Combinatorial Bounds for List Decoding

Venkatesan Guruswami [*] Johan Håstad [†] Madhu Sudan [‡] David Zuckerman [§]

Abstract

Informally, an error-correcting code has "nice" list-decodability properties if every Hamming ball of "large" radius has a "small" number of codewords in it. Here, we report linear codes with non-trivial list-decodability: i.e., codes of large rate that are nicely list-decodable, and codes of large distance that are *not* nicely list-decodable.

1 Introduction

List decoding was introduced independently by Elias [3] and Wozencraft [15] as a relaxation of the "classical" notion of decoding by allowing a list of codewords as answers. The decoding is considered successful as long as the correct message is included in the list. Formally, an instance of the list decoding problem for a code $\mathbf{C}$ consists of a received word $\mathbf{r}$ and an error bound e, and the goal is to output a list of *all* codewords of $\mathbf{C}$ that are within a Hamming distance of e from $\mathbf{r}$.

Early works in this area [3, 15, 10] focused on the probability of error in this more general decoding model. Research in the eighties applied this notion in a more adversarial setting and investigated what happens if the error is effected by an adversary or a "jammer", as opposed to a probabilistic channel. Works of Zyablov and Pinsker [16], Blinovskii [1] and Elias [4] applied in this setting. (The paper by Elias [4] also gives a very good summary of the prior work and history.) The basic (purely combinatorial) question raised in this setting is: How many errors could still be recovered from, with lists of small size? Two basic parameters of interest thus are the number of errors and the allowed size of the output list. These parameters are usually studied as a function of some of the more classical parameters of error-correcting codes. How large can the rate of a code be if we want small list sizes for a certain number of errors? And how do codes of large minimum distance perform with respect to list decoding? Recently there has been rejuvenated interest in this line of work thanks to the development of some efficient algorithms for list decoding in [12, 6, 11, 7]. These algorithms decode with polynomial sized lists (and sometimes with constant sized lists) for a number of errors that is much more than half the minimum distance of the code, and investigations of the tightness of the algorithms have led Hoholdt and Justesen [9] to re-initiate the investigation of the combinatorial bounds on list decoding.

[*] Laboratory for Computer Science, 200 Technology Square, Cambridge, MA 02139, USA. Email: venkat@theory.lcs.mit.edu.

[†] Department of Numerical Analysis and Computer Science, Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Email: johanh@nada.kth.se. Supported in part by the Göran Gustafsson foundation.

[‡] Laboratory for Computer Science, 200 Technology Square, Cambridge, MA 02139, USA. Email: madhu@mit.edu. Supported in part by a Sloan Foundation Fellowship and NSF Career Award CCR-9875511.

[§] Department of Computer Science, University of Texas, Austin, TX 78712, USA. Email: diz@cs.utexas.edu. Currently on leave at the University of California, Berkeley. Supported in part by NSF Grant CCR-9912428, NSF NYI Grant CCR-9457799, and a David and Lucile Packard Fellowship for Science and Engineering.

In this paper we continue the investigation of bounds on list decoding. In particular, we investigate codes that exhibit non-trivial list decoding performance. Specifically, we report the existence of linear codes of large rate that are nicely list-decodable, and codes of large minimum distance which are not nicely list-decodable (the precise quantitative versions of these results are stated in the next section). To motivate this study we first fix some standard notation and then define two fundamental questions (parameters) to study in the context of list decoding.

2 Definitions and Main Results

For a prime power q, let $\mathbb{F}_q$ denote a finite field of cardinality q. An $[n,k]_q$ (linear) code C is a k-dimensional vector space in $\mathbb{F}_2^n$. Unless explicitly mentioned otherwise, throughout this paper, we will only be interested in linear codes and will moreover restrict ourselves to the binary case (when $q = 2$).

For vectors $x, y \in \mathbb{F}_2^n$, let $\mathrm{distance}(x,y)$ denote the Hamming distance between them. The minimum distance of a code C, denoted $\mathrm{distance}(C)$, is the quantity $\min_{x,y \in C, x \neq y}\{\mathrm{distance}(x,y)\}$.

Since the main thrust of this paper is the asymptotic performance of the codes, we define analogs of the quantities above for infinite families of codes. An infinite family of (binary) codes is a family $\mathcal{C} = \{C_i | i \in \mathbb{Z}^+\}$ where C_i is an $[n_i, k_i]_2$ code with $n_i > n_{i-1}$. We define the *rate* of an infinite family of codes $\mathcal{C}$ to be $\mathrm{rate}(\mathcal{C}) = \inf_i \left\{\frac{k_i}{n_i}\right\}$, and the *(relative) distance* of an infinite family of codes $\mathcal{C}$ to be $\Delta(\mathcal{C}) = \inf_i \left\{\frac{\mathrm{distance}(C_i)}{n_i}\right\}$.

We now define the list decoding radius of a code. For non-negative integer r and $x \in \mathbb{F}_2^n$, let $B(x,r)$ denote the ball of radius r around x, i.e., $B(x,r) = \{y \in \mathbb{F}_2^n | \mathrm{distance}(x,y) \leq r\}$.

Definition 1 (List decoding Radius) *For an* $[n,k]$ *binary code* C*, and list size* ℓ*, the* list of ℓ decoding radius *of* C*, denoted* $\mathrm{radius}(C,\ell)$ *is defined to be*

$$\mathrm{radius}(C,\ell) = \max\left\{e | \forall x \in \mathbb{F}_2^n,\ \ |B(x,e) \cap C| \leq \ell\right\}.$$

For a function $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$*, and an infinite family of codes* $\mathcal{C}$*, let the list of* ℓ *decoding radius of* $\mathcal{C}$*, denoted* $\mathrm{Rad}(\mathcal{C},\ell)$ *be the quantity*

$$\mathrm{Rad}(\mathcal{C},\ell) = \inf_i \left\{\frac{\mathrm{radius}(C_i, \ell(n_i))}{n_i}\right\}.$$

It is interesting to study the list decoding radius of infinite families of codes as a function of their distance and rate, when the list size is either bounded by a constant, or grows as polynomially in the length of the code. Within this scope the broad nature of the two main questions are: (1) Do there exist codes of large rate with large list decoding radius for a fixed function ℓ? and (2) Do there exist codes of large distance with small list decoding radius for a given function ℓ? Note that the other two questions are uninteresting: specifically, it is possible to construct codes of small rate that have small list decoding radius (for example, by taking a linear code and just adding few codewords at a small distance to some existing codeword); and it is possible to construct codes of small distance that have large list decoding radius even for lists of size 2 (for example by taking a code of large minimum distance and adding one codeword at a small distance to some existing codeword). In what follows we introduce some formal parameters to study the above questions.

2.1 List decoding radius vs. Rate of the code

Definition 2 (Upper bound on list decoding radius) *For real rate $0 \leq R \leq 1$ and list size $l : \mathbb{Z}^+ \to \mathbb{Z}^+$, the* upper bound on list of ℓ decoding radius *for (binary) codes of rate R, denoted $U_\ell(R)$, is defined to be*

$$U_\ell(R) = \sup_{\mathcal{C} \mid \text{rate}(\mathcal{C}) \geq R} \text{Rad}(\mathcal{C}, \ell).$$

Note that the reason for the term "upper bound" is that $U_\ell(R)$ is the list decoding radius of the *best code* (i.e. one with largest possible list decoding radius) among codes that have at least a certain rate. The case where the list size function is a constant or a polynomial is of special interest to us, and we consider the following definitions.

Definition 3 *For real rate $0 \leq R \leq 1$ and constant c, the quantity $U_c^{\text{const}}(R)$ is defined to be $U_\ell(R)$ where $\ell(n) = c$. The quantity $U_c^{\text{poly}}(R)$ is defined to be $\sup_{c_1} U_{\ell_{c_1}}(R)$ where $\ell_{c_1}(n) = c_1 n^c$. The quantity $U^{\text{const}}(R)$ (resp. $U^{\text{poly}}(R)$) will denote the quantity $\limsup_{c\to\infty}\{U_c^{\text{const}}(R)\}$ (resp. $\limsup_{c\to\infty}\{U_c^{\text{poly}}(R)\}$).*

Thus the quantities $U^{\text{const}}(R)$ and $U^{\text{poly}}(R)$ denote the maximum possible value of the (relative) list decoding radius for lists of constant and polynomial size, respectively. These quantities are actually surprisingly well-understood. Zyablov and Pinsker [16] showed that $U^{\text{const}}(R) = U^{\text{poly}}(R) = H^{-1}(1 - R)$ which matches the Gilbert-Varshamov bound (here $H(\cdot)$ is the binary entropy function: $H(x) = -x \lg x - (1 - x) \lg(1 - x)$). The behavior of the upper bound on list decoding radius for lists of size c, for specific constants c, however, was not known completely. The results of Zyablov and Pinsker [16], stated in our notation, showed that

$$U_c^{\text{const}}(R) \geq H^{-1}(1 - \frac{1}{\lg(c+1)} - R)\ , \tag{1}$$

(this result clearly implies the above-mentioned result $U^{\text{const}}(R) = H^{-1}(1 - R)$) and Elias [4] had shown (again rephrased in our notation) that

$$U_c^{\text{const}}(R) \geq \frac{1}{2}\Big(1 - \sqrt{1 - \frac{2(c-1)}{c} H^{-1}(1 - R)}\Big)\ . \tag{2}$$

Here we improve the bounds significantly and show the following. In particular this answers the main open question posed by Elias [4] on whether the bound (1) can be improved.

Theorem 4 *For each fixed integer $c \geq 1$, and rate $0 < R < 1$, $U_c^{\text{const}}(R) \geq H^{-1}(1 - \frac{1}{c} - R)$.*

We prove this theorem using the probabilistic/greedy method. We stress that our result is only existential and not constructive. Surprisingly our proof does not give a high-probability result; and it is open whether a random code satisfies the above trade-off between rate and list decoding radius with high probability (the result of Equation (1) above is a high-probability result). For general non-linear codes, the bound of Theorem 4 is shown in [4] and in fact holds with high probability for a random code.

All the above results (including ours from Theorem 4) provide lower bounds on $U_c^{\text{const}}(\cdot)$ (except for the simple upper bound $U_c^{\text{const}}(R) \leq U^{\text{poly}}(R) \leq H^{-1}(1 - R)$). The results of Blinovskii [1] imply non-trivial upper bounds on $U_c^{\text{const}}(\cdot)$ for fixed constants c. His results are not easily stated, but they do imply that $U_c^{\text{const}}(R)$ is *strictly less than* $H^{-1}(1-R)$ for every c, R, and thus in order to achieve the Gilbert-Varshamov bound, we *have to* allow an unbounded number of elements in the list.

Theorem 5 [Follows from [1]] *For every $c \geq 1$ and $0 < R < 1$, we have $U_c^{\text{const}}(R) < H^{-1}(1 - R)$.*

Comment on Theorem 4: Suppose we want a binary code family $\mathcal{C}$ with $\mathrm{Rad}(\mathcal{C}, \ell) = (1/2 - \varepsilon)$ for some constant $\varepsilon > 0$ and where ℓ is the constant function $\ell(n) = c \; \forall n$. Then Theorem 4 tells us that such code families with rate $\Omega(\varepsilon^2)$ exist for a list size of $c = O(\varepsilon^{-2})$. We remark that the result of Elias (Equation 2) would give code families of rate $\Omega(\varepsilon^4)$ for a list size $c = O(\varepsilon^{-2})$, while that of Zyablov and Pinsker (Equation 1) would give code families of rate $\Omega(\varepsilon^2)$ but with a much worse list size of $c = 2^{O(\varepsilon^{-2})}$. One can also show using the results of Blinovskii [1] that in order to have such families $\mathcal{C}$ with $\mathrm{rate}(\mathcal{C}) > 0$, we must have $c = \Omega(\varepsilon^{-2})$. In this sense, the result of Theorem 4 is asymptotically the best possible (for example, the $1/c$ "loss" term cannot be improved to $1/c^{1+\gamma}$ for any positive γ).

2.2 List decoding radius vs. Distance of the code

Next we move on to lower bounds on the list decoding radius. As mentioned earlier, it makes sense to study this as a function of the minimum distance of the code. A large minimum distance implies a large list decoding radius by existing combinatorial bounds (see for example [5]), and we want to find the smallest possible list decoding radius for a code of (at least) a certain minimum distance. This motivates the next definition.

Definition 6 (Lower bound on list decoding radius) *For a distance $0 \le \delta \le 1$, and list size $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, the* lower bound on list of ℓ decoding radius *for (binary) codes of distance δ, denoted $L_{\ell,q}(\delta)$, is defined to be*

$$L_\ell(\delta) = \inf_{\mathcal{C} \mid \Delta(\mathcal{C}) \ge \delta} \mathrm{Rad}(\mathcal{C}, \ell).$$

Note that both in the case of the upper bound function U_ℓ and the lower bound function L_ℓ one could allow the arguments, i.e., rate and distance to be functions of n, in which case the supremum would be taken over codes $\mathcal{C}$ that satisfy $\mathrm{rate}(C_i) \ge R(n_i) \cdot n_i$ (or in the case of the lower bound function, we would take the infimum over codes that satisfy $\Delta(C_i) \ge \delta(n_i) \cdot n_i$).

As in the case of the upper bound function, we introduce notation to study the special cases when the list size is a constant or grows as a polynomial.

Definition 7 *For real distance $0 \le \delta \le 1/2$ and constant c, the quantity $L_c^{\mathrm{const}}(\delta)$ is defined to be $L_\ell(\delta)$ where $\ell(n) = c$. The quantity $L_c^{\mathrm{poly}}(\delta)$ is defined to be $\sup_{c_1} L_{\ell_{c_1}}(\delta)$ where $\ell_{c_1}(n) = c_1 n^c$. The quantity $L^{\mathrm{const}}(\delta)$ (resp. $L^{\mathrm{poly}}(\delta)$) will denote the quantity* $\limsup_{c \to \infty}\{L_c^{\mathrm{const}}(\delta)\}$ *(resp.* $\limsup_{c\to\infty}\{L_c^{\mathrm{poly}}(\delta)\}$*).*

Note that we restrict $\delta \le 1/2$ since binary codes with minimum distance $\delta > 1/2$ have only a polynomial number of codewords and are thus not interesting. It is clear that $L_1(\delta) = \delta/2$. It is also easy to see that $L^{\mathrm{poly}}(\delta) \le \delta$ (since there exist codes of minimum distance δ with super-polynomially many codewords at minimum distance.) Thus all lower bounds of interest lie in the range $[\delta/2, \delta]$. The exact values are, however, mostly unknown. The main motivation for our work is the following conjecture.

Conjecture 8 *For every $0 < \delta < 1/2$, $L^{\mathrm{const}}(\delta) = L^{\mathrm{poly}}(\delta) = \frac{1}{2} \cdot (1 - \sqrt{1 - 2\delta})$.*

Evidence in support of the conjecture comes piecemeal. Firstly, it is known that $L^{\mathrm{poly}}(\delta) \ge L^{\mathrm{const}}(\delta) \ge L_2^{\mathrm{const}}(\delta) \ge \frac{1}{2} \cdot \left(1 - \sqrt{1 - 2\delta}\right)$ (see for example [5]). Upper bounds on L^{poly} and L^{const} are not as well studied. Justesen and Hoholdt [9] demonstrate some MDS code families $\mathcal{C}$ of distance δ with $\mathrm{Rad}(\mathcal{C}, c) \le (1 - \sqrt{1 - \delta})$, but this does not apply for codes over any fixed size alphabet, and in particular for binary codes.

The quantity $L^{\mathrm{poly}}(\delta)$ is even less well understood. When δ is either very large (of the form $1/2 - o(1)$) or very small (of the form $o(1)$), there is some evidence confirming this bound. In particular, Dumer et al. [2] construct a family of linear codes $\mathcal{C}$, for any $\varepsilon > 0$, for which $\delta(n) = n^{\varepsilon-1}$ and $L^{\mathrm{poly}}(\delta) \leq \delta/(2-\varepsilon)$ which matches the conjecture above reasonably closely. In unpublished work, Ta-Shma and Zuckerman [13] show that for every $\varepsilon > 0$, $L^{\mathrm{poly}}(\frac{1}{2}(1 - n^{\varepsilon-1/2})) \leq \frac{1}{2}(1 - \frac{1}{3\varepsilon}n^{\varepsilon-1/2})$. This seems to show that the tangent of the curve $L^{\mathrm{poly}}(\delta)$ has infinite slope as $\delta \to 1/2$, which is consistent with the conjecture above (and thus mild evidence in favor of the conjecture). One additional reason for believing in the conjecture is that if the definition of codes is extended to allow non-linear codes, then indeed it is known that the conjecture is true (see for example [5]). All this evidence adds support to the conjecture, however remains far from proving it. In fact till this paper it was not even known if $L_c^{\mathrm{poly}}(\delta) < \delta$. The following theorem resolves this question.

Theorem 9 *For every c and every δ, we have $L_c^{\mathrm{poly}}(\delta) < \delta$.*

Further, for the case $\delta = \frac{1}{2} \cdot (1 - o(1))$, we actually get close to proving the above conjecture. The theorem below follows from Lemma 12 which is stated and proved in Section 3.3.

Theorem 10 *For $\delta = \frac{1}{2}(1 - o(1))$ for some (explicitly given) $o(1)$ function, for every $\varepsilon > 0$, we have $L^{\mathrm{poly}}(\delta) \leq \frac{1}{2}[1 - (1-2\delta)^{1/2+\varepsilon}]$.*

Organization of the Paper. We study the lower bound functions $L^{\mathrm{poly}}(\delta)$ and $L_c^{\mathrm{poly}}(\delta)$ in Section 3 and prove Theorems 9 and 10. In Section 4, we study the function $U_c^{\mathrm{const}}(R)$ and prove Theorem 4.

3 List Decoding Radius and Minimum Distance

We now prove upper bounds on the function $L^{\mathrm{poly}}(\delta)$ claimed in Theorems 9 and 10. We will first prove Theorem 10 which shows that when $\delta = \frac{1}{2} \cdot (1 - o(1))$, one "almost" has a proof of Conjecture 8. A modification of this proof will also yield the proof of Theorem 9. We first review the basic definitions and concepts from (Discrete) Fourier analysis that will be used in our proof.

3.1 Fourier analysis and Group characters

For this section, it will be convenient to represent Boolean values by $\{1, -1\}$ with 1 standing for FALSE and -1 for TRUE. This has the nice feature that XOR just becomes multiplication. Thus a binary code of blocklength m will be a subset of $\{1, -1\}^m$. There are 2^t linear functions $\chi_\alpha : \{0,1\}^t \to \{1, -1\}$ on t-variables, one for each $\alpha \in \{0,1\}^t$. The function χ_α is defined by $\chi_\alpha(x) = (-1)^{\alpha \cdot x} = (-1)^{\sum \alpha_i x_i}$. Fixing some representation of the field $\mathrm{GF}(2^t)$ as elements of $\{0,1\}^t$, the linear functions χ_α are the *additive characters* of the field $\mathrm{GF}(2^t)$, and can also be indexed by elements $\alpha \in \mathrm{GF}(2^t)$. We will do so in the rest of the paper. We also have, for each $y \in \mathrm{GF}(2^t)$, $\sum_\alpha \chi_\alpha(y)$ equals 0 if $y \neq 0$ and 2^t if $y = 0$, where the summation is over all $\alpha \in \mathrm{GF}(2^t)$.

We can define an inner product $\langle f, g \rangle$ for functions $f, g : \mathrm{GF}(2^t) \to \mathbb{R}$ as $\langle f, g \rangle = 2^{-t} \sum_x f(x)g(x)$. The linear functions form an orthonormal basis for the space of real-valued functions on $\mathrm{GF}(2^t)$ with respect to this inner product. Thus every real-valued function on $\mathrm{GF}(2^t)$, and in particular every Boolean function $f : \mathrm{GF}(2^t) \to \{1, -1\}$ can be written in terms of the χ_α's as: $f(x) = \sum_{\alpha \in \mathrm{GF}(2^t)} \hat{f}_\alpha \chi_\alpha(x)$. The coefficient f_α is called the *Fourier coefficient* of f with respect to α and satisfies $\hat{f}_\alpha = \langle f, \chi_\alpha \rangle = 2^{-t} \sum_x f(x)\chi_\alpha(x)$. If we define the distance between functions f, g as $\Delta(f, g) = \Pr_x [f(x) \neq g(x)]$, then $\hat{f}_\alpha = 1 - 2\Delta(f, \chi_\alpha)$. The Fourier coefficients of a Boolean function also satisfy the Plancherel's identity $\sum_\alpha \hat{f}_\alpha^2 = 1$.

Hadamard code: For any integer t, the Hadamard code Had_t of dimension t maps t bits (or equivalently elements of $\mathrm{GF}(2^t)$) into $\{1, -1\}^{2^t}$ as follows: For any $x \in \mathrm{GF}(2^t)$, $\mathrm{Had}_t(x) = \langle \chi_\alpha(x) \rangle_{\alpha \in \mathrm{GF}(2^t)}$.

3.2 Idea behind the Construction

Since our aim is to prove lower bounds on the list decoding radius we must construct codes with large minimum distance with a large number of codewords in a ball of desired radius. The specific codes we construct are obtained by concatenating an outer Reed-Solomon code over a finite field $F = \mathrm{GF}(2^t)$ with the Hadamard code Had_t of blocklength 2^t and dimension t. Thus the messages of this code will be degree ℓ polynomials over $\mathrm{GF}(2^t)$ for some ℓ, and such a polynomial P is mapped into the codeword $\langle \mathrm{Had}_t(P(z_1)), \ldots, \mathrm{Had}_t(P(z_{2^t}))\rangle$ where $z_1, z_2, \ldots, z_{2^t}$ is some enumeration of the elements in $\mathrm{GF}(2^t)$.

It is easy to see that this code has blocklength 2^{2t} and minimum distance $\frac{1}{2}(1-\frac{\ell}{n})2^{2t}$. If $\ell = (1-2\delta)n$, then the relative minimum distance is δ, and for future reference we denote this code by $\text{RS-}\textsc{Had}_t(\delta)$.

To construct the *received word* (which will be the center of the Hamming ball with a lot of codewords), consider the following. Suppose we could pick an appropriate subset S of $\mathrm{GF}(2^t)$ and construct a Boolean function $f : \mathrm{GF}(2^t) \to \{1, -1\}$ that has large Fourier coefficient $\hat{f}_\alpha$ with respect to α for $\alpha \in S$. Let $\mathbf{v} \in \{1,-1\}^{2^t}$ be the 2^t-dimensional vector consisting of the values of f on $\mathrm{GF}(2^t)$. The word $\mathbf{v}^{|F|}$, i.e., $\mathbf{v}$ repeated $|F|$ times will be the "received word" (the center of the Hamming ball which we want to show has several codewords). Since f has large Fourier support on S, $\mathbf{v}^{|F|}$ will have good agreement with all codewords that correspond to messages (polynomials) P that satisfy $P(z_i) \in S$ for many field elements z_i. By picking for the set S a multiplicative subgroup of $\mathrm{GF}(2^t)$ suitable size, we can ensure that there are several such polynomials, and hence several codewords in the concatenated code with good agreement with $\mathbf{v}^{|F|}$.

The main technical component of our construction and analysis is the following Theorem which asserts the existence of Boolean functions f with large support on subgroups S of $\mathrm{GF}(2^t)$. We will defer the proof of the theorem to Section 3.5, and first use it to prove Theorems 10 and 9.

Theorem 11 *There exist infinitely many integers s with the following property: For infinitely many integers t, there exists a multiplicative subgroup S of $\mathrm{GF}(2^t)$ of size s such that the following holds: For every $\beta \neq 0$ in $\mathrm{GF}(2^t)$ there exists a function $f : \mathrm{GF}(2^t) \to \{1,-1\}$ with $\sum_{\alpha \in \beta S} \hat{f}_\alpha \geq \sqrt{\frac{s}{3}}$. (Here βS denotes the coset $\{\beta x : x \in S\}$ of S.)*

Remarks: Our proof of the above theorem in fact gives the following additional features which we make use of in our applications of the theorem.

1. The integers s exists with good density; in particular for any integer $k \geq 4$, there exists an s, with $k \leq s < 3k$, that satisfies the requirements of Theorem 11.

2. We can also add the condition that there exist infinitely many t including one that is at most s, and the theorem still holds.

For any subset $S \subseteq \mathrm{GF}(2^t)$, one can show that $\sum_{\alpha \in S} \hat{f}_\alpha$ is at most $|S|^{1/2}$ using Plancherel's identity and Cauchy-Schwartz, and Theorem 11 shows that we can achieve a sum of $\Omega(|S|^{1/2})$ infinitely often for appropriate multiplicative subgroups S.

3.3 Proof of Theorem 10

We now employ Theorem 11 to prove Theorem 10. We in fact prove the following Lemma which clearly establishes Theorem 10.

Lemma 12 *For every $\varepsilon > 0$, there exist infinitely many t such that the following holds: Let $N = 2^{2t}$. There exists a vector $\mathbf{r} \in \{1,-1\}^N$ and $\delta = \frac{1}{2}(1-(\log N)^{-e})$ for some $e < 1$, such that the number of codewords C of the code $\text{RS-}\textsc{Had}_t(\delta)$ with $\mathrm{distance}(\mathbf{r}, C) \leq \frac{N}{2}(1-(1-2\delta)^{1/2+\varepsilon})$ is at least $N^{\Omega(\log^\varepsilon N)}$.*

Proof: Let s, t be any pair of integers guaranteed by Theorem 11 with $t \leq s$ (we are using one of the remarks following Theorem 11 here). Let S be a multiplicative subgroup of $\mathrm{GF}(2^t)$ of size s and $f : \mathrm{GF}(2^t) \to \{1, -1\}$ a function such that

$$\sum_{\alpha \in S} \hat{f}_\alpha \geq \sqrt{\frac{s}{3}} \,. \tag{3}$$

Let $n = 2^t$, $N = 2^{2t}$ and $p = (n-1)/s$. Then $S \cup \{0\}$ consists of all elements in $\mathrm{GF}(2^t)$ which are p'th powers of some element of $\mathrm{GF}(2^t)$.

We first fix the "received word" $\mathbf{r}$. Let $\mathbf{v} \in \{1, -1\}^n$ be the vector $\langle f(x) \rangle_{x \in \mathrm{GF}(2^t)}$ of all values of f. Then $\mathbf{r} = \mathbf{v}^n$, i.e. the vector $\mathbf{v}$ repeated $n = 2^t$ times, one for each position of the outer Reed-Solomon code.

Let δ be a parameter to be specified later and $\ell = (1-2\delta)n$. Consider the binary code $\mathbf{C} = \text{RS-}\textsc{Had}_t(\delta)$ obtained by concatenating a Reed-Solomon code of dimension $\ell + 1 = (1-2\delta)n + 1$ over $\mathrm{GF}(2^t)$ with Had_t. $\mathbf{C}$ has blocklength N and minimum distance δN. We now want to demonstrate several codewords in $\mathbf{C}$ that are "close" to $\mathbf{r}$. We prove this picking codewords in $\mathbf{C}$ at random from some distribution and showing that the agreement with $\mathbf{r}$ is "large" with good probability.

Let $m = \lfloor \ell/p \rfloor$ and consider a message (degree ℓ polynomial over $\mathrm{GF}(2^t)$) P of $\mathbf{C}$ which is of the form $P(x) = R(x)^p$ for a *random* polynomial R of degree at most m over $\mathrm{GF}(2^t)$. The Reed-Solomon encoding $(b_1, b_2, \ldots, b_n)$ of P satisfies $b_i \in S \cup \{0\}$ for every i, $1 \leq i \leq n$. Moreover, for each i and each $a \in S$, we have $\mathbf{Pr}[b_i = a] = p/n$, and $\mathbf{Pr}[b_i = 0] = 1/n$. Further, the choices of b_i are *pairwise independent*. Now, using Equation (3) and a simple probabilistic analysis one can show that, for large enough N, with probability at least $1/2$ (over the choice of the polynomial R), the encoding of P differs from $\mathbf{r}$ in at most $(\frac{1}{2} - \frac{1}{2\sqrt{4s}})N$ codeword positions.

We thus get at least $\frac{1}{2}n^m$ codewords in a ball of radius $(\frac{1}{2} - \frac{1}{2\sqrt{4s}})N$. We now pick parameters suitably to conclude the result. We have $t = \frac{1}{2}\log N$ and $s \geq t$. Picking $m \leq s^\varepsilon$, we have $(1-2\delta) \simeq m/s = s^{\varepsilon - 1} \geq \log^e N$ for some constant $e < 1$, and we can thus have a minimum distance of $\delta = \frac{1}{2}(1 - (\log N)^{-e})$ for some $e < 1$. Also for ε small enough, $(1-2\delta)^{1/2+\varepsilon} \simeq s^{(\varepsilon-1)(1/2+\varepsilon)} \leq (4s)^{-1/2}$ and thus we have $\Omega(n^m) = N^{\Omega(\log^\varepsilon N)}$ codewords in a Hamming ball of radius $\frac{N}{2}(1 - (1-2\delta)^{1/2+\varepsilon})$. □

3.4 Proof of Theorem 9

We now turn to obtaining upper bounds on $L_c^{\mathrm{poly}}(\delta)$ for a fixed constant c. One way to achieve this would be to pick $m \simeq 2c$ in the above proof, and then pick $s \simeq 2c/(1-2\delta)$ and this would give (roughly) $L_c^{\mathrm{poly}}(\delta) \leq \frac{1}{2}(1 - \left(\frac{1-2\delta}{6c}\right)^{1/2})$. However this upper bound is better than δ only for δ large enough, specifically for $\delta > \frac{1}{2} - \frac{1}{12c}$. We thus have to modify the construction of Lemma 12 in order to prove Theorem 9. We prove the following lemma which will in turn imply Theorem 9. Since our goal was only to establish Theorem 9, we have not attempted to optimize the exact bounds in the lemma below.

Lemma 13 *For every c and every δ, we have*

$$L_c^{\mathrm{poly}}(\delta) \leq \min_{0 \leq \alpha \leq 1/2 - \delta} \Big\{ (\delta + \alpha)(1 - (\frac{\alpha}{12(2c+1)})^{1/2}) \Big\}.$$

Proof of Lemma 13: To prove the claimed upper bound on $L_c^{\mathrm{poly}}(\delta)$, we will closely follow the construction from the proof of Lemma 12. Let $0 < \delta < 1/2$, $0 \leq \alpha \leq (1/2 - \delta)$, and c be given. Define $\alpha' = 2\alpha$ and pick an integer s, $2(2c+1)/\alpha' \leq s < 6(2c+1)/\alpha'$ such that the conditions of Theorem 11 are met (we know such an s exists by the remarks following Theorem 11). Let t be *any* integer for which a subgroup S of $\mathrm{GF}(2^t)$ exists as guaranteed by Theorem 11 (there are once again infinitely many such values of t).

Now we describe the actual construction for a particular δ, α', s, t. Let $n = 2^t$, $N = n^2$ and $p = (n-1)/s$. Let S be a multiplicative subgroup of $\mathrm{GF}(2^t)$ of size s. As in the proof of Lemma 12, the code will again be $\text{RS-}\textsc{Had}_t(\delta)$ (the messages of the code will thus be polynomials over $\mathrm{GF}(2^t)$ of degree at most $\ell = (1-2\delta)n$ and the code has blocklength N). The only change will be in the construction of the received word $\mathbf{r}$. Let $B = (1 - 2\delta - \alpha')n$. We set the first B blocks of $\mathbf{r}$ to be all zeroes. The last $(n - B)$ blocks of $\mathbf{r}$ will be vectors $\mathbf{v}^{(i)} \in \{1, -1\}^{2^t}$, $B < i \le n$, defined as follows. Let $z_1, z_2, \ldots, z_B$ be the B elements of $\mathrm{GF}(2^t)$ that correspond to the first B positions of the Reed-Solomon code. For $B < i \le n$, define $\beta_i = (z_i - z_1) \cdots (z_i - z_B)$, and $f^{(i)} : \mathrm{GF}(2^t) \to \{1, -1\}$ be a function with large Fourier support on the coset $\beta_i S$ of S as guaranteed by Theorem 11; i.e. $\sum_{\alpha \in \beta_i S} \hat{f}^{(i)}_\alpha \ge \sqrt{s/3}$. We set $\mathbf{v}^{(i)}$ to be the table of values of $f^{(i)}$.

Let $m = 2c + 1$. We will consider the messages corresponding to polynomials of the form $P(x) = (x - z_1) \cdots (x - z_B) R(x)^p$ where $z_1, \ldots, z_B$ of $\mathrm{GF}(2^t)$ are the B elements of $\mathrm{GF}(n)$ that correspond to the first B positions of the Reed-Solomon code and R is a random degree m polynomial. Note that $\text{degree}(P) = B + pm = \ell - \alpha' n + \frac{n-1}{s}(2c+1) \le \ell$ since we picked $s \ge 2(2c+1)/\alpha'$. Note that the encoding of every such P agrees with $\mathbf{r}$ in the first B blocks.

Using arguments similar to those in the proof of Lemma 12, one can show that with high probability (say at least $1/2$), the codeword P differs from $\mathbf{r}$ in at most $E = (n - B)(\frac{1}{2} - \frac{1}{2\sqrt{4s}})n$ positions, and thus there are at least $\frac{1}{2}n^m$ codewords within a ball of radius E around $\mathbf{r}$. Since $N = n^2$, $m = 2c + 1$ and $s < 6(2c+1)/\alpha'$, we have $\omega(N^c)$ codewords in a Hamming ball of radius $N(\delta + \alpha'/2)(1 - \sqrt{\frac{\alpha'}{24(2c+1)}})$, and recalling that $\alpha' = 2\alpha$, the claimed result follows. □ *(Lemma 13)*

Proof of Theorem 9: We want to prove $L_c^{\text{poly}}(\delta) < \delta$. Note that when $\delta > \frac{1}{2} - \frac{1}{48(2c+1)}$, setting $\alpha = 1/2 - \delta$ gives

$$L_c^{\text{poly}}(\delta) \le \frac{1}{2}(1 - (\frac{1-2\delta}{24(2c+1)})^{1/2}) < \delta\,.$$

When $\delta \le \frac{1}{2} - \frac{1}{48(2c+1)}$, setting $\alpha = \delta^2/48(2c+1)$ (this is a valid setting since it is less than $1/2 - \delta$), we have $L_c^{\text{poly}}(\delta) \le \delta + \alpha - \delta(\frac{\alpha}{12(2c+1)})^{1/2} < \delta$. Thus we have $L_c^{\text{poly}}(\delta) < \delta$ in either case. □ *(Theorem 9)*

3.5 Proof of Theorem 11

The proof proceeds in several steps. We first state the following lemma which shows that if a subset S of $\mathrm{GF}(2^t)$ satisfies a certain property, then there exists a Boolean function $f : \mathrm{GF}(2^t) \to \{1, -1\}$ such that $\sum \hat{f}_\alpha$ is large when summed over $\alpha \in S$. The proof of the lemma is omitted here for reasons of space and can be found in the full version of the paper.

Lemma 14 *For any integer t, let S be any subset of elements of the field $\mathrm{GF}(2^t)$ such that no four (distinct) elements of S sum up to 0. Then there exists a function $f : \mathrm{GF}(2^t) \to \{1, -1\}$ with $\sum_{\alpha \in S} \hat{f}_\alpha \ge \sqrt{|S|/3}$.*

Given the statement of Lemma 14, we next turn to constructing subgroups of $\mathrm{GF}(2^t)$ with the property that no four (or fewer) distinct elements of the subgroup sum up to 0. To construct such subgroups, we make use of the following simple lemma about the existence of certain kinds of cyclic codes.

Lemma 15 *Let $k \ge 4$ be any integer. Then there exists an integer s in the interval $[k, 3k)$ such that a maximal binary BCH code of blocklength s and minimum distance at least 5 exists.*

Proof: Let s be an integer of the form $2^f - 3$ in the range $[k, 3k)$ (such an integer clearly exists). Let β be the primitive s'th root of unity over $\mathrm{GF}(2)$ and let h be the minimal polynomial of β over $\mathrm{GF}(2)$. Clearly, $h(\beta^{2^i}) = 0$ for all $i \geq 1$, and hence $h(\beta^2) = h(\beta^4) = 0$. Since $\beta^{2^f} = \beta^3$, we also have $h(\beta^3) = 0$. Now the consider the cyclic code C_h of blocklength s with generator polynomial h. It is clearly maximal since h, being the minimal polynomial of β, is irreducible over $\mathrm{GF}(2)$. Also $h(\beta^i) = 0$ for $i = 1, 2, 3, 4$. Using the BCH bound on designed distance (see, for example, Section 6.6 of [14]), this implies that the minimum distance of C_h is at least 5, as desired. □

Lemma 16 *Let $k \geq 4$ be any integer. Then there exists an integer s in the interval $[k, 3k)$ with the following property. For infinitely many integers t, including some integer which is at most s, there exists a multiplicative subgroup S of $\mathrm{GF}(2^t)$ of size s such that no four or fewer distinct elements of S sum up to 0 (in $\mathrm{GF}(2^t)$). Moreover, for any non-zero $\beta \in \mathrm{GF}(2^t)$ this property holds for the coset βS as well.*

Proof: Given k, let $k \leq s < 3k$ be an integer for which there exists a binary BCH code C of blocklength s as guaranteed by Lemma 15 exists. Such a code is generated by an irreducible polynomial h where $h(x)|(x^s - 1)$. Let $t = \mathrm{degree}(h)$; clearly $t \leq s$. Consider the finite field $F = \mathbb{F}_2[X]/(h(X))$ which is isomorphic to $\mathrm{GF}(2^t)$, and consider the subgroup S of size s of F comprising of $\{1, X, X^2, X^3, \ldots, X^{s-1}\}$. The fact that C has distance at least 5 implies that $\sum_{i \in G} X^i$ is not divisible by $h(X)$ for any set G of size at most 4, and thus no four or fewer distinct elements of S sum up to 0 in the field F. This gives us one value of $t \leq s$ for which the conditions of Lemma 16 are met, but it is easy to see that any multiple of t also works, since the same S is also a (multiplicative) subgroup of $\mathrm{GF}(2^{kt})$ for all $k \geq 1$. The claim about the cosets also follows easily, since if $a_1 + a_2 + a_3 + a_4 = 0$ where each $a_i \in \beta S$, then $\beta^{-1}a_1 + \beta^{-1}a_2 + \beta^{-1}a_3 + \beta^{-1}a_4 = 0$ as well, and since $\beta^{-1}a_i \in S$, this contradicts the property of S. □

We now have all the ingredients necessary to easily deduce Theorem 11.

Proof of Theorem 11: Theorem 11 now follows from Lemma 14 and Lemma 16. Note also that the statement of Lemma 16 implies the remarks made after the statement of Theorem 11. □ *(Theorem 11)*

4 List Decoding Radius vs. Rate

We now prove Theorem 4.

Proof of Theorem 4: For each fixed integer $c \geq 1$ and $0 < p < 1/2$, we use the probabilistic method to guarantee the existence of a binary linear code $\mathbf{C}$ of blocklength n, with at most c codewords in any ball of radius $e = pn$, and whose rate is $k = \lfloor (1 - H(p) - 1/c)n \rfloor$, for all large enough n. This clearly implies the lower bound on U_c^{const} claimed in the statement of the Theorem.

The code $\mathbf{C} = \mathcal{C}_k$ will be built iteratively in k steps by randomly picking the k basis vectors in turn. Initially the code $\mathcal{C}_0$ will just consist of the all-zeroes codeword $b_0 = 0^n$. The code $\mathcal{C}_i$, $1 \leq i \leq k$, will be successively built by picking a random (non-zero) basis vector b_i that is linearly independent of $b_1, \ldots, b_{i-1}$, and setting $\mathcal{C}_i = \mathrm{span}(b_1, \ldots, b_i)$. Thus $\mathbf{C} = \mathcal{C}_k$ is an $[n, k]_2$ linear code. We will now analyze the list of c decoding radius of the codes $\mathcal{C}_i$, and the goal is to prove that the list of c decoding radius of $\mathbf{C}$ is at least e.

The key to analyzing the list of c decoding radius is the following potential function $S_{\mathcal{C}}$ defined for a code $\mathcal{C}$ of blocklength n:

$$S_{\mathcal{C}} = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 2^{\frac{n}{c} \cdot |B(x,e) \cap \mathcal{C}|} . \tag{4}$$

For notational convenience, we denote $S_{\mathcal{C}_i}$ be S_i. Also denote by T_x^i the quantity $|B(x, e) \cap \mathcal{C}_i|$, so that $S_i = 2^{-n} \sum_x 2^{T_x^i}$.

Let $B = |B(0,e)|$; then $B \leq 2^{H(p)n}$ where $H(p)$ is the binary entropy function of p. Clearly

$$S_0 = 1 - B/2^n + B2^{n/c}/2^n \leq 1 + 2^{n\left(H(p)-1+1/c\right)} . \tag{5}$$

Now once $\mathcal{C}_i$ has been picked with the potential function S_i taking on some value, say $\hat{S}_i$, the potential function S_{i+1} for $\mathcal{C}_{i+1} = \mathrm{span}(\mathcal{C}_i \cup \{b_{i+1}\})$ is a random variable depending upon the choice of b_{i+1}. We consider the expectation $\mathbf{E}[S_{i+1}|S_i = \hat{S}_i]$ taken over the random choice of b_{i+1} chosen uniformly from outside $\mathrm{span}(b_1,\ldots,b_i)$. It is not difficult to show that (see the full version of the paper for details), $\mathbf{E}[S_{i+1}|S_i = \hat{S}_i] = \hat{S}_i^2$ if the expectation is taken over all choices of b_{i+1} in $\{0,1\}^n$. We now use the simple fact that the expectation of a positive random variable taken over b_{i+1} chosen randomly from outside $\mathrm{span}(b_1,\ldots,b_i)$ is at most $(1-2^{i-n})^{-1}$ times the expectation taken over b_{i+1} chosen uniformly at random from $\{0,1\}^n$. We thus get

$$\mathbf{E}[S_{i+1}|S_i = \hat{S}_i] \leq \frac{\hat{S}_i^2}{(1-2^{i-n})} \tag{6}$$

Applying (6) repeatedly for $i = 0,1,\ldots,k-1$ and using the value of S_0 from Equation (5), we can show, after some calculations, that there exists an $[n,k]$ binary linear code $\mathbf{C}$ with

$$S_{\mathbf{C}} = S_k \leq 2(1 + 2 \cdot 2^{k+(H(p)-1+1/c)n}) \leq 6 \tag{7}$$

(the last inequality follows since $k = \lfloor (1-H(p)-1/c)n \rfloor$). By the definition of the potential S_k in Equation (4), this implies that $2^{n/c \cdot |B(x,e)\cap \mathbf{C}|} \leq 6 \cdot 2^n < 2^{n+3}$, or $|B(x,e) \cap \mathbf{C}| < (1+\frac{3}{n})c$ for every $x \in \{0,1\}^n$. If $n > 3c$, this implies $|B(x,e)\cap\mathbf{C}| < c+1$ for every x, implying that the list of c decoding radius of $\mathbf{C}$ is at least e, as desired. □ *(Theorem 4)*

Remark: It turns out that the statement of Theorem 4 can be adapted to guarantee the existence of certain (linear) codes which serve as excellent (for purposes of list decoding) inner codes in a concatenation scheme with an outer Reed-Solomon code. This in turn gives an *efficiently constructible* family of binary linear codes of rate ε^4 which can be efficiently list decoded from up to a $(\frac{1}{2} - O(\varepsilon))$ fraction of errors. This improves upon the results claimed in [7] (the best rate achieved by [7] for such families of codes was ε^6). Details of this construction will appear in the full version of this paper.

References

[1] V. M. Blinovskii. Bounds for codes in the case of list decoding of finite volume. *Prob. Information Transmission*, **22** (1), pp. 11-25 (in Russian), 1986; pp. 7-19 (in English), 1986.

[2] I. Dumer, D. Micciancio and M. Sudan. Hardness of approximating the minimum distance. *Proc. of STOC 1999*.

[3] P. Elias. List decoding for noisy channels. *Wescon Convention Record*, Part 2, Institute of Radio Engineers (now IEEE), pp. 94-104, 1957.

[4] P. Elias. Error-correcting codes for List decoding. *IEEE Trans. Info. Theory*, **37** (1), pp. 5-12, 1991.

[5] O. Goldreich, R. Rubinfeld and M. Sudan. Learning polynomials with queries: the highly noisy case. *Proc. of FOCS 95*.

[6] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and Algebraic-geometric codes. *IEEE Trans. on Information Theory*, 45 (1999), pp. 1757-1767. Preliminary version appeared in *Proc. of FOCS'98*.

[7] V. Guruswami and M. Sudan. List decoding algorithms for certain concatenated codes. *Proc. of STOC 2000*.

[8] G. H. Hardy, J. E. Littlewood, G. Pólya. *Inequalities*, 2nd Edition, Cambridge University Press, 1952.

[9] J. Justesen and T. Hoholdt. Bounds on list decoding of MDS codes. Preprint, 1999.

[10] C. E. Shannon, R. G. Gallager and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. *Information and Control*, **10**, pp. 65-103 (Part I), pp. 522-552 (Part II), 1967.

[11] M. A. Shokrollahi and H. Wasserman. List decoding of algebraic-geometric codes. *IEEE Trans. on Information Theory*, Vol. 45, No. 2, March 1999, pp. 432-437.

[12] M. Sudan. Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180-193, March 1997.

[13] A. Ta-Shma and D. Zuckerman. Personal Communication, April 1999.

[14] J. H. van Lint. *Introduction to Coding Theory*, Graduate Texts in Mathematics **86**, (Third Edition) Springer-Verlag, Berlin, 1999.

[15] J. M. Wozencraft. List Decoding. *Quarterly Progress Report*, Research Laboratory of Electronics, MIT, Vol. 48 (1958), pp. 90-95.

[16] V. V. Zyablov and M. S. Pinsker. List cascade decoding. In *Prob. Information Transmission*, **17** (4), pp. 29-34 (in Russian), 1981; pp. 236-240 (in English), 1982.