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. Specifically, on the positive side, we show that there exist codes of rate R and block length n that have at most c codewords in every Hamming ball of radius $H^{-1}(1-R-1/c)\cdot n$. This answers the main open question from the work of Elias [8]. This result also has consequences for the construction of concatenated codes of good rate that are list decodable from a large fraction of errors, improving previous results of [13] in this vein. Specifically, for every $\varepsilon > 0$, we present a polynomial time constructible asymptotically good family of binary codes of rate $\Omega(\varepsilon^4)$ that can be list decoded in polynomial time from up to a fraction $(1/2-\varepsilon)$ of errors, using lists of size $O(\varepsilon^{-2})$.**

On the negative side, we show that for every δ and c, there exists $\tau < \delta$, $c_1 > 0$ and an infinite family of linear codes $\{\mathcal{C}_i\}_i$ such that if n_i denotes the block length of $\mathcal{C}_i$, then $\mathcal{C}_i$ has minimum distance at least $\delta \cdot n_i$ and contains more than $c_1 \cdot n_i^c$ codewords in some Hamming ball of radius $\tau \cdot n_i$. While this result is still far from known bounds on the list-decodability of linear codes, it is the first to bound the "radius for list-decodability by a polynomial-sized list" away from the minimum distance of the code.

Keywords— **Error-correcting codes, List decoding, Concatenated codes, Reed-Solomon code.**

I. Introduction

LIST decoding was introduced independently by Elias [7] and Wozencraft [24] as a relaxation of the "classical" notion of decoding by allowing the decoder to output a *list* of codewords as answers. The decoding is considered successful as long as the correct message is included in the list. Early work by Elias and Wozencraft [7], [24] analyzed the probability of error in this model and used random coding arguments to explore the average decoding error probability of block codes at low rates for the binary symmetric channel. List decoding was also used by Shannon, Gallager and Berlekamp [17] in exploring low rate average error bounds for general discrete memoryless channels, and Ahlswede [1] showed that it enables one to determine capacity of a wide class of communication channels.

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 [25], Blinovsky [3], [4], and Elias [8] applied in this setting. (The paper by Elias [8] also gives a very good summary of the prior work and history.) The basic question raised in this setting was: How many errors could still be recovered from, with lists of small size? Two basic parameters 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 [19], [12], [18], [13]. These algorithms decode with polynomial sized lists (and sometimes with constant sized lists) for much more than half the minimum distance of the code, and investigations of the tightness of the algorithms have led Høholdt and Justesen [16] to re-initiate the investigation of the combinatorial bounds on list decoding.

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.

Our results also has consequences for the construction of concatenated codes of good rate that are list decodable from a large fraction of errors, improving previous results of [13] in this vein. Specifically, for every $\varepsilon > 0$, we present a polynomial time constructible asymptotically good family of binary codes of rate $\Omega(\varepsilon^4)$ that can be list decoded in polynomial time from up to a fraction $(1/2-\varepsilon)$ of errors, using lists of size $O(\varepsilon^{-2})$.

II. 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 $\mathbb{F}_2^n$. We refer to n as the blocklength of the code and to k as the dimension of the code. Unless explicitly mentioned otherwise, we will only be interested in linear codes in this paper and will moreover restrict ourselves to the binary case (when $q = 2$).

For two strings x, y of length n over an arbitrary alphabet Σ, let $\Delta(x,y)$ denote the Hamming distance between them, i.e., the number of coordinates where x and y differ. Denote by $\delta(x,y) = \frac{\Delta(x,y)}{n}$ be the relative (fractional) distance between x and y. The minimum distance of a code C, denoted $\mathrm{dist}(C)$, is the quantity $\min_{x,y\in C, x\neq y}\{\Delta(x,y)\}$. The relative distance of the code C, denoted $\delta(C)$, is analogously defined.

A preliminary version of this paper appears in the *Proceedings of the Annual Allerton Conference on Communication, Control and Computing*, Monticello, Illinois, October 2000, pp. 603-612.

Venkatesan Guruswami's address is University of California at Berkeley, Computer Science Division, Berkeley, CA 94720. Email: `venkat@lcs.mit.edu`. The work was done while the author was at MIT and was supported in part by an IBM Graduate Fellowship and NSF CCR-9875511.

Johan Håstad's address is 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 and NSF grant CCR-9987077.

Madhu Sudan's address is Laboratory for Computer Science, 200 Technology Square, Cambridge, MA 02139, USA. Email: `madhu@mit.edu`. Supported in part by a Sloan Foundation Fellowship, NSF Career Award CCR-9875511, NSF Grant CCR-9912342, and NTT Award MIT2001-04.

David Zuckerman's address is Department of Computer Science, University of Texas, Austin, TX 78712, USA. Email: `diz@cs.utexas.edu`. Most of this work was done while this author was 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.

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

$$\mathrm{rate}(\mathcal{C}) = \liminf_i \left\{ \frac{k_i}{n_i} \right\} .$$

We define the *(relative) distance* of an infinite family of codes $\mathcal{C}$ to be

$$\Delta(\mathcal{C}) = \liminf_i \left\{ \frac{\mathrm{dist}(C_i)}{n_i} \right\} .$$

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., $B(x,r) = \{y \in \mathbb{F}_2^n | \Delta(x,y) \le r\}$. For integers e, ℓ, a code $C \subseteq \mathbb{F}_2^n$ is said to be (e,ℓ)*-list decodable* if every ball of radius e has at most ℓ codewords, i.e. $\forall\, x \in \mathbb{F}_2^n$, $|B(x,e) \cap C| \le \ell$.

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 the *maximum* value of e for which C is (e,ℓ)-list decodable.

Definition 2: (List Decoding Radius for code and function families) For an infinite family of codes $\mathcal{C}$ and a function $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, define the *list of ℓ decoding radius* of $\mathcal{C}$, denoted $\mathrm{Rad}(\mathcal{C},\ell)$, to be

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

For an infinite family of codes $\mathcal{C}$ and a *family* of integer-valued functions $\mathcal{F}$, the list decoding radius of $\mathcal{C}$ w.r.t $\mathcal{F}$, also denoted $\mathrm{Rad}(\mathcal{C},\mathcal{F})$ by abuse of notation, is defined as

$$\mathrm{Rad}(\mathcal{C},\mathcal{F}) = \sup_{\ell \in \mathcal{F}} \mathrm{Rad}(\mathcal{C},\ell)$$

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 a polynomial 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, the linear code that is spanned by a small number of standard basis vectors has small rate, but the entire code is contained in a small ball around the all zeroes 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.

A. List decoding radius vs. Rate of the code

Definition 3 (Upper bound on list decoding radius) For real rate $0 \le R \le 1$ and list size $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, the *upper bound on list of ℓ decoding radius* for (binary) codes of rate R, denoted $U_\ell(R)$, is defined to be

$$U_\ell(R) = \sup_{\mathcal{C} \,|\, \mathrm{rate}(\mathcal{C}) \ge R} \mathrm{Rad}(\mathcal{C},\ell).$$

Similarly, for a family of integer-valued functions $\mathcal{F}$, one defines the quantity

$$U_{\mathcal{F}}(R) = \sup_{\ell \in \mathcal{F}} U_\ell(R) \ .$$

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 *all* codes that have at least a certain rate. The case where the list size function is a constant, or growing polynomially is of special interest to us, and we consider the following definitions.

Definition 4: For real rate $0 \le R \le 1$ and constant c, the quantity $U_c^{\mathrm{const}}(R)$ is defined to be $U_\ell(R)$ where $\ell(n) = c$. The quantity $U_c^{\mathrm{poly}}(R)$ is defined to be $U_{\mathcal{F}_c}(R)$ where $\mathcal{F}_c$ is the family of functions $\{\ell_{c_1} : \mathbb{Z}^+ \to \mathbb{Z}^+$ where $\ell_{c_1}(n) = c_1 n^c\}$. The quantity $U^{\mathrm{const}}(R)$ (resp. $U^{\mathrm{poly}}(R)$) will denote the quantity $\limsup_{c\to\infty}\{U_c^{\mathrm{const}}(R)\}$ (resp. $\limsup_{c\to\infty}\{U_c^{\mathrm{poly}}(R)\}$).

Thus the quantities $U^{\mathrm{const}}(R)$ and $U^{\mathrm{poly}}(R)$ 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. The first to pin this quantity down were Zyablov and Pinsker [25]. Zyablov and Pinsker showed that

$$U^{\mathrm{const}}(R) = U^{\mathrm{poly}}(R) = H^{-1}(1-R).$$

Here $H(\cdot)$ is the binary entropy function and $H^{-1}(\cdot)$ is its inverse. Specifically,

$$H(x) = -x \lg x - (1-x)\lg(1-x)$$

where $\lg x$ denotes the logarithm of x to base 2. Further, for $0 \le y \le 1$, $H^{-1}(y)$ denotes the unique z in the range $0 \le z \le 1/2$ such that $H(z) = y$.

The behavior of the upper bound on list decoding radius for lists of size c, for specific constants c, however, was not known completely. This quantity has been investigated significantly in [25], [3], [4], [8], [23], [5] and below we attempt to describe their results and how it motivates our study. We start by noting that $U_c^{\mathrm{const}}(R)$ is monotonic in c, and is thus always at least $H^{-1}(1-R)/2$ which is the Gilbert-Varshamov bound. The results of Zyablov and Pinsker [25], stated in our notation, showed that

$$U_c^{\mathrm{const}}(R) \ge H^{-1}\left(1 - \frac{1}{\lg(c+1)} - R\right) , \tag{1}$$

(this result implies the above-mentioned result $U^{\mathrm{const}}(R) = H^{-1}(1-R)$). The dependence on c above is weaker than what what one can hope for and so the question merited further study. Blinovsky [3] (see also [4]) initiated a systematic study of this quantity for specific choices of c. His focus however was on small values of c and the lower bounds in his result were obtained using non-linear codes. In more recent work [5] shows how the techniques from his prior work may be used to get lower bounds on $U_c^{\mathrm{const}}(R)$ for linear codes as well. Other researchers to focus on $U_c^{\text{const}}(R)$ for small c include Wei and Feng [23]. The results of [3], [4], [5], [23] have a complex dependence on c and so it is hard to extract the asymptotic behavior of $U_c^{\text{const}}(R)$ as a function of c. The only other result with a nice asymptotic relationship between $U_c^{\text{const}}(R)$ and R and c is that of Elias [8] who shows:

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

The two results with analytic forms, specifically (1) and (2), are incomparable to one another. Note that we are interested in relating three parameters: the rate R, the list-size c, and the list-decoding radius $U_c^{\text{const}}(R)$. The bound (2) has a better dependence on the list-size, but a weaker dependence on the rate than the bound (1). A setting which brings this incomparability out very well and also motivates our result (Theorem 5 below) is the following. Consider binary linear codes which have a list-of-c decoding radius $(1/2-\varepsilon)$ for some constant c (that may depend on ε). The bound (1) guarantees the existence of such codes of rate $\Omega(\varepsilon^2)$ with a list size $c = 2^{O(\varepsilon^{-2})}$. While the rate is good (in fact, optimal up to constant factors), the list size is very high. On the other hand, the bound (2) guarantees the existence of such codes of rate $\Omega(\varepsilon^4)$ with a list size $c = O(1/\varepsilon^2)$. Here we strengthen the bounds and show the following result which, for the case for a list decoding radius of $(1/2-\varepsilon)$, combines the optimal rate $\Omega(\varepsilon^2)$ with a list size of $O(1/\varepsilon^2)$. In particular, our result answers the main open question posed by Elias [8] on whether the bound (1), specifically its dependence on the list size c, can be improved.

Theorem 5: For each fixed integer $c \geq 1$, and rate $0 < R < 1$, $U_c^{\text{const}}(R) \geq H^{-1}\big(1 - R - \frac{1}{c}\big)$.

To see why this is the right form for the bound $U_c^{\text{const}}(R)$, we survey some of the known upper bounds on this quantity.

A.1 Upper bounds on $U_c^{\text{const}}(R)$

All the above results (including ours from Theorem 5) provide lower bounds on $U_c^{\text{const}}(\cdot)$ (except for the simple upper bound $U_c^{\text{const}}(R) \leq U^{\text{poly}}(R) \leq H^{-1}(1-R)$). Blinovsky [3] also gave non-trivial upper bounds on $U_c^{\text{const}}(\cdot)$ for fixed constants c. Specifically, he obtains the following result:

$$U_c^{\text{const}}(R) \leq \lambda - \frac{c'+2}{c'+1}\binom{2c'}{c'}\frac{(\lambda(1-\lambda))^{c'+1}}{(c'+2) - 2(2c'+1)\lambda(1-\lambda)} , \qquad (3)$$

where $c' = \lceil c/2 \rceil$ and $\lambda = H^{-1}(1-R)$. (For the special case of $c = 2$, the exact upper bound was later improved in [2].) The above bound applies to non-linear codes as well. While this form of the result is hard to parse, it does imply the following theorem:

Theorem 6: [Follows from [3]] For every $c \geq 1$ and $0 < R < 1$, we have $U_c^{\text{const}}(R) < H^{-1}(1-R)$.

A careful interpretation of the bound (3) above gives a hint that Theorem 5 has the right behavior as a function of c. To get this perspective, let us again focus on the case of a family of binary codes $\mathcal{C}$ with $\text{Rad}(\mathcal{C}, \ell) \geq (1/2-\varepsilon)$ for some constant $\varepsilon > 0$ and where ℓ is the constant function $\ell(n) = c \; \forall n$. Then Theorem 5 tells us that such code families with rate $\Omega(\varepsilon^2)$ exist for a list size of $c = O(\varepsilon^{-2})$. On the other hand, the bound (3) implies that in order to have $\text{rate}(\mathcal{C}) > 0$, we must have $c = \Omega(\varepsilon^{-2})$. Indeed if we want $\text{Rad}(\mathcal{C}, c) \geq 1/2 - \varepsilon$, then Equation (3) implies $\lambda \geq (1/2-\varepsilon)$ and thus $\lambda(1-\lambda) \geq 1/4 - \varepsilon^2$. Therefore the second term in the right hand side of Equation (3) is at least

$$\Omega\Big(\frac{(1-4\varepsilon^2)^{c'+1}}{\sqrt{c'}(2+4c'\varepsilon^2)}\Big)$$

using Stirling's approximation $\binom{2c'}{c'} = \Theta(\frac{4^{c'}}{\sqrt{c'}})$. On the other hand, this term must be at most $O(\varepsilon)$, since we want $U_c^{\text{const}}(R) \geq 1/2 - \varepsilon$. Together these facts imply that $c' = \Omega(\varepsilon^{-2})$, as desired.

In this sense, the result of Theorem 5 is (nearly) the best possible, and in particular the $1/c$ loss term in the bound for $U_c^{\text{const}}(R)$ cannot be improved asymptotically (for instance, it cannot be improved to $1/c^{1+\gamma}$ for a positive γ). In fact, since the upper bound of Equation (3) holds even for general codes, Theorem 5 cannot be improved substantially even if one allows general, non-linear codes.

We remark that an account of the results discussed above in a slightly different notation which studies the rate as a function of list decoding radius (instead of studying the list decoding radius as a function of the rate) appears in [10, Chap. 5]. The presentation there also gives more detailed descriptions of the various results in the literature and their interconnections.

B. *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 [9]), 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 7 (Lower bound on list decoding radius) For a distance $0 \leq \delta \leq 1$, and list size $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, the *lower bound on list-of-ℓ decoding radius* for (binary) codes of relative distance δ, denoted $L_\ell(\delta)$, is defined to be

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

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 $\dim(C_i) \geq R(n_i) \cdot n_i$ (or in the case of the lower bound function, we would take the infimum over codes that satisfy $\text{dist}(C_i) \geq \delta(n_i) \cdot n_i$).

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

Note that we restrict $\delta < 1/2$ since binary codes with relative distance $\delta \geq 1/2$ have at most a linear number of codewords and are thus not very interesting. It is clear that $L_1(\delta) = \delta/2$. It is also easy to see that $L^{\rm poly}(\delta) \le \delta$ (since there exist codes of relative distance δ with super-polynomially many codewords in ball of radius close to the 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 9: For every $0 < \delta < 1/2$, $L^{\rm const}(\delta) = L^{\rm poly}(\delta) = \frac{1}{2} \cdot \left(1 - \sqrt{1-2\delta}\right)$.

Evidence in support of the conjecture comes piecemeal. Firstly, it is known that

$$L^{\rm poly}(\delta) \ge L_1^{\rm poly}(\delta) \ge \frac{1}{2} \cdot \left(1 - \sqrt{1-2\delta}\right)$$

and

$$L_c^{\rm const}(\delta) \ge \frac{1}{2} \cdot \left(1 - \sqrt{1-2\delta+2\delta/c}\right)$$

(see, for example, [9], [14] for a proof of these facts). Upper bounds on $L^{\rm poly}$ and $L^{\rm const}$ are not as well studied. Justesen and Høholdt [16] demonstrate some MDS code families $\mathcal{C}$ of distance δ with $\mathrm{Rad}(\mathcal{C}, c) \le (1 - \sqrt{1-\delta})$ for every constant c for certain values of δ, but this does not apply for codes over any fixed size alphabet, and in particular for binary codes.

The quantity $L^{\rm poly}(\delta)$ 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. [6] construct a family of linear codes $\mathcal{C}$, for any $\varepsilon > 0$, for which $\delta(n) = n^{\varepsilon-1}$ and $L^{\rm poly}(\delta) \le \delta/(2-\varepsilon)$ which matches the conjecture above reasonably closely. We give a simple probabilistic argument to show the following:

Theorem 10: For every $\varepsilon > 0$, there exists an infinite family of binary codes $\mathcal{C}$ and a function $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$ that grows faster than any polynomial such that every member of $C \in \mathcal{C}$ with block length n satisfies

$$\frac{(n/2 - \Delta(C))}{(n/2 - \mathrm{radius}(C, \ell(n)))} \le 3\varepsilon.$$

This seems to show that the tangent of the curve $L^{\rm poly}(\delta)$ 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 [9]). All this evidence adds support to the conjecture, however remains far from proving it. In fact until this paper it was not even known if $L_c^{\rm poly}(\delta) < \delta$. The following theorem resolves this question.

Theorem 11: For every integer $c \ge 1$ and every δ, $0 < \delta < 1/2$, we have $L_c^{\rm poly}(\delta) < \delta$.

Further, for the case $\delta = \frac{1}{2} \cdot (1 - o(1))$, we actually get close to proving the above conjecture. This is done in the theorem below which informally states that if

$$\delta(n) = \frac{1}{2}\left(1 - \Theta((\log n)^{\varepsilon-1})\right),$$

then

$$L^{\rm poly}(\delta) \le \frac{1}{2}[1 - (1-2\delta)^{1/2+\varepsilon}],$$

for arbitrarily small ε. (Of course, the above does not make sense formally since $L^{\rm poly}(\delta)$ was defined as a limit of a series and not a function of n. The following theorem makes the assertion formally, in slightly more cumbersome detail.) The theorem below follows from Lemma 14 which is stated and proved in Section III-C.

Theorem 12: For every ε, $0 < \varepsilon < 1/2$, for some $\delta : \mathbb{Z}^+ \to \mathbb{Z}^+$ satisfying $\delta(n) = \frac{1}{2}\left(1 - \Theta((\log n)^{\varepsilon-1})\right)$ and some super-polynomial function $\ell : \mathbb{Z}^+ \to \mathbb{Z}^+$, there exists an infinite family of codes $\mathcal{C}$ such that for every $C \in \mathcal{C}$ of block length n, the relative minimum distance of C is at least $\delta(n)$ and the list of $\ell(n)$ decoding radius of C is at most $\frac{1}{2}[1 - (1-2\delta)^{1/2+\varepsilon}]$.

In a recent result, Guruswami [11] has made significant progress towards resolving Conjecture 9 — he resolves this conjecture assuming a well-known number-theoretic conjecture. We discuss this result further in Section VI.

Remark: For codes over an alphabet of size q for large enough q, it turns out that $L^{\rm poly}(q, \delta) < \delta$ for certain values of δ can be easily deduced from existing results on codes that beat the Gilbert-Varshamov bound (here $L^{\rm poly}(q, \delta)$ denotes the quantity analogous to $L^{\rm poly}(\delta)$ for the case of q-ary codes). Indeed, it is easy to show that for any code that lies above the GV bound, the expected number of codewords at a Hamming distance of at most d from a random received word, where d is the minimum distance of the code, is exponential. Since q-ary codes that beat the GV bound are known for all square prime powers $q \ge 49$ (specifically certain algebraic-geometric codes achieve this [21]), it follows that for certain $q \ge 49$ and certain values of δ, we indeed have $L^{\rm poly}(q, \delta) < \delta$. However, our focus is on binary codes, and since the GV bound is the best current asymptotic trade-off between rate and distance known for binary codes, the above approach does not give anything for binary codes.

C. Organization of the Paper

We study the lower bound functions $L^{\rm poly}(\delta)$ and $L_c^{\rm poly}(\delta)$ in Section III and prove Theorems 10, 11, and 12. In Section IV, we study the function $U_c^{\rm const}(R)$ and prove Theorem 5. We then prove an adaptation of Theorem 5 (Lemma 22) in Section V, and then use it to construct binary linear codes with very high (algorithmic) list decodability.

III. List decoding Radius and Minimum Distance

We now prove upper bounds on the function $L^{\rm poly}(\delta)$ claimed in Theorems 11 and 12. We will first prove Theorem 12 which shows that when $\delta = \frac{1}{2} \cdot (1 - o(1))$, one "almost" has a proof of Conjecture 9. A modification of this proof will also yield the proof of Theorem 11. We first review the basic definitions and concepts from (Discrete) Fourier analysis that will be used in some of our proofs.

A. 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 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 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

$$\langle f, g\rangle = 2^{-t} \sum_x f(x) g(x).$$

We call this inner product the normalized inner product, in contrast to the unnormalized inner product $\sum_x f(x)g(x)$. The functions χ_α form an orthonormal basis for the space of real-valued functions on $\mathrm{GF}(2^t)$ with respect to the normalized 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:

$$f(x) = \sum_{\alpha \in \mathrm{GF}(2^t)} \hat{f}_\alpha \chi_\alpha(x)\ . \qquad (4)$$

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

$$\Delta(f, g) = \mathop{\mathbf{Pr}}_x \big[f(x) \neq g(x)\big],$$

then

$$\hat{f}_\alpha = 1 - 2\Delta(f, \chi_\alpha).$$

The Fourier coefficients of a Boolean function also satisfy 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 $x \in \mathrm{GF}(2^t)$, $\mathrm{Had}_t(x) = \langle \chi_\alpha(x)\rangle_{\alpha \in \mathrm{GF}(2^t)}$.

B. 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 extended 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 $\mathrm{GF}(2^t)$.

Let $n = 2^t$. It is easy to see that this code has blocklength 2^{2t} and minimum distance $\frac{1}{2}\big(1 - \frac{\ell}{n}\big)2^{2t}$. If $\ell = (1-2\delta)n$, then the relative minimum distance is δ, and for future reference we denote this code by 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 $\mathrm{GF}(2^t)$ of 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 III-E, and first use it to prove Theorems 12 and 11.

Theorem 13: 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

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

Here $\beta \cdot S$ 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 13.
2. We can also add the condition that there exist infinitely many t including one that lies in the range $s/2 \leq t \leq s$, 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 13 shows that we can achieve a sum of $\Omega(|S|^{1/2})$ infinitely often for appropriate multiplicative subgroups S.

C. Proof of Theorem 12

We now employ Theorem 13 to prove Theorem 12. We in fact prove the following Lemma which clearly establishes Theorem 12.

Lemma 14: For every ε, $0 < \varepsilon < 1/2$, there exist infinitely many integers t such that the following holds: Let $N = 2^{2t}$. There exists a vector $\mathbf{r} \in \{1,-1\}^N$ and $\delta = \frac{1}{2}\big(1 - \Theta((\log N)^{\varepsilon-1})\big)$, such that the number of codewords C of the code RS-$\textsc{Had}_t(\delta)$ with

$$\Delta(\mathbf{r}, C) \leq \frac{N}{2}\big(1 - (1-2\delta)^{1/2+\varepsilon}\big)$$

is at least $N^{\Omega(\log^{\varepsilon} N)}$.

Proof: Let s, t be any pair of integers guaranteed by Theorem 13 with $t \le s \le 2t$ (we are using one of the remarks following Theorem 13 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

$$\sum_{\alpha \in S} \hat{f}_\alpha \ge \sqrt{\frac{s}{3}}\,. \tag{5}$$

Let $n = 2^t$, $N = 2^{2t}$ and $p = (n-1)/s$. Note that $s = \Theta(\log N)$ since we have $t \le s \le 2t$. 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 $\mathbf{C} = \text{RS-H}\textsc{ad}_t(\delta)$ obtained by concatenating an extended 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, $1 \le i \le n$. It is easy to see that for each i and each $a \in S$, we have $\mathbf{Pr}[b_i = a] = p/n$, and $\mathbf{Pr}[b_i = 0] = 1/n$. Moreover, the choices of b_i are *pairwise independent*.

Now, by definition of the Fourier coefficient, for each i, the Hadamard codeword $\mathrm{Had}_t(b_i)$ and the vector $\mathbf{v}$ we constructed above have an unnormalized inner product equal to $n \cdot \hat{f}_{b_i}$ (or equivalently, agree on a fraction $\frac{1+\hat{f}_{b_i}}{2}$ of positions). For any i, $1 \le i \le n$, the expected value of $\hat{f}_{b_i}$ satisfies

$$\begin{aligned} \frac{p}{n}\sum_{\alpha \in S} \hat{f}_\alpha + \frac{1}{n}\hat{f}_0 &\ge \frac{(n-1)}{ns}\sum_{\alpha \in S} \hat{f}_\alpha - \frac{1}{n} \ge \frac{1}{s}\sum_{\alpha \in S} \hat{f}_\alpha - \frac{2}{n} \\ &\ge \frac{1}{\sqrt{3s}} - \frac{2}{n}\,, \end{aligned} \tag{6}$$

where the last inequality follows from Equation (5). Let X denote the random variable which is the unnormalized inner product of the codeword (encoding the message $R(x)^p$ for a random polynomial R of degree at most m) with the received vector $\mathbf{r} = \mathbf{v}^n$. By linearity of expectation and using (6), we have

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[n\hat{f}_{b_i}] \ge \frac{N}{\sqrt{3s}} - 2\sqrt{N} \ge \frac{1.1N}{\sqrt{4s}} \tag{7}$$

for large enough N (since $s = \Theta(\log N)$). Now, for each i, $1 \le i \le n$,

$$\mathbf{E}[\hat{f}_{b_i}^2] \le \frac{p}{n}\sum_{\alpha \in S \cup \{0\}} \hat{f}_\alpha^2 \le \frac{1}{s}\,.$$

Since the b_i's are evaluations of the polynomial $R(x)^p$ at the n field elements for a random R, they are pairwise independent. Thus the variance of the random variable X is bounded from above by

$$\mathbf{E}[X^2] = \sum_{i=1}^{n} \mathbf{E}[(n\hat{f}_{b_i})^2] \le \frac{N^{3/2}}{s}\,. \tag{8}$$

We now use Chebyshev's inequality to prove that the inner product X is greater than $N/\sqrt{4s}$ with probability at least $1/2$. Indeed

$$\begin{aligned} \mathbf{Pr}[X \le \frac{N}{\sqrt{4s}}] &\le \mathbf{Pr}[X - \mathbf{E}[X] \le -\frac{N}{10\sqrt{4s}}] \\ &\le \mathbf{Pr}[|X - \mathbf{E}[X]| \ge \frac{N}{10\sqrt{4s}}] \\ &\le \frac{400s\,\mathbf{E}[X^2]}{N^2} \le \frac{400}{\sqrt{N}} \\ &< \frac{1}{2} \qquad \text{(for large enough } N\text{)}, \end{aligned}$$

where we have used the lower bound on $\mathbf{E}[X]$ from Equation (7) and the upper bound on $\mathbf{E}[X^2]$ from Equation (8).

Hence the codewords encoding at least $\frac{1}{2} \cdot n^m$ of the polynomials of the form $R(x)^p$ where R is a polynomial of degree at most m, differ from $\mathbf{r}$ in at most $(\frac{1}{2} - \frac{1}{2\sqrt{4s}})N$ codeword positions.

We now pick parameters (namely m, δ) suitably to conclude the result. Recall that $s = \Theta(\log N)$. Picking $m = s^{\varepsilon}$, we have

$$(1 - 2\delta) = \frac{\ell}{n} = \Theta(\frac{\ell}{ps}) = \Theta(\frac{m}{s}) = \Theta\big((\log N)^{\varepsilon - 1}\big)\,.$$

Thus the minimum distance δ (for our choice of m) satisfies $\delta = \frac{1}{2}\big(1 - \Theta((\log N)^{\varepsilon-1})\big)$.

Also we have

$$(1 - 2\delta)^{1/2+\varepsilon} \simeq s^{(\varepsilon-1)(1/2+\varepsilon)} \le (4s)^{-1/2}$$

for large enough N (since $\varepsilon < 1/2$). Thus there exist $\Omega(n^m) = N^{\Omega(\log^{\varepsilon} N)}$ codewords of $\text{RS-H}\textsc{ad}_t(\delta)$ all of which lie in a Hamming ball of radius $\frac{N}{2}(1 - (1 - 2\delta)^{1/2+\varepsilon})$. Since Theorem 13 implies that there are infinitely many choices for t that we could use, we also have infinitely many choices of blocklengths N available for the above construction, and the proof is thus complete. ■

D. Proof of Theorem 11

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) $L_c^{\mathrm{poly}}(\delta) \le \frac{1}{2}\Big(1 - \Big(\frac{1-2\delta}{6c}\Big)^{1/2}\Big)$. 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 14 in order to prove Theorem 11. We prove the following lemma which will in turn imply Theorem 11. Since our goal was only to establish Theorem 11, we have not attempted to optimize the exact bounds in the lemma below.

Lemma 15: For every c and every δ, we have

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

Proof: To prove the claimed upper bound on $L_c^{\rm poly}(\delta)$, we will closely follow the construction from the proof of Lemma 14. Let $0 < \delta < 1/2$, $0 \le \alpha \le (1/2 - \delta)$, and c be given. Define $\alpha' = 2\alpha$ and pick an integer s,

$$2(2c+1)/\alpha' \le s < 6(2c+1)/\alpha'$$

such that the conditions of Theorem 13 are met (we know such an s exists by the remarks following Theorem 13). Let t be *any* integer for which a subgroup S of $\mathrm{GF}(2^t)$ exists as guaranteed by Theorem 13 (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 $p = (n-1)/s$. As in the proof of Lemma 14, the code will again be $\text{RS-H}\textsc{ad}_t(\delta)$ (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 $\mathbf{r}$. Now, instead of using as received word the vector $\mathbf{v}^n$ (recall that $\mathbf{v}$ was the table of values of the Boolean function f with large Fourier support on a multiplicative subgroup S of $\mathrm{GF}(2^t)$), we will set the first $B = (\ell - \alpha' n) = (1 - 2\delta - \alpha')n$ blocks of $\mathbf{r}$ to be all zeroes. The last $(n-B)$ blocks of $\mathbf{r}$ will be vectors $\mathbf{v}^{(i)}$, $B < i \le n$, which will be specified shortly.

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

$$\mathrm{degree}(P) = B + pm = \ell - \alpha' n + \frac{n-1}{s}(2c+1) \le \ell$$

since we picked $s \ge 2(2c+1)/\alpha'$. By the choice of P, the codeword $(b_1, b_2, \ldots, b_n)$ corresponding to P (which we abuse notation and also denote by P) will agree with $\mathbf{r}$ in the first nB positions (as both begin with a string of nB zeroes). At each of the remaining $(n-B)$ blocks, we will have $b_i \in S_i \cup \{0\}$ where S_i is a coset S (recall that S is s-element multiplicative subgroup of $\mathrm{GF}(2^t)$ consisting of all the p'th powers). Specifically $S_i = \beta_i S$ where $\beta_i = (z_i - z_1)\cdots(z_i - z_B)$. Now, for $B < i \le n$, define $\mathbf{v}^{(i)} \in \{1,-1\}^{2^t}$ to the value of the functions $f^{(i)}$ where $f^{(i)} : \mathrm{GF}(2^t) \to \{1,-1\}$ is a function with $\sum_{\alpha \in S_i} \hat{f}_\alpha^{(i)} \ge \sqrt{s/3}$ as guaranteed by Theorem 13.

Using arguments similar to those in the proof of Lemma 14, one can show that with probability at least $1/2$, the codeword corresponding to the polynomial P differs from $\mathbf{r}$ in at most $E = (n-B)\big(\frac{1}{2} - \frac{1}{2\sqrt{4s}}\big)n$ positions. Thus there are at least $\frac{1}{2}n^m$ codewords of $\text{RS-H}\textsc{ad}_t(\delta)$ that lie 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

$$N(\delta + \alpha'/2)\big(1 - \sqrt{\frac{\alpha'}{24(2c+1)}}\big),$$

and recalling that $\alpha' = 2\alpha$, the claimed result follows. To conclude, we just reiterate that by Theorem 13, for the picked value of s, there are infinitely many values of t (and therefore the blocklength N) for which the code $\text{RS-H}\textsc{ad}_t(\delta)$ has the claimed properties. Thus we get an infinite family of codes with the requisite property, and the proof is complete. ∎

We now turn to the proof of of Theorem 11.

Proof: (of Theorem 11) We want to prove $L_c^{\rm poly}(\delta) < \delta$. Note that when $\delta > \frac{1}{2} - \frac{1}{48(2c+1)}$, setting $\alpha = 1/2 - \delta$ gives

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

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

$$L_c^{\rm poly}(\delta) \le \delta + \alpha - \delta(\frac{\alpha}{12(2c+1)})^{1/2} < \delta.$$

Thus we have $L_c^{\rm poly}(\delta) < \delta$ in either case. ∎

E. Proof of Theorem 13

The proof proceeds in several steps. We first prove 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 $\alpha \in S$.

Lemma 16: For any integer t, let S be an arbitrary 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 $\sum_{\alpha \in S} \hat{f}_\alpha \ge \sqrt{\frac{|S|}{3}}$.

Proof: For any set S, the following simple claim identifies the "best" function f for our purposes.

Claim: Define the function $g : \mathrm{GF}(2^t) \to \mathbb{R}$ by $g(x) = \sum_{\alpha \in S} \chi_\alpha(x)$. Then the maximum value of $\sum_{\alpha \in S} \hat{f}_\alpha$ achieved by a boolean function f is exactly $2^{-t} \cdot \sum_x |g(x)|$.

Proof: Indeed

$$\begin{aligned} 2^t \sum_{\alpha \in S} \hat{f}_\alpha &= \sum_{x, \alpha \in S} f(x)\chi_\alpha(x) = \sum_x f(x) \sum_{\alpha \in S} \chi_\alpha(x) \\ &= \sum_x f(x) g(x) \le \sum_x |g(x)| \end{aligned}$$

with equality holding when f is defined as $f(x) = \mathrm{sign}(g(x))$. ∎

Thus the above claim "removes" the issue of searching for an f by presenting the "best" choice of f, and one only needs to analyze the behavior of the above character sum function g, and specifically prove a lower bound on $\sum_x |g(x)|$.[1]

To get a lower bound on $\sum_x |g(x)|$, we employ Hölder's inequality which states that

$$\sum_x |h_1(x) h_2(x)| \le \left(\sum_x |h_1(x)|^p\right)^{1/p} \left(\sum_x |h_2(x)|^q\right)^{1/q},$$

[1] It can be shown that the representation of the field (as a vector space of dimension t over $\mathrm{GF}(2)$) does not affect the value distribution of g, and thus we can pick an arbitrary representation of the field, and the result will be the same.

for every positive p and q that satisfy $\frac{1}{p}+\frac{1}{q}=1$. Applying this with $h_1(x)=|g(x)|^{2/3}$, $h_2(x)=|g(x)|^{4/3}$, $p=3/2$ and $q=3$ gives

$$\left(\sum_x |g(x)|\right)^{2/3}\left(\sum_x g(x)^4\right)^{1/3} \geq \sum_x g^2(x). \qquad (9)$$

This inequality is also a consequence of log convexity of the power means (see Hardy, Littlewood, Polya [15]; Theorem 18).

Now $\sum_x g^2(x)=\sum_{\alpha_1,\alpha_2}\sum_x \chi_{\alpha_1+\alpha_2}(x)$ which equals $|S|\cdot 2^t$ (the inner sum equals 2^t whenever $\alpha_1=\alpha_2$ and 0 otherwise, and there are $|S|$ pairs (α_1,α_2) with $\alpha_1=\alpha_2$). Note that this also follows from Plancherel's identity.

Similarly

$$\sum_x g^4(x)=\sum_{\alpha_1,\alpha_2,\alpha_3,\alpha_4\in S}\sum_x \chi_{\alpha_1+\alpha_2+\alpha_3+\alpha_4}(x)$$

equals $N_{4,S}\cdot 2^t$ where $N_{4,S}$ is the number of 4-tuples in $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)\in S^4$ that sum up to 0. But the property satisfied by S, no four distinct elements of S sum up to 0, and hence the only such 4-tuples which sum up to 0 are those which have two of the α's equal. There are at most $3|S|^2$ such 4-tuples $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ with two of the α's equal. Hence $N_{4,S}\leq 3|S|^2$, and hence $\sum_x g^4(x)\leq 3|S|^2 2^t$. Plugging this into Equation (9) we get, when $f(x)=\mathrm{sign}(g(x))$,

$$\sum_{\alpha\in S}\hat{f}_\alpha=\frac{1}{2^t}\sum_x |g(x)|\geq\sqrt{\frac{|S|^3}{3|S|^2}}=\sqrt{\frac{|S|}{3}}.$$

■

Given the statement of Lemma 16, 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. For completeness sake, we quickly review the necessary facts about cyclic codes. A binary cyclic code of blocklength n is an ideal in the ring

$$R=\mathbb{F}_2[X]/(X^n-1).$$

It is characterized by its generator polynomial $g(X)$ where $g(X)|(X^n-1)$. The codewords correspond to polynomials in R that are multiples of $g(X)$ (the n coefficients of each such polynomial form the codeword symbols). A (binary) cyclic code is said to be maximal if its generator polynomial is irreducible over $\mathrm{GF}(2)$. A BCH code is a special kind of cyclic code whose generator polynomial is defined to be the minimal polynomial that has roots $\beta,\beta^2,\ldots,\beta^{d-1}$. Here β is a primitive n'th root of unity over $\mathrm{GF}(2)$, and d is the "designed distance" of the code.

Lemma 17: Let $k\geq 4$ 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 [22]), this implies that the minimum distance of C_h is at least 5, as desired. ■

Lemma 18: 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 lies in the range $s/2\leq t\leq s$, 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 17 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 18 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 $k\geq 1$. In particular we can repeatedly double t until it lies in the range $s/2\leq t\leq s$ (note that we had $t\leq s$ to begin with). 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 13.

Proof: (of Theorem 13) Theorem 13 now follows from Lemma 16 and Lemma 18. Note also that the statement of Lemma 18 implies the remarks made after the statement of Theorem 13. ■

F. Proof of Theorem 10

We begin by bounding the expected number of codewords in a random ball of an MDS code. Recall that an MDS code is an $[n,k]$ code whose minimum distance equals (the optimum value of) $(n-k+1)$.

Lemma 19: For any MDS $[n,k]_q$ code C and $a\geq k$,

$$\frac{1}{e}\binom{n}{a}q^{k-a}\leq \mathop{\mathbf{E}}_x[|B(x,n-a)\cap C|]\leq\binom{n}{a}q^{k-a}.$$

Proof: The upper bound follows from the claim that for any set S_a of a positions, the expected number of codewords which agree with x on S_a is at most q^{k-a}. To show this claim, first fix a subset $S_k\subseteq S_a$ of k of these positions. For each x, there is a unique codeword w_x that agrees with x on S_k. The probability that w_x agrees with x on S_a therefore equals q^{k-a}.

The lower bound follows from a similar claim: that for any set S_a of a positions, the probability that a codeword agrees with x

on S_a and disagrees with x outside of S_a is at least q^{k-a}/e. This claim is true because the probability that w_x above agrees with x on S_a and disagrees with x outside of S_a equals $q^{k-a}(1-1/q)^{n-a}$. For an MDS code, $n < q+k-1$, so $n-a \leq n-k < q-1$ so $(1-1/q)^{n-a} > 1/e$. ■

Corollary 20: For any constants $\varepsilon, \gamma > 0$, for large enough n, $L_n^{\rm poly}(1-n^{\varepsilon-1}) \leq 1-(1-\gamma)n^{\varepsilon-1}/\varepsilon$, where $L_q^{\rm poly}$ denotes the analog of $L^{\rm poly}$ for q-ary codes.

Proof: Use an MDS $[n,k]_q$ code with $n = q$ and $k = n^\varepsilon$, such as a Reed-Solomon code. Then

$$\binom{n}{a} q^{k-a} \geq \left(\frac{n}{a}\right)^a n^{k-a} = \frac{n^k}{a^a}.$$

Letting $a = (1-\gamma)n^\varepsilon/\varepsilon$, for large enough n we have $a^a \leq n^{(1-\gamma/2)n^\varepsilon}$, and the expected number of codewords in a ball of radius $n-a$ is $\Omega(n^{\frac{\gamma}{2}n^\varepsilon})$, yielding the corollary. ■

Proof: (of Theorem 10) We show that the family of codes $\mathcal{C}$ that we construct satisfies the property that every member $C \in \mathcal{C}$ with block length n satisfies
1. The relative minimum distance of C is at least $\frac{1}{2}\left(1-n^{\varepsilon-1/2}\right)$.
2. The list-of-$\ell(n)$ decoding radius of C is at most $\frac{1}{2}\left(1-\frac{1}{3\varepsilon}n^{\varepsilon-1/2}\right)$.

This suffices to prove the theorem.

The codes C in our family are concatenations of Reed-Solomon codes with Hadamard codes. For such a concatenated code C to have block length n, the RS code must have block length $\sqrt{n}$, and the relative minimum distance of C is half the relative minimum distance of the RS code. The theorem then follows from Corollary 20 for $\ell(n)$ growing exponentially in n. ■

IV. List Decoding Radius vs. Rate

We now prove Theorem 5.

Proof: (of Theorem 5) 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 dimension is $k = \lfloor (1-H(p)-1/c)n \rfloor$, for all large enough n. This clearly implies the lower bound on $U_c^{\rm const}$ claimed in the statement of the Theorem.

The code $\mathbf{C} = C_k$ will be built iteratively in k steps by randomly picking the k basis vectors in turn. Initially the code C_0 will just consist of the all-zeroes codeword $b_0 = 0^n$. The code 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 $C_i = {\rm span}(b_1, \ldots, b_i)$. Thus $\mathbf{C} = C_k$ is an $[n,k]_2$ linear code. We will now analyze the list of c decoding radius of the codes 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_C defined for a code C of blocklength n:

$$S_C = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 2^{\frac{n}{c} \cdot |B(x,e) \cap C|} . \quad (10)$$

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

Let $B = |B(0,e)| = |B(0,pn)|$; then $B \leq 2^{H(p)n}$ where $H(p)$ is the binary entropy function of p (see for example Theorem (1.4.5) in [22, Chapter 1]). Clearly

$$S_0 = 1 - B/2^n + B2^{n/c}/2^n \leq 1 + 2^{n\left(H(p)-1+1/c\right)} . \quad (11)$$

Now once 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 $C_{i+1} = {\rm span}(C_i \cup \{b_{i+1}\})$ 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 ${\rm span}(b_1, \ldots, b_i)$.

$$\begin{aligned}
\mathbf{E}[S_{i+1}] &= 2^{-n} \sum_x \mathbf{E}[2^{n/c \cdot T_x^{i+1}}] \\
&= 2^{-n} \sum_x \mathbf{E}[2^{n/c \cdot \left(|B(x,e) \cap C_i| + |B(x,e) \cap (C_i + b_{i+1})|\right)}] \\
&= 2^{-n} \sum_x \left(2^{n/c \cdot T_x^i} \mathop{\mathbf{E}}_{b_{i+1}} [2^{n/c \cdot T^i_{x+b_{i+1}}}]\right) \quad (12)
\end{aligned}$$

where in the second and third steps we used the fact that if $z \in B(x,e) \cap C_{i+1}$, then either $z \in B(x,e) \cap C_i$, or $z + b_{i+1} \in B(x,e) \cap C_i$. To estimate the quantity (12), first note that if we did not have the condition that b_{i+1} was chosen from outside ${\rm span}(b_1, \ldots, b_i)$ (12) would simply equal $\hat{S}_i^2$. This follows from the fact that x and $x + b_{i+1}$ are independent and the definition of $\hat{S}_i$. Now we use the simple fact that the expectation of a positive random variable taken over b_{i+1} chosen randomly from outside ${\rm span}(b_1, \ldots, b_i)$ is at most $(1-2^{i-n})^{-1}$ times the expectation taken over b_{i+1} chosen uniformly at random from $\{0,1\}^n$. Hence, we get that

$$\mathbf{E}[S_{i+1}] \leq \frac{\hat{S}_i^2}{(1-2^{i-n})}. \quad (13)$$

Applying (13) repeatedly for $i = 0, 1, \ldots, k-1$, we conclude that there exists an $[n,k]$ binary linear code $\mathbf{C}$ with

$$\begin{aligned}
S_{\mathbf{C}} = S_k &\leq \frac{S_0^{2^k}}{\prod_{i=0}^{k-1}(1-2^{i-n})^{2^{k-i}}} \\
&\leq \frac{S_0^{2^k}}{(1-2^{k-n})^k} \leq \frac{S_0^{2^k}}{1-k2^{k-n}} \quad (14)
\end{aligned}$$

since $(1-x)^a \geq 1 - ax$ for $x, a \geq 0$. Combining (14) with (11), we have

$$S_k \leq (1-k2^{k-n})^{-1}\left(1 + 2^{n(H(R)-1+1/c)}\right)^{2^k}$$

and using $(1+x)^a \leq (1+2ax)$ for $ax \ll 1$, this gives

$$S_k \leq 2(1 + 2 \cdot 2^{k+(H(p)-1+1/c)n}) \leq 6, \quad (15)$$

where the last inequality follows since $k = \lfloor (1-H(p)-1/c)n \rfloor$. By the definition of the potential S_k (10), this implies that

$$2^{n/c \cdot |B(x,e) \cap \mathbf{C}|} \leq 6 \cdot 2^n < 2^{n+3},$$

or

$$|B(x,e) \cap \mathbf{C}| \leq (1 + \frac{3}{n})c$$

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. ∎

Remark: One can also prove Theorem 5 with the additional property that the relative minimum distance $\Delta(R)$ of the code (in addition to its list decoding radius for list size c) also satisfies $\Delta(R) \geq H^{-1}(1-R-1/c)$. This can be done, for example, by conditioning the choice of the random basis vector b_{i+1} in the above proof so that $\mathrm{span}(b_1, b_2, \ldots, b_{i+1})$ does not contain any vector of weight less than pn. It is easy to see that with this modification, Equation (13) becomes

$$\mathbf{E}[S_{i+1}] \leq \frac{\hat{S}_i^2}{(1-2^{i+H(p)n-n})}.$$

Using exactly similar calculations as in the above proof, we can then guarantee a code $\mathbf{C}$ of dimension $k = \lfloor (1-H(p)-1/c)n \rfloor$ and minimum distance at least pn such that $S_{\mathbf{C}} = O(1)$.

V. Application to highly list decodable codes

We now apply the proof technique from the previous section to give constructions of concatenated codes that are list decodable from very high noise and yet have good rate. We first describe the setting that we are interested in, which is the same as the one that was considered in [13].

Given $\varepsilon > 0$, we are interested in asymptotically good family of binary linear codes $\mathcal{C}_\varepsilon$ that can be list decoded *efficiently* for up to a fraction $(1/2-\varepsilon)$ of errors. The goal is to give explicit (polynomial time) constructions of such code families with a reasonable rate. Such codes have a variety of applications some of which are discussed in [13], [20]. The best earlier result, due to [13], gives constructions with a rate of $\Omega(\varepsilon^6)$ (the construction is an algebraic-geometric code concatenated with any inner code like the Hadamard code that has large minimum distance). Note that if we did not care about efficient constructibility or efficient list decoding, then Theorem 5 guarantees that such code families exist with rate $\Omega(\varepsilon^2)$, and this is the best possible asymptotically.

Using the codes guaranteed by Theorem 5 as inner codes in a concatenation scheme with outer Reed-Solomon code, we can show that a rate of $\Omega(\varepsilon^6)$ can be achieved *without* relying on algebraic-geometric codes, thus "simplifying" the construction in [13]. This does not, however, improve the quantitative aspects of the earlier result. Instead we prove an adaptation of Theorem 5 that guarantees the existence of codes that have certain properties tailor-made for the weighted list decoding algorithm for Reed-Solomon codes from [12] to work well. Using such codes as inner codes in a concatenation scheme with outer Reed-Solomon code, gives us code families of rate $\Omega(\varepsilon^4)$ that are efficiently list decodable from a $(1/2-\varepsilon)$ fraction of errors. This is summarized in the following theorem, which is the main result of this section.

Theorem 21: There exist absolute constants $b, d > 0$ such that for each fixed $\varepsilon > 0$, there exists a polynomial time constructible code family $\mathcal{C}$ with the following properties:
1. $\mathrm{rate}(\mathcal{C}) \geq \frac{\varepsilon^4}{b}$
2. $\mathrm{Rad}(\mathcal{C}, d\varepsilon^{-2}) \geq \frac{1}{2} - \varepsilon$
3. $\Delta(\mathcal{C}) \geq (\frac{1}{2} - \varepsilon)$
4. There is a *polynomial time list decoding algorithm* for $\mathcal{C}$ that corrects up to a fraction $(1/2-\varepsilon)$ of errors.

The above theorem will follow from Theorem 24, which is stated and proved in Section V-B.

A. An "inner code" construction

A.1 Existence of a good code

We now prove the existence of codes that will serve as excellent inner codes in our later concatenated code construction. The proof is an adaptation of that of Theorem 5. We will then show how such a code can be constructed in $2^{O(n)}$ time (where n is the blocklength) using an iterative greedy procedure.

Lemma 22: There exist absolute constants $\sigma, A > 0$ such that for any $\varepsilon > 0$ there exists a binary linear code family $\mathcal{C}$ with the following properties:
1. $\mathrm{rate}(\mathcal{C}) = \sigma\varepsilon^2$
2. For every code $C \in \mathcal{C}$ and every $x \in \{0,1\}^n$ where n is the blocklength of C, we have

$$\sum_{\substack{\mathbf{c} \in C \\ \delta(x,\mathbf{c}) \leq (1/2-\varepsilon)}} \big(1 - 2\delta(x,\mathbf{c})\big)^2 \leq A\,. \qquad (16)$$

Proof: For every large enough n, we will prove the existence of a binary linear code C_k of blocklength n and dimension $k \geq \sigma\varepsilon^2 n$ which satisfies Condition (16) for every $x \in \{0,1\}^n$.

The proof will follow very closely the proof of Theorem 5 and in particular we will again build the code C_k iteratively in k steps by randomly picking the k basis vectors $b_1, b_2, \ldots, b_k$ in turn. Define $C_i = \mathrm{span}(b_1, \ldots, b_i)$ for $0 \leq i \leq k$. The key to our proof is the following potential function W_C defined for a code C of blocklength n (compare with the potential function (10) from the proof of Theorem 5):

$$W_C = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 2^{\frac{n}{A}\sum_{\mathbf{c} \in C : \delta(x,\mathbf{c}) \leq (1/2-\varepsilon)} (1-2\delta(x,\mathbf{c}))^2}\,. \qquad (17)$$

(The constant A will be fixed later in the proof, and we assume that $A > \ln 4$.) Denote the random variable W_{C_i} by the shorthand W_i, and for $x \in \{0,1\}^n$, define

$$R_x^i = \sum_{\substack{\mathbf{c} \in C_i \\ \delta(x,\mathbf{c}) \leq (1/2-\varepsilon)}} (1-2\delta(x,\mathbf{c}))^2\,, \qquad (18)$$

so that $W_i = 2^{-n}\sum_x 2^{\frac{n}{A} \cdot R_x^i}$.

Now, exactly as in the proof of Theorem 5, we have $R_x^{i+1} = R_x^i + R_{x+b_{i+1}}^i$, and using this it is straightforward to show that $\mathop{\mathbf{E}}_{b_{i+1}}[W_{i+1} | W_i = \hat{W}_i] = \hat{W}_i^2$ over the choice of b_{i+1} uniformly at random from $\{0,1\}^n$, and it is therefore easy to argue that

$$\mathbf{E}[W_{i+1} | W_i = \hat{W}_i] \leq \frac{\hat{W}_i^2}{1-2^{i-n}}\,. \qquad (19)$$

when the expectation is taken over a random choice of b_{i+1} outside $\mathrm{span}(b_1, \ldots, b_i)$. Applying (19) repeatedly for $i = 0, 1, \ldots, k-1$, we conclude that there exists an $[n,k]$ binary linear code $C = C_k$ with

$$W_C = W_k \leq \frac{W_0^{2^k}}{1-k2^{k-n}}\,. \qquad (20)$$

If we could prove, for example, that $W_C = O(1)$, then this would imply, using (17), that $R_x^k \le A$ for every $x \in \{0,1\}^n$ and thus C would satisfy Condition (16), as desired. To show this, we need an estimate (upper bound) on W_0, to which we turn next.

Define $A = (1/2 - \varepsilon)n$. Since C_0 consists of only the all-zeroes codeword, we have $R_x^0 = (1 - 2\mathrm{wt}(x)/n)^2$ if $\mathrm{wt}(x) \le a$ and $R_x^0 = 0$ otherwise (here we use $\mathrm{wt}(x) = \Delta(x, \mathbf{0})$ to denote the Hamming weight of x). Let us denote 2^x by $\exp_2(x)$. We now have

$$\begin{aligned} W_0 &= 2^{-n} \sum_{x \in \{0,1\}^n} \exp_2\big(\frac{n}{A} R_x^0\big) \\ &\le 1 + 2^{-n} \sum_{i=0}^{A} \binom{n}{i} \exp_2\big(\frac{n}{A}(1 - \frac{2i}{n})^2\big) \\ &\le 1 + n 2^{-n} \exp_2\Big(\max_{0 \le i \le A} \Big\{ H(\frac{i}{n}) n + \frac{4n}{A}(\frac{1}{2} - \frac{i}{n})^2 \Big\}\Big) \\ &\le 1 + n 2^{un} \end{aligned} \qquad (21)$$

where

$$u \stackrel{\text{def}}{=} \max_{0 \le y \le (1/2 - \varepsilon)} \Big\{ H(y) - 1 + \frac{4}{A}\Big(\frac{1}{2} - y\Big)^2 \Big\} .$$

We now claim that for every y, $0 \le y \le 1/2$, we have $H(y) \le 1 - \frac{2}{\ln 2}(\frac{1}{2} - y)^2$. One way to prove this is to consider the Taylor expansion around $1/2$ of $H(y)$, which is valid for the range $0 \le y \le 1/2$. We have $H'(1/2) = 0$ and $H''(1/2) = -4/\ln 2$. Also it is easy to check that all odd derivatives of $H(y)$ at $y = 1/2$ are zero while the even derivatives are non-positive. Thus

$$H(y) \le H(1/2) - H''(1/2)\frac{(1/2 - y)^2}{2} = 1 - \frac{2}{\ln 2}\big(\frac{1}{2} - y\big)^2 .$$

Therefore

$$\begin{aligned} u &\le \max_{0 \le y \le (1/2 - \varepsilon)} \Big(\frac{4}{A} - \frac{2}{\ln 2}\Big)\Big(\frac{1}{2} - y\Big)^2 \\ &= -4\Big(\frac{1}{\ln 4} - \frac{1}{A}\Big)\varepsilon^2 , \end{aligned} \qquad (22)$$

since $A > \ln 4$. Combining (20), (21) and (22), it is now easy to argue that we have $W_C = W_k = O(1)$ as long as $k < -un$, which will be satisfied if $k < 4(\frac{1}{\ln 4} - \frac{1}{A})\varepsilon^2 n$. Thus the statement of the lemma holds, for example, with $A = 2$ and $\sigma = 0.85$. ■

Remark: Arguing exactly as in the remark following the proof of Theorem 5, one can also add the condition $\Delta(\mathcal{C}) \ge (1/2 - \varepsilon)$ to the claim of Lemma 22. The proof will then pick b_{i+1} randomly from among all choices such that $\mathrm{span}(b_1, b_2, \ldots, b_{i+1}) \cap B(\mathbf{0}, (\frac{1}{2} - \varepsilon)n) = \emptyset$.

A.2 A greedy construction of the "inner" code

We now discuss how a code guaranteed by Lemma 22 can be constructed in a greedy fashion. We will refer to some notation that was used in the proof of Lemma 22. The algorithm works as follows:

Algorithm GREEDY-INNER:

Parameters: Dimension k; $\varepsilon, A > 0$ (where A is the absolute constant from Lemma 22)

Output: A binary linear code $C = \textsc{Greedy}(k, \varepsilon)$ with dimension k, blocklength $n = O(k/\varepsilon^2)$ and minimum distance $(1/2 - \varepsilon)n$ such that for every $x \in \{0,1\}^n$, Condition (16) holds.

1. Start with $b_0 = \mathbf{0}$.
2. For $i = 1, 2, \ldots, k$:
 - Let $U_i = \{x \in \{0,1\}^n : \mathrm{span}(b_1, b_2, \ldots, b_{i-1}, x) \cap B(\mathbf{0}, (1/2 - \varepsilon)n) = \emptyset\}$.
 - Pick $b_i \in U_i$ that *minimizes* the potential function $W_i = 2^{-n} \sum_x 2^{\frac{n}{A} \cdot R_x^i}$, where R_x^i is as defined in Equation (18) (break ties arbitrarily)
3. Output $C = \mathrm{span}(b_1, b_2, \ldots, b_k)$.

The following result easily follows from the proof of Lemma 22 and since each of the k iterations of the for loop above can be implemented to run in $2^{O(n)}$ time.

Lemma 23: Algorithm GREEDY-INNER constructs a code $\textsc{Greedy}(k, \varepsilon)$ with the desired properties in $k \cdot 2^{O(n)}$ time.

B. A concatenated code construction

The statement of Theorem 21 follows immediately from the concatenated code construction guaranteed by the following theorem.

Theorem 24: There exist absolute constants $b, d > 0$ such that for every integer K and every $\varepsilon > 0$, there exists a concatenated code $C_K \stackrel{\text{def}}{=} \mathrm{RS} \oplus \textsc{Greedy}(m, \varepsilon/2)$ (for a suitable parameter m) that has the following properties:

1. C_K is a linear code of dimension K, blocklength $N \le \frac{bK}{\varepsilon^4}$, and minimum distance at least $(\frac{1}{2} - \varepsilon)N$.
2. The generator matrix of C_K can be constructed in $N^{O(\varepsilon^{-2})}$ time.
3. C_K is $((\frac{1}{2} - \varepsilon)N, d/\varepsilon^2)$-list decodable; i.e. any Hamming ball of radius $(1/2 - \varepsilon)N$ has at most $O(\varepsilon^{-2})$ codewords of C_K.
4. There exists a polynomial time list decoding algorithm for C_K that can correct up to $(1/2 - \varepsilon)N$ errors.

Proof: The code C_K is constructed by concatenating an outer Reed-Solomon code over $\mathrm{GF}(2^m)$ of blocklength $n_0 = 2^m$ and dimension $k_0 = K/m$ (for some integer m which will be specified later in the proof) with an inner code $C_{\text{inner}} = \textsc{Greedy}(m, \varepsilon/2)$ (as guaranteed by Lemma 23). Since the blocklength of C_{inner} is $n_1 = O(\frac{m}{\varepsilon^2})$, the concatenated code C_K has dimension K and blocklength

$$N = O\Big(\frac{n_0 m}{\varepsilon^2}\Big) \qquad (23)$$

and minimum distance D at least

$$D \ge \Big(1 - \frac{K}{m n_0}\Big)\Big(\frac{1}{2} - \frac{\varepsilon}{2}\Big) . \qquad (24)$$

For ease of notation, we often hide constants using the big-Oh notation in what follows, but in all these cases the hidden constants will be absolute constants that do not depend upon ε. Note that since C_{inner} is constructible in $2^{O(n_1)} = 2^{O(m/\varepsilon^2)}$ time, and $m = \log n_0$, the generator matrix for C_K can be constructed in $N^{O(\varepsilon^{-2})}$ time. This proves Property 2 claimed in the theorem.

We will now present a polynomial time list decoding algorithm for C_K to recover from a fraction $(1/2-\varepsilon)$ of errors with a small ($O(\varepsilon^{-2})$) list size. This will clearly establish both Properties 3 and 4 claimed in the theorem.

Let $y \in \{0,1\}^N$ be any received word. We wish to find a list of all codewords $\mathbf{c} \in C_K$ such that $\Delta(y,\mathbf{c}) \le 1/2-\varepsilon$. For $1 \le i \le n_0$, denote by y_i (resp. $\mathbf{c}_i$) the portion of y (resp. $\mathbf{c}$) that corresponds to the i^{th} codeword position of the outer Reed-Solomon code. For $1 \le i \le n_0$ and $\alpha \in \mathrm{GF}(2^m)$, define

$$w_{i,\alpha} = \max\Big\{\Big(\frac{1}{2}-\frac{\varepsilon}{2}-\Delta(y_i, C_{\text{inner}}[\alpha])\Big), 0\Big\} \tag{25}$$

(here $C_{\text{inner}}[\alpha]$ denotes the inner encoding of α interpreted as an m-bit string). By the property of C_{inner} guaranteed by Lemmas 22 and 23, we have, for each i, $1 \le i \le n_0$,

$$\sum_{\alpha\in\mathrm{GF}(2^m)} w_{i,\alpha}^2 \le B' , \tag{26}$$

for some absolute constant B'.

Now, consider the following decoding algorithm for C_K. First, the inner codes are decoded by a brute force procedure that goes over all codewords. Specifically, for each position i of the outer Reed-Solomon code, the inner decoder passes a list of *all* field elements α with the respective weights $w_{i,\alpha}$ defined in Equation (25). The weight $w_{i,\alpha}$ may be interpreted as the reliability information for the possibility that the i'th symbol of the outer codeword was the field element α. The inner decoding takes $O(2^m) = O(n_0)$ time for each of the n_0 inner codes, and thus the total time required to perform this step is $\mathrm{poly}(N)$. We now have to perform decoding of the outer Reed-Solomon code taking into account these weights. For this we use a weighted (or "soft-decision") list decoding algorithm for Reed-Solomon codes from [12], similar to its use in [13] for decoding the Reed-Solomon concatenated with the Hadamard code. This algorithm guarantees to find, in time polynomial in n_0 and $1/\gamma$, a list of *all* codewords $\mathbf{c} \in C_K$ that satisfy

$$\sum_{i=1}^{n_0} w_{i,\mathbf{c}_i} \ge \sqrt{\Big(n_0 - \frac{n_0 - K/m + 1}{1+\gamma}\Big)\cdot \sum_{i,\alpha} w_{i,\alpha}^2} \tag{27}$$

where $\gamma > 0$ is a parameter to be set later, and by abuse of notation $w_{i,\mathbf{c}_i} = w_{i,\alpha_i}$ where $\alpha_i \in \mathrm{GF}(2^m)$ is such that $C_{\text{inner}}[\alpha_i] = \mathbf{c}_i$. Moreover, it is also known that there will be at most $(1+1/\gamma)$ codewords $\mathbf{c}$ that satisfy Condition (27) for *any* choice of weights $w_{i,\alpha}$, and thus the algorithm will output a list of at most $O(1/\gamma)$ codewords.

Using (25) and (26), we have that Condition (27) will be satisfied if

$$\sum_{i=1}^{n_0}\Big(\frac{1}{2}-\frac{\varepsilon}{2}-\frac{\Delta(y_i,\mathbf{c}_i)}{n_1}\Big) \ge \sqrt{\Big(\gamma n_0 + \frac{K}{m}\Big)\cdot n_0 B'}$$

which is equivalent to

$$\Delta(y,\mathbf{c}) \le N\Big(\frac{1}{2}-\frac{\varepsilon}{2}-\sqrt{B'\Big(\gamma+\frac{K}{mn_0}\Big)}\Big) \tag{28}$$

and, as long as we pick $\gamma \le \frac{\varepsilon^2}{8B'}$ and m such that $\frac{K}{mn_0} = \frac{K}{m2^m} \le \frac{\varepsilon^2}{8B'}$, we can hence conclude that Condition (27) is satisfied provided

$$\Delta(y,\mathbf{c}) \le \Big(\frac{1}{2}-\varepsilon\Big)N.$$

Thus we have a decoding algorithm that outputs a list of all $O(1/\gamma) = O(\varepsilon^{-2})$ codewords that differ from y in at most $(1/2-\varepsilon)N$ positions. Finally, by our choice of m, we have $mn_0 = O(K/\varepsilon^2)$, and plugging this into (23) and (24), we have that the blocklength N of C_K satisfies $N = O(K/\varepsilon^4)$ and the distance D satisfies $D \ge (1/2-\varepsilon)N$, as desired. ■

Discussion: The time required to construct a code with the properties claimed in Theorem 24, though polynomial for every fixed ε, grows as $N^{O(\varepsilon^{-2})}$. Thus these codes are not *uniformly constructive* (i.e. are constructible in $O(f(\varepsilon)n^c)$ time for a fixed constant c, independent of ε, for some arbitrary function f). If one uses the best known algebraic-geometric codes (which in particular beat the Gilbert-Varshamov bound) as the outer code instead of Reed-Solomon codes, one can carry out the code construction of Theorem 24 in $2^{O(\varepsilon^{-2}\log(1/\varepsilon))}N^c$ time for a fixed constant c (the constant c will depend upon the time required to construct the outer algebraic-geometric code). This is not entirely satisfying since the construction complexity of such algebraic-geometric codes that beat the Gilbert-Varshamov bound is still quite high. It is an interesting open question to find an alternative, simpler construction of uniformly constructive codes which meet the requirements of Theorem 24.

VI. Concluding Remarks

In this paper, we reported codes with non-trivial list decoding properties. One of our results was to show the existence of *linear* codes that have an arbitrarily large polynomial number of codewords in a Hamming ball of relative radius strictly less than the relative distance. While it is easy to show that non-linear codes with this property exist (by a simple random coding argument), the situation for linear codes is more tricky. Recently, the techniques used in Section III of this paper were used together with some new ideas to prove that, under a widely believed number-theoretic conjecture, the result of Conjecture 9 holds [11] (see also [10, Chap. 4]). However, this does not subsume the result of Theorem 11 in this paper, since our result holds unconditionally without the need for any unproven number-theoretic conjecture.

We also demonstrated the existence of codes of good rate with a small number of codewords in a Hamming ball of large radius (Theorem 5). Our proof, however, was highly non-constructive and does not even give a high probability result. It is an open question whether a random linear code satisfies the property claimed in Theorem 5 with high probability.

We then showed that the statement of Theorem 5 can be adapted to guarantee the existence of certain linear codes which serve as good (for purposes of list decoding) inner codes in a concatenation scheme with an outer Reed-Solomon code. This in turn gave us an *efficiently constructible* family of binary linear codes of rate $\Omega(\varepsilon^4)$ and relative distance at least $(1/2-\varepsilon)$, which can be efficiently list decoded from up to a $(\frac{1}{2}-\varepsilon)$ fraction of errors, using lists of size $O(\varepsilon^{-2})$. This improves upon the results claimed in [13] (the best rate achieved by [13] for such families of codes was $\Omega(\varepsilon^6)$). The time required to construct such a code, though polynomial for every fixed ε, grows exponentially in $1/\varepsilon$, and it will be desirable to, if possible, bring this down to polynomial in both N and $1/\varepsilon$.

ACKNOWLEDGEMENTS

We thank Amnon Ta-Shma and Alex Russell for helpful discussions about Theorem 10.

REFERENCES

[1] R. Ahlswede. Channel capacities for list codes. *J. Appl. Probability*, 10 (1973), pp. 824-836.

[2] A. Ashikhmin, A. Barg, and S. Litsyn. A new upper bound on codes decodable into size-2 lists. In Ingo Althofer *et al.*, editor, *Numbers, Information and Complexity*, pages 239–244. Boston: Kluwer Publishers, 2000.

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

[4] V. M. Blinovsky. *Asymptotic Combinatorial Coding Theory*. Kluwer Academic Publishers, Boston, 1997.

[5] V. M. Blinovsky. Lower Bound for the Linear Multiple Packing of the Binary Hamming Space. *Journal of Combinatorial Theory, Series A*, Vol. 92, No. 1, October 2000, pp. 95-101.

[6] I. Dumer, D. Micciancio and M. Sudan. Hardness of approximating the minimum distance of a linear code. *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science (FOCS)*, New York, NY, October 1999, pp. 475-484.

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

[8] P. Elias. Error-correcting codes for List decoding. *IEEE Transactions on Information Theory*, 37(1):5–12, 1991.

[9] O. Goldreich, R. Rubinfeld and M. Sudan. Learning polynomials with queries: the highly noisy case. *SIAM Journal on Discrete Mathematics*, 13(4):535–570, November 2000.

[10] V. Guruswami. *List Decoding of Error-Correcting Codes*. Ph.D thesis, Massachusetts Institute of Technology, August 2001.

[11] V. Guruswami. Limits to list decodability of linear codes. *Manuscript*, November 2001.

[12] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and Algebraic-geometric codes. *IEEE Trans. on Information Theory*, 45 (1999), pp. 1757-1767.

[13] V. Guruswami and M. Sudan. List decoding algorithms for certain concatenated codes. *Proceedings of the 32nd ACM Symposium on the Theory of Computing (STOC)*, Portland, OR, May 2000, pp. 181-190.

[14] V. Guruswami and M. Sudan. *Extensions to the Johnson Bound*. Manuscript, February 2001.

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

[16] J. Justesen and T. Høholdt. Bounds on list decoding of MDS codes. *IEEE Transactions on Information Theory*, 47(4):1604–1609, May 2001.

[17] 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.

[18] M. A. Shokrollahi and H. Wasserman. List decoding of algebraic-geometric codes. *IEEE Trans. on Information Theory*, 45(2):432–437, March 1999.

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

[20] M. Sudan. List Decoding: Algorithms and Applications. *SIGACT News*, Vol. 31, March 2000, pp. 16–27.

[21] M. A. Tsfasman, S. G. Vlădut and T. Zink. Modular curves, Shimura curves, and codes better than the Varshamov-Gilbert bound. *Math. Nachrichten*, 109:21–28, 1982.

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

[23] V. K. Wei and G. L. Feng. Improved lower bounds on the sizes of error-correcting codes for list decoding. *IEEE Trans on Info Theory*, 40(2):559–563, 1994.

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

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