^n_{\mathrm{RLC}}(\mathcal{P}_\tau) \geq R^n_{\mathrm{RLC}}(\mathcal{P}_{\tau'})$.

Finally, (4) yields

$$R^n_{\mathrm{RLC}}(\mathcal{P}_{\tau'}) \geq R^{\mathbb{E}}_{\mathrm{RLC}}(\tau') - o(1) = 1 - \frac{H_2(\tau')}{d(\tau')} - o(1)$$
$$= 1 - \frac{1}{2} - o(1) = \frac{1}{2} - o(1)$$

and we conclude that

$$R^n_{\mathrm{RLC}}(\mathcal{P}_\tau) \geq \frac{1}{2} - o(1) > \frac{1}{3} = R^{\mathbb{E}}_{\mathrm{RLC}}(\tau)$$

for large n.

In Example II.5, the bound of $R^{\mathbb{E}}_{\mathrm{RLC}}(\tau)$ was not tight, in that the rate can actually be much higher than we would expect from a first-moment argument. The reason was that there was some linear map A so that $\tau' = A\tau$ had a larger value of $R^{\mathbb{E}}_{\mathrm{RLC}}(\tau')$. We will show below that this is the only reason that $R^{\mathbb{E}}_{\mathrm{RLC}}(\tau)$ might not be the right answer. To make this precise, we introduce the following definition.

Definition II.6 (Implied distribution)**.** *Let* τ *be a distribution over* $\mathbb{F}_q^\ell$ *and let* $A \in \mathbb{F}_q^{m\times \ell}$ *be a rank* m *matrix for some* $m \leq \ell$. *The distribution of the random vector* Au, *where* u *is randomly sampled from* τ, *is said to be* τ*-implied. We denote the set of* τ*-implied distributions by* $\mathcal{I}_\tau$.

Note that whenever $\tau' \in \mathcal{I}_\tau$, a linear code satisfying $\mathcal{P}_{\tau'}$ must also satisfy $\mathcal{P}_\tau$. Indeed, in the setting of Definition II.6 assume that C contains a matrix $M \in \mathcal{M}_{n,\tau}$. By linearity, C also contains the matrix MA^T, which belongs to $\mathcal{M}_{n,\tau'}$. Hence, not satisfying $\mathcal{P}_\tau$ implies not satisfying $\mathcal{P}_{\tau'}$. Consequently, $R^n_{\mathrm{RLC}}(\mathcal{P}_\tau) \geq R^n_{\mathrm{RLC}}(\mathcal{P}_{\tau'})$.

Inequality (4) now yields the stronger bound

$$R_{\mathrm{RLC}}(\mathcal{P}_\tau)^n \geq \max_{\tau'\in\mathcal{I}_\tau} R^{\mathbb{E}}_{\mathrm{RLC}}(\tau') - o(1). \qquad (5)$$

Lemma II.7 below essentially says that (5) is tight, and that $\mathcal{P}_\tau$ is sharp for random linear codes. We prove this Lemma in Section 3 of the full version.

Lemma II.7 (Sharp threshold for $\mathcal{P}_\tau$ for random linear codes)**.** *Let* $\ell \in \mathbb{N}$ *and let* τ *be a distribution over* $\mathbb{F}_q^\ell$. *Denote* $R^*_\tau = \max_{\tau'\in\mathcal{I}_\tau} R^{\mathbb{E}}_{\mathrm{RLC}}(\tau')$. *Then*

$$R^n_{\mathrm{RLC}}(\mathcal{P}_\tau) = R^*_\tau \pm o_{n\to\infty}(1).$$

for $n \in \mathcal{L}_\tau$. *Moreover,* $\mathcal{P}_\tau$ *is sharp for random linear codes. Specifically, fix any* $\varepsilon > 0$, *and let* C *be a random linear code of rate* R *and length* $n \in \mathcal{L}_\tau$. *The following holds:*

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

We now can conclude a more general result.

Theorem II.8 (Sharp thresholds for local properties for random linear codes)**.** *Fix* $\ell \in \mathbb{N}$. *Let* $\mathcal{P} = (P_n)_{n\in\mathbb{N}}$ *be an* ℓ*-local property and let* $(T_n)_{n\in\mathbb{N}}$ *be as in Observation II.1. Then* $\mathcal{P}$ *is sharp for random linear codes and*

$$R^n_{\mathrm{RLC}}(\mathcal{P}) = \min_{\tau\in T_n} \max_{\tau'\in\mathcal{I}_\tau} R^{\mathbb{E}}_{\mathrm{RLC}}(\tau') \pm o_{n\to\infty}(1).$$

Remark II.9 (Probability of satisfying $\mathcal{P}$ in Theorem II.8)**.** *Fix* $\varepsilon > 0$. *An inspection of the proof of Theorem II.8 shows that* $C^n_{\mathrm{RLC}}(R^n_{\mathrm{RLC}}(\mathcal{P}) - \varepsilon)$ *satisfies* $\mathcal{P}$ *with probability* $1 - 2^{-\Omega(n)}$. *Likewise,* $C^n_{\mathrm{RLC}}(R^n_{\mathrm{RLC}}(\mathcal{P}) + \varepsilon)$ *satisfies* $\mathcal{P}$ *with probability* $2^{-\Omega(n)}$.

Remark II.10 (Relationship to random graphs)**.** *Lemma II.7 has an analog in the theory of random graphs. Fix a constant-sized graph* H *and let* G *be a random graph in the* $G(n,p)$ *model. A natural problem is to determine the threshold for the appearance of* H *as a sub-graph of* G. *The answer (see for example [Bol01, Sec. 4.2]) is that a copy of* H *is likely to occur in* G *whenever* p *is large enough so that every subgraph of* H *has, in expectation,* $\omega(1)$ *copies as subgraphs of* G. *To complete the analogy, equate* H *with* τ, *and a subgraph of* H *with a* τ*-implied distribution.*

We also mention the recent breakthrough result of Frankston et al., which studies this relationship between thresholds and expectations of sub-structures in a more general framework [FKNP19]. However, since the properties that they study are not necessarily local, it is impossible for that work to precisely pinpoint the thresholds, as we do in our work.

C. Probability that a matrix is contained in a random s-LDPC code

The second building block shows that given a matrix $M \in \mathbb{F}_q^{n\times \ell}$, the probability that M is contained in a random s-LDPC code is not much larger than that of appearing in a random linear code, provided that M is δ-*smooth* (defined below).

Definition II.11 (Smooth distribution)**.** *Let $\delta > 0$. We say that a distribution τ over $\mathbb{F}_q^\ell$ is δ-smooth if $\Pr_{v\sim\tau}[\langle u, v\rangle \neq 0] \geq \delta$ for all $u \in \mathbb{F}_q^\ell \setminus \{0\}$. If $M \in \mathbb{F}_q^{n\times\ell}$ is such that τ_M is δ-smooth, we also say that M is δ-smooth.*

Remark II.12 (Relationship to distance)**.** *In coding-theoretic terms, τ_M is δ-smooth if and only if the code $\{Mu : u \in \mathbb{F}_q^\ell\}$ has relative distance at least δ and M is full-rank. 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 row distribution is contained in a random LDPC code with sufficiently large sparsity parameter. We prove this lemma in Section 4 of the full version.

Lemma II.13 (Probability that a random LDPC code contains a matrix)**.** *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 δ-smooth. Then the probability p that M is contained in a random s-LDPC code of length n and rate R satisfies*

$$p \leq q^{-(1-\varepsilon)\cdot(1-R)\cdot\ell\cdot n}.$$

Given a smooth distribution τ, in light of Fact II.2, Lemma II.13 says that the expected number of matrices from $\mathcal{M}_{n,\tau}$ in a random s-LDPC code is not much larger than this number for a random linear code. If we ignore the constraint that τ must be smooth, then together with Lemma II.7 the above would imply Theorem I.8. Indeed, if a distribution τ is unlikely to appear in a random linear code then Lemma II.7 shows that some τ-implied distribution τ' appears $o(1)$ times in expectation in the random linear code. By Lemma II.13, τ' appears $o(1)$ times in the random LDPC code as well, so the LDPC code is unlikely to contain τ'. Thus, it is also unlikely to contain τ. (Of course, we cannot ignore the constraint that τ must be smooth; we will address this in our next building block discussed in Section II-D).

The proof of Lemma II.13 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 $[n]$.[11] Thus, the probability that a matrix $M \in \mathcal{M}_{n,\tau}$ is in C can be 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.

D. Distance of random s-LDPC codes

As noted above, the first two building blocks show that for any δ-smooth distribution $\tau \sim \mathbb{F}_q^\ell$, a random LDPC code of rate slightly below $R^n_{\mathrm{RLC}}(\mathcal{P}_\tau)$ is unlikely to contain τ. The third and final building block shows that we may restrict our attention to δ-smooth distributions.

As noted in Remark II.12, 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 at least δ, it does not contain any matrices that are not δ-smooth. 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.[12] Theorem II.14 generalizes this result to s-LDPC codes over any alphabet. Below, $h_q(x)$ is the q-ary entropy function (as in (1)).

Theorem II.14 (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 II.15 (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*

[11] 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 significantly affect the probability of the event of interest.

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

the first author [LM20]. 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 1's, 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 in the full version is very similar to the corresponding proof in [Gal62] at a technical level.

E. Proof of Theorem I.8

Theorem I.8 is proven in the full version as an immediate consequence of the building blocks above.

ACKNOWLEDGEMENTS

JM would like to thank Yael Hacohen and Nati Linial for useful conversations. NRe would like to thank Venkat Guruswami for helpful feedback on a draft of this work.

JM is partially supported by NSF grants CCF-1814603 and CCF-1563742. A significant portion of this work was accomplished while JM was a postdoctoral fellow at the Weizmann Institute, 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 and ISF grant 735/20. 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.

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.

[DHK+19] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta Shma. List decoding with double samplers. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2134–2153. SIAM, 2019.

[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.

[FKNP19] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *arXiv preprint arXiv:1910.13433*, 2019.

[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.

[GLM+20] Venkatesan Guruswami, Ray Li, Jonathan Mosheiff, Nicolas Resch, Shashwat Silas, and Mary Wootters. Bounds for list-decoding and list-recovery of random linear codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*, 2020.

[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.

[LM20] Nati Linial and Jonathan Mosheiff. On the weight distribution of random binary linear codes. *Random Structures & Algorithms*, 56(1):5–36, 2020.

[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.

[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.

[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.