List Decoding from Erasures: Bounds and Code Constructions *

Venkatesan Guruswami†

University of California at Berkeley
Computer Science Division
Berkeley, CA 94720.
Email: venkat@lcs.mit.edu

Abstract

We consider the problem of list decoding from erasures. We establish lower and upper bounds on the rate of a (binary linear) code that can be list decoded with list size L when up to a fraction p of its symbols are adversarially erased. Our results show that in the limit of large L, the rate of such a code approaches the "capacity" $(1-p)$ of the erasure channel. Such nicely list decodable codes are then used as inner codes in a suitable concatenation scheme to give a uniformly constructive family of asymptotically good binary linear codes of rate $\Omega(\varepsilon^2/\log(1/\varepsilon))$ that can be efficiently list decoded using lists of size $O(1/\varepsilon)$ from up to a fraction $(1-\varepsilon)$ of erasures, for arbitrary $\varepsilon > 0$. This improves previous results from [19] in this vein, which achieved a rate of $\Omega(\varepsilon^3 \log(1/\varepsilon))$.

Keywords: Error-correcting codes, Linear codes, Decoding algorithms, List decoding, Erasure channel, Generalized Hamming Weights.

1 Introduction

The list decoding problem for an error-correcting code consists of outputting a list of all codewords that lie within a certain Hamming distance of the received word [10, 30]. The decoding is considered successful as long as the correct codeword is included in the list. If the decoder is restricted to output lists of size L, the decoding is called list-of-L decoding. List decoding with even moderate-sized lists allows one to correct significantly more errors than possible with unambiguous decoding (cf. [11, 18, 15]).

In this paper, we consider the problem of list decoding from *erasures*, i.e. list decoding when a certain fraction of the symbols of the received word are *adversarially* erased. Specifically, we are interested in codes with non-trivial list decoding performance – i.e. codes of "large" rate that are list decodable using "small" lists from a "large" number of erasures. Such a study has already been carried out with some success for the case of errors [11, 31, 6, 15]. Though the case of erasures is algorithmically always easier to handle, it raises several interesting combinatorial questions which

*A preliminary version of this paper appears in the Proceedings of the 21st Conference on *Foundations of Software Technology and Theoretical Computer Science* (FSTTCS), Bangalore, India, December 13-15, 2001.

†Supported by a Miller Research Fellowhip. Most of this work was done when the author was at the Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

this paper addresses. Also, the simpler situation of erasures makes better trade-offs (between rate and list decodability) possible than for the errors case. For example, for binary codes the maximum fraction of errors one can hope to correct from is $(1/2-\varepsilon)$, where as decoding up to a fraction $(1-\varepsilon)$ of erasures is possible. The rate (as a function of ε) that one can achieve for the erasures case also turns out be better than what is possible for the errors case.

1.1 Definitions and Notation

We focus on linear codes throughout this paper. For simplicity we also restrict our attention to binary codes though all our results go through for larger alphabets as well. Recall that a binary linear code C of block length n, dimension k and minimum distance d, denoted an $[n,k,d]_2$ code, is a k-dimensional subspace of $\mathbb{F}_2^n$ such that the minimum Hamming weight of a non-zero element (called codeword) of C equals d. When we omit the distance, we refer to the code as an $[n,k]_2$ code. The rate of the code is defined to be the ratio k/n and the relative distance is defined to be d/n. The main thrust of this paper is the asymptotic performance of codes, and we therefore focus on infinite families $\mathcal{C}=\{C_i\}_{I\geq 1}$ of $[n_i,k_i,d_i]_2$ codes of increasing block length $n_i\to\infty$. The rate $R(\mathcal{C})$ of such a family is defined to be $R(\mathcal{C})\stackrel{\text{def}}{=}\liminf_i\left\{\frac{k_i}{n_i}\right\}$ and its relative distance is defined to be $\delta(\mathcal{C})\stackrel{\text{def}}{=}\liminf_i\left\{\frac{d_i}{n_i}\right\}$.

We next define the list decoding radius for recovering from erasures. For $y\in\{0,1\}^n$ and $T\subseteq\{1,2,\ldots,n\}$, define $[y]_T\in\{0,1\}^{|T|}$ to be the projection of y onto the coordinates in T.

Definition 1 ((s,L)-erasure list-decodability) *A (binary) code C of block length n is said to be* (s,L)-erasure list-decodable *if for every $r\in\{0,1\}^{n-s}$ and every set $T\subseteq\{1,2,\ldots,n\}$ of size $(n-s)$, we have $|\{c\in C:[c]_T=r\}|\leq L$. In other words, given any received word with at most s erasures, the number of codewords consistent with the received word is at most L.*

Note that a code of minimum distance d is $(d-1,1)$-erasure list-decodable, but is *not* $(d,1)$-erasure list-decodable.

Definition 2 (Erasure List Decoding Radius) *For an integer $L\geq 1$ and a code C, the* list-of-L erasure decoding radius *of C, denoted* $\text{radius}_L(C)$*, is defined to be the maximum value of s for which C is (s,L)-erasure list-decodable.*

Definition 3 (ELDR for Code Families) *For an infinite family $\mathcal{C}=\{C_i\}_{i\geq 1}$ of codes and an integer $L\geq 1$, the* list-of-L erasure decoding radius *of $\mathcal{C}$, denoted $\mathsf{ELDR}_L(\mathcal{C})$, is defined to be*

$$\mathsf{ELDR}_L(\mathcal{C})\stackrel{\text{def}}{=}\liminf_i\left\{\frac{\text{radius}_L(C_i)}{n_i}\right\}\ , \tag{1}$$

where n_i is the block length of C_i.

We now define the function which measures the trade-off achievable between rate and erasure list decoding radius.

Definition 4 *For an integer L and $0\leq p\leq 1$, the maximum rate of a linear code family with list-of-L erasure decoding radius at least p, denoted $R_L(p)$, is defined as*

$$R_L(p)\stackrel{\text{def}}{=}\sup_{\mathcal{C}:\ \mathsf{ELDR}_L(\mathcal{C})\geq p}R(\mathcal{C})\ . \tag{2}$$

When the supremum above is over all code families, including non-linear ones, we denote the above quantity by $R^{\text{nl}}(p)$.

1.2 Comment on combinatorial vs. algorithmic erasure list-decodability

For linear codes, list decoding from erasures is algorithmically easy as it just amounts to finding the space of solutions to a linear system. Thus, if a linear code is (s, L)-erasure list-decodable for some small L, then a list of size at most L can also be *efficiently output* given any received word with at most s erasures. (This is not true for the case when there are errors, where algorithmic list decodability is potentially a lot more difficult to achieve than combinatorial list decodability – in fact there are known constructions of codes with good combinatorial list decodability, but for which efficient list decoding algorithms are not known.) Since our focus in this paper is on recovery from erasures, we only focus on linear codes with good combinatorial erasure list decodability, and this automatically implies a polynomial time list decoding algorithm from a large number of erasures. But for our main explicit code construction, we do mention how the decoding time can be improved to near-quadratic time from the obvious cubic time required to solve a linear system.

1.3 Relation to Generalized Hamming Weights

The notion of erasure list-decodability of linear codes has been implicitly studied in the literature under the name of *generalized Hamming weights* (this was pointed out to us by Ralf Koetter). For a set $D \subseteq \mathbb{F}_2^n$, define the support of D as $\mathrm{supp}(D) = \{i : \exists\ \langle x_1, x_2, \ldots, x_n \rangle \in D,\ x_i \neq 0\}$.

Definition 5 (Generalized Hamming weight) *The* r'th generalized Hamming weight *of a linear code C, denoted $d_r(C)$, is defined to be the size of the smallest support of an r-dimensional subcode of C.*

Note that $d_1(C)$ equals the traditional minimum distance of C. The concept of generalized Hamming weights was first introduced in [29] with some cryptographic applications in mind, and has since received an enormous amount of attention; see, for example, [8, 5], and in particular the expository paper [26] and the pointers therein. It is known to have a wealth of unexpected applications such as in cryptography (which was the original motivation for its introduction in [29]), the study of t-resilient functions, state complexity of trellis diagrams, and in the design of codes for the switching multiple-access channel. However, surprisingly, its relation to list decoding, as detailed in the simple lemma below, seems not to have been made explicit before in the literature.

Lemma 1 *Let C be any linear code. Then, C is (e, L)-erasure list-decodable if and only if $d_r(C) > e$ where $r = 1 + \lfloor \lg L \rfloor$.*

Proof: Suppose C is (e, L)-erasure list-decodable. Let $c_1, c_2, \ldots, c_r$ be r linearly independent codewords in C which span a subcode C' of dimension r. We wish to show that $\mathrm{supp}(C') > e$. Consider the received word $\mathbf{r}$ that has all positions in $\mathrm{supp}(C')$ erased and 0's in the remaining positions. Clearly each of the $2^r > L$ codewords in C' is consistent with $\mathbf{r}$, which together with the (e, L)-erasure list-decodability of C implies that $\mathrm{supp}(C') > e$.

Conversely, suppose $d_r(C) > e$. Let $\mathbf{r}$ be a received word with e erasures; let T be the set of unerased positions. The set of codewords of C consistent with $\mathbf{r}$ is an affine space consisting of solutions to the linear system $C(x)_{|T} = \mathbf{r}$. Let this set be $\{c_1, \ldots, c_M\}$; we wish to prove $M \leq L$. Consider the set $D = \{c_i - c_1 : 1 \leq i \leq M\}$. Then D is a linear subspace of C with $\mathrm{supp}(D) \cap T = \emptyset$, and hence $|\mathrm{supp}(D)| \leq e$. Using $d_r(C) > e$, this means $\dim(D) \leq r - 1 = \lfloor \lg L \rfloor$, implying $M = |D| = 2^{\dim(D)} \leq 2^{\lfloor \lg L \rfloor} \leq L$. □

In light of the above connection, results on list decoding presented in this paper inherit the applications of generalized Hamming weights and thus have interest beyond the subject of list decoding. However, since our primary motivation came from list decoding, we state all our results using the terminology of list decodable codes, and if the reader feels the need, she should have no difficulty in translating them into the language of generalized Hamming weights.

1.4 Main Results

1.4.1 Combinatorial Bounds

One of the objectives of this paper is to establish lower and upper bounds on the function $R_L(p)$. By the relation to generalized Hamming weights, we are able to obtain such bounds by simply appealing to existing results in the literature. Since the list-of-1 ELDR of a code family equals its relative distance δ, $R_1(p) = R(\delta)$ is the central function in coding theory that studies the trade-off between the rate and minimum distance of a code. As a consequence of the combinatorial results (borrowed from the literature on generalized Hamming weights) that we state, it turns out that in the limit of large L, the function $R_L(p)$ tends to $1 - p$, thus matching the Singleton bound. This result has the following nice interpretation. It is a classical result in coding theory that the capacity of the erasure channel where each codeword symbol is randomly and independently erased with probability p, equals $(1 - p)$ [9]. Thus our results show that using list decoding with large enough (but constant-sized) lists, one can approach capacity even if the symbols that are erased are *adversarially* (and not randomly) chosen. The upper bound on $R_L(p)$ also shows that $R_L(p) < 1-p$ for every p and every fixed list size L (this result holds even for general, non-linear codes). Thus one *needs* unbounded list size in order to approach the capacity of the adversarial erasure channel using list decoding. We remark that analogous statements also hold for the errors case from earlier results in [11, 31, 6, 15].

1.4.2 Code Constructions

For any constant $\varepsilon > 0$, our lower bound on $R_L(p)$ implies the existence of a binary linear code family $\mathcal{C}_\varepsilon = \{C_i\}_{i \geq 1}$ of rate $\Omega(\varepsilon/\lg(1/\varepsilon))$ such that each C_i is $((1-\sigma)n_i, O(1/\sigma))$-erasure list-decodable for every σ satisfying $\varepsilon \leq \sigma \leq 1$. One can use such codes as inner codes in a concatenated scheme with outer code being any code of rate $\Omega(\varepsilon)$ and relative distance $1 - O(\varepsilon)$ (like Reed-Solomon codes, or, if one wants codes over constant alphabet size, algebraic-geometric codes or the expander based codes of [2]). Such a concatenation scheme gives a *uniformly constructive* family of binary linear codes of rate $\Omega(\varepsilon^2/\lg(1/\varepsilon))$ that can be list decoded from a fraction $(1-\varepsilon)$ of erasures using lists of size $O(1/\varepsilon)$.[1] The previous best result in this vein from [19] achieved a rate of $\Omega(\varepsilon^3)$, by exploiting the fact that any family of codes with relative distance greater than $(1 - \varepsilon)/2$ has a small list size for decoding from a fraction $(1 - \varepsilon)$ erasures (see Section 2.1 for some bounds of this nature). The best construction of such high distance codes achieves a rate of $\Omega(\varepsilon^3)$ (see [2]). We stress that the novelty of our result is that it is **not** obtained by appealing to this distance to ELDR relation, and indeed no polynomial time constructions of binary code families of relative distance $(1/2 - O(\varepsilon))$ and rate about ε^2 are known. In fact such a construction (which will asymptotically match the Gilbert-Varshamov bound at low rates) will be a *major* breakthrough in coding theory.

Moreover, by using Reed-Solomon codes as outer codes, we can get a code with slightly worse construction time, but which can be encoded in near-linear time and decoded in near-quadratic time. The decoding is based on a list decoding algorithm due to [18] and its fast implementation due to [24, 25].

[1] The term "uniformly constructive" means a polynomial time construction where the exponent of the polynomial is an absolute constant independent of ε.

2 Combinatorial Results on Erasure List-Decodability

2.1 Erasure list-decodability and minimum distance

We now study the erasure list-decodability of a code *purely* as a function of its minimum distance. The two results below together show that, purely as a function of the relative distance, the best bound on the erasure list decoding radius of a binary code of relative distance δ is 2δ. This is the analog of the "Johnson bound on list decoding radius" for the case of erasures.

Lemma 2 *Let C be a binary code of block length n and relative distance δ, $0 < delta < 1/2$. Then, for any $\varepsilon > 0$, C is $((2-\varepsilon)\delta n, \frac{2}{\varepsilon})$-erasure list-decodable.*

Proof: Define $s = (2-\varepsilon)\delta n$, $t = n - s$, and let $\mathbf{r} \in \{0,1\}^t$ be a received word with s erasures. Without loss of generality, assume that the first s symbols of $\mathbf{r}$ have been erased. Let c_i, $1 \le i \le M$, be all the codewords of C that agree with $\mathbf{r}$ in the last $(n-s)$ positions. Hence they all agree with each other in the last t positions. For each i, $1 \le i \le M$, define $\tilde{c}_i$ to be the truncation of c_i to the first s positions. We have

$$\mathrm{HamDist}(\tilde{c}_i, \tilde{c}_j) \ge \delta n = (2-\varepsilon)^{-1} s \ . \tag{3}$$

Thus we have M strings of length s with fractional distance between any pair strictly larger than $1/2$. It is a folklore fact that there can be at most a constant (independent of s) number of such strings. One way to prove this is as follows. Associate an s-dimensional unit vector v_i with each $\tilde{c}_i$ by setting its ℓ'th component $v_{i,\ell}$, $1 \le \ell \le s$, as follows: $v_{i,\ell} = 1/\sqrt{s}$ if $\tilde{c}_{i,\ell} = 1$ and $v_{i,\ell} = -1/\sqrt{s}$ otherwise. Clearly, the dot product of v_i and v_j satisfies

$$\langle v_i, v_j \rangle = \sum_{\ell=1}^{s} v_{i,\ell} v_{j,\ell} = 1 - \frac{2 \cdot \mathrm{HamDist}(\tilde{c}_i, \tilde{c}_j)}{s} \ ,$$

which together with Equation (3) implies that

$$\langle v_i, v_j \rangle \le \frac{-\varepsilon}{2-\varepsilon}$$

for $1 \le i < j \le M$. By a standard geometric fact (see, for example, [7, §10, Lemma 4]), this means that the number of such vectors M is at most $1 + \frac{2-\varepsilon}{\varepsilon} = \frac{2}{\varepsilon}$. Hence $M \le \frac{2}{\varepsilon}$, and therefore C is $(s, \frac{2}{\varepsilon})$-erasure list-decodable. $\square$

Lemma 3 *For every δ, $0 < \delta < 1/2$, and all small enough $\varepsilon > 0$, the following holds. For all large enough n, there exists a binary linear code C of relative distance at least δ which is* **not** *$((2+\varepsilon)\delta n, 2^{\Omega(\varepsilon^2 \delta n)})$-erasure list-decodable.*

Proof: Define $n' = (2+\varepsilon)\delta n$. Pick a binary linear code C' of block length n' that meets the Gilbert-Varshamov bound and has minimum distance at least $\delta n = n'/(2+\varepsilon)$. The relative distance δ' of this code is at most $(1/2 - O(\varepsilon))$, and therefore the code has rate $r \ge 1 - H_q(\delta') = \Omega(\varepsilon^2)$. Now form a code C of block length n from C' by just padding each codeword with $(n - n')$ zeroes. The minimum distance of C is still at least δn, and hence its relative distance is at least δ. Moreover, consider the received word $\mathbf{r}$ for C whose first n' positions are erased and the last $(n-n')$ positions contain zeroes. Clearly all codewords in C agree with $\mathbf{r}$ in the unerased positions. Thus, C is not $(n', |C|-1)$-erasure list-decodable. Recalling that $n' = (2+\varepsilon)\delta n$ and $|C| = |C'| \ge 2^{rn'} = 2^{\Omega(\varepsilon^2 \delta n)}$, we get the claimed result. $\square$

The above two results indicate that in order to construct binary codes which are erasure list-decodable up to a fraction $(1 - \varepsilon)$ of erasures, it suffices to construct codes which have relative distance $(1/2 - O(\varepsilon))$. Moreover, if one only uses the distance to bound the erasure list decoding radius, then a relative distance of $(1/2 - O(\varepsilon))$ is also necessary. Since there is an upper bound of $O(\varepsilon^2 \log(1/\varepsilon))$ on the rate of such large distance binary codes (the Linear Programming bound [23]), this indicates that the best rate for codes erasure list-decodable from a fraction $(1 - \varepsilon)$ of erasures that one can obtain by this method is also $O(\varepsilon^2 \log(1/\varepsilon))$. However, it turns out that a much better rate of $\Omega(\varepsilon)$ is achievable for such codes by directly studying the trade-off between erasure list-decodability and rate. The detailed investigation of such a trade-off is the subject of the next section.

2.2 Lower bound on $R_L(p)$

All logarithms in this paper are to the base 2 and will be denoted by $\lg x$. For a real number x, $\lfloor x \rfloor$ denotes the largest integer less than or equal to x, and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

The connection to generalized Hamming weights from Lemma 1, together with existential lower bounds in the literature for the rate as a function of generalized Hamming weights [29, 21, 26], gives the following when stated in our notation.

Theorem 4 *For every integer $L \geq 1$ and every p, $0 \leq p \leq 1$, we have*

$$R_L(p) \geq 1 - p - \frac{H(p)}{\lceil \lg(L+1) \rceil} . \tag{4}$$

The above bound, which is the generalization of the Gilbert-Varshamov bound for list sizes greater than 1, follows from Part (i) of the stronger claim of the lemma below. The second part of the lemma will be used in the application to concatenated codes in Section 3. It asserts the existence of codes that exhibit a "gradual" increase in list size as more and more codeword symbols are erased, up to a fraction $(1 - \varepsilon)$ of erasures. Since Part (ii) is not stated explicitly in the literature on generalized Hamming weights and since the proof is fairly simple, we include a proof for the sake of completeness.

Lemma 5 *For all large enough n, the following hold:*

(i) For every integer $L \geq 1$ and every p, $0 \leq p \leq 1$, there exists an $[n, k]_2$ linear code C with $k = \lfloor (1 - p - \frac{H(p)}{\lceil \lg(L+1) \rceil})n - \sqrt{n} \rfloor$ that is (pn, L)-erasure list-decodable.

(ii) There exists absolute constants $A, B > 0$ such that for every ε, $0 \leq \varepsilon \leq 1$, there exists an $[n, k]_2$ linear code C with $k = \lfloor \frac{\varepsilon n}{\lg(A/\varepsilon)} \rfloor$ that is $(s, \frac{Bn}{n-s})$-erasure list-decodable for every *$s \leq (1 - \varepsilon)n$.*

Before proving the above lemma, we state the following folklore combinatorial lemma gives a useful linear-algebraic characterization of when a linear code is (s, L)-erasure list-decodable.

Lemma 6 *An $[n, k]_2$ linear code C is (s, L)-erasure list-decodable if and only if its $n \times k$ generator matrix G has the property that every $(n - s) \times k$ sub-matrix of G has rank at least $(k - \lfloor \lg L \rfloor)$.*

Proof: Let $T \subseteq \{1, 2, \ldots, n\}$ with $|T| = n - s$ and $r \in \{0, 1\}^{n-s}$, the number of codewords $c \in C$ with $[c]_T = r$ is precisely the number of solutions $x \in \{0, 1\}^k$ to the system $G_T x = r$ where G_T is the sub-matrix of G consisting of all rows indexed by elements in T. By standard linear algebra, the number of solutions x to the linear system $G_T x = \mathbf{0}$ is precisely 2^ℓ where $\ell = k - \mathrm{rank}(G_T)$, and for any $r \in \{0,1\}^{n-s}$, the number of solutions x to $G_T x = r$ is at most 2^ℓ (in fact, it is always either 0 or 2^ℓ). Hence C is (s, L)-erasure list-decodable if and only if for every $T \subseteq \{1, \ldots, n\}$ with $|T| = n - s$, G_T has rank at least $k - \lfloor \lg L \rfloor$. □

Proof of Lemma 5: We first prove Part (i) of the lemma. The proof is based on the probabilistic method. We will pick a code C generated by a random $n \times k$ generator matrix G where k is as specified in the statement of the lemma.[2] We will prove that such a random code is (pn, L)-erasure list-decodable with high probability. By Lemma 6, this amounts to estimating the probability that a random $n \times k$ matrix over $\mathbb{F}_2$ has the property that every $t \times k$ sub-matrix has rank at least $(k - \lfloor \lg L \rfloor)$, where $t = (1-p)n$. For any J', $0 \le J' \le k$, the probability that a random matrix M of order $t \times k$ has rank $(k - J')$ is at most $2^{kJ'} 2^{-tJ'}$. This follows since for a fixed subspace S of $\mathbb{F}_2^k$ of dimension $(k - J')$, the probability that all t rows of M lie in S is at most $2^{-J't}$, and furthermore the number of subspaces of $\mathbb{F}_2^k$ of dimension $(k - J')$ is at most $2^{kJ'}$ as one can specify such a subspace as the null-space of a $J' \times k$ matrix over $\mathbb{F}_2$. Let $J \stackrel{\text{def}}{=} \lceil \lg(L+1) \rceil = 1 + \lfloor \lg L \rfloor$. Then, for $t \ge k$, the probability that a random matrix M of order $t \times k$ has rank $(k - J)$ or less is at most

$$\sum_{J'=J}^{k} 2^{(k-t)J'} \le k 2^{(k-t)J} .$$

Now, by a union bound, the probability that some $t \times k$ sub-matrix of G has rank at most $(k - J)$ is at most

$$\begin{aligned}
\binom{n}{t} \cdot k 2^{(k-t)J} &= \binom{n}{(1-p)n} \cdot k 2^{(k-t)J} \\
&\le k \cdot 2^{H(p)n} \cdot 2^{kJ - (1-p)nJ} \\
&\le k \cdot 2^{H(p)n} \cdot 2^{\left((1-p)J - H(p)\right)n} \cdot 2^{-J\sqrt{n}} \cdot 2^{-(1-p)nJ} \\
&= k \cdot 2^{-J\sqrt{n}} \\
&= o(1) ,
\end{aligned}$$

where in the second step we have used the fact that $\binom{n}{(1-p)n} = \binom{n}{pn} \le 2^{H(p)n}$ (cf. [28, Chap. 1, Theorem 1.4.5]), and in the third step we substituted the value of k from the statement of Part (i) of the lemma.

Hence an $n \times k$ matrix G every $t \times k$ sub-matrix of which has rank at least $k - J + 1 = k - \lfloor \lg L \rfloor$ exists, and by Lemma 6, therefore, there exists a (pn, L)-erasure list-decodable code of block length n and dimension k. This proves Part (i) of the lemma.

To prove Part (ii), we apply the above proof for $p = (1 - \varepsilon)$, and $k = \lfloor \frac{\varepsilon}{\lg(8/\varepsilon)} n \rfloor$. Let us estimate the probability that for a fixed s, where $n/2 \le s \le (1-\varepsilon)n$, C is not $(s, \frac{8n}{n-s})$-erasure list-decodable. By Lemma 6, this happens only if some $(n-s) \times k$ sub-matrix of G has rank less than $k - \lg(\frac{8n}{n-s})$. Let $\sigma = (n-s)/n$; we have $\varepsilon \le \sigma \le 1/2$. As in the proof of Part (i), the probability that some $(n-s) \times n$ sub-matrix of G has rank less than $k - \lg(8/\sigma)$ is at most

$$\binom{n}{n-s} \cdot k \cdot 2^{(k-n+s)\lg(8/\sigma)} \le k \cdot \left(\frac{e}{\sigma}\right)^{\sigma n} \cdot 2^{\left((\varepsilon/\lg(8/\varepsilon)) - \sigma\right)\lg(8/\sigma)n}$$

[2] We will assume for simplicity that C has dimension k, i.e. G has full column rank, since this happens with very high probability.

$$
\begin{aligned}
&= k \cdot 2^{-n\left(\sigma \lg(8/e) - \frac{\varepsilon \lg(8/\sigma)}{\lg(8/\varepsilon)}\right)} \\
&= 2^{-\Theta(n)}
\end{aligned}
$$

where the last step follows since $\sigma \geq \varepsilon$, and in the first step we used the inequality $\binom{n}{\sigma n} \leq (e/\sigma)^{\sigma n}$ for $\sigma \leq 1/2$.

Now, by a union bound, the probability that for some s, $n/2 \leq s \leq (1-\varepsilon)n$, C is not $(s, \frac{8n}{n-s})$-erasure list-decodable is also exponentially small. Hence there exists a linear code C of block length n and rate $\varepsilon/\lg(8/\varepsilon)$ that is $(s, \frac{8n}{n-s})$-erasure list-decodable for every s that satisfies $n/2 \leq s \leq (1-\varepsilon)n$. Since the list size for $s < n/2$ erasures is at most the list size for $n/2$ erasures, such a code is also $(s, \frac{16n}{n-s})$-erasure list-decodable for every s, $0 \leq s \leq (1-\varepsilon)n$. This proves Part (ii) of the claim (with the choice $A = 8$ and $B = 16$ for the absolute constants). □

Remark: Note that the above lemma not only guarantees the existence of codes with the required properties, but also proves that a random code has these properties with very high (i.e. $1 - o(1)$) probability.

2.3 Upper bounds on $R_L(p)$

We now turn to upper bounds on $R_L(p)$. It is easy to prove that for any fixed L (in fact even for a list size L that is allowed to grow as a polynomial function of the block length), we must have $R_L(p) \leq 1-p$. Indeed, let C be a code of block length n and rate r, and let $T = \{1, 2, \ldots, (1-p)n\}$. Pick a *random* $y \in \{0,1\}^{(1-p)n}$ and consider the set S_y of codewords $c \in C$ that satisfy $[c]_T = y$. The expected number of such codewords equals $2^{rn}2^{-(1-p)n}$, and hence there must exist a $y \in \{0,1\}^n$ for which $|S_y| \geq 2^{(r-(1-p))n}$. Since we want $|S_y| \leq L \leq \mathrm{poly}(n)$, we must have $r \leq (1-p)$. Hence $R_L(p) \leq 1-p$. Below, we are interested in a better upper bound on $R_L(p)$, which in particular bounds it strictly away from $(1-p)$ for every fixed L (and thereby shows that one requires unbounded list size in order to approach the "capacity" $(1-p)$ of the erasure channel). Such a bound is stated in Theorem 7 below. The first upper is a generalization of the Elias-Bassalygo bound for the rate vs. minimum distance trade-off, while the second upper bound is a generalization of the Plotkin bound. These bounds are proved in [8]; see also the account in [26, Sec. V.B].

Theorem 7 ([8]) *For every integer $L \geq 1$ and every p, $0 \leq p \leq 1 - 2^{-L}$, we have*

$$
R_L(p) \leq \min\left\{1 - H(\lambda), 1 - \frac{p}{1-2^{-r}}\right\}, \tag{5}
$$

where $r = 1 + \lfloor \lg L \rfloor$ and where λ is the unique root of the equation $\lambda^{r+1} + (1-\lambda)^{r+1} = 1 - p$ in the range $0 \leq \lambda \leq 1/2$. For $p \geq 1 - 2^{-r}$ (this condition is satisfied if $p \geq 1 - \frac{1}{2L}$), we have $R_L(p) = 0$.

Corollary 8 *For every integer L and every p, $0 < p < 1$, $R_L(p) < 1 - p$.*

The results in [5] improve the bound of Theorem 7 for small values of r (specifically $r \leq 3$). However, their bounds are complicated to state and analytically estimate for larger values of r.

Remark: For $L = 1$, the two bounds proven in Theorem 7 are precisely the Elias-Bassalygo and Plotkin bounds on the rate of a code in terms of its minimum distance, which state that $R(\delta) \leq 1 - H(\frac{1-\sqrt{1-2\delta}}{2})$ and $R(\delta) \leq 1 - 2\delta$, respectively. For large L, the bound $R_L(p) \leq 1 - p/(1-2^{-r})$ is better than the other bound $R_L(p) \leq 1 - H(\rho)$ *except for* very small p (less than about $r/2^r$).

Now, imagine the situation $p = 1 - \varepsilon$ for some very small $\varepsilon > 0$ (i.e. we want to decode from a very large fraction of erasures). For this case, we have the following corollary that follows immediately from Theorem 7.

Corollary 9 *Let $\varepsilon > 0$ and $\mathcal{C}^{\mathrm{lin}}$ be a binary linear code family of positive rate with $\mathsf{ELDR}_L(\mathcal{C}^{\mathrm{lin}}) \geq 1 - \varepsilon$. Then $L = \Omega(1/\varepsilon)$.*

Note that the above bound is optimal up to constant factors, since Lemma 5 implies the existence of linear codes of positive rate which can be decoded from a fraction $(1 - \varepsilon)$ of erasures using lists of size $O(1/\varepsilon)$.

Subsequent to our work, Guruswami and Indyk [17] investigate constructions of good non-linear codes for list decoding from erasures. One of their results shows that there *exist* non-linear binary code families of positive rate (in fact, rate $\Omega(\varepsilon)$), which can be list decoded from a fraction $(1 - \varepsilon)$ of erasures using lists of size only $O(\log(1/\varepsilon))$. Again, the bound on list size is optimal up to constant factors since a bound similar to Theorem 7 also holds for general (not necessarily linear) codes provided one replaces all occurrences of r in the bounds by L (this fact follows from the proofs in [8]). Moreover, in light of Corollary 9, this implies a *provable separation* between the asymptotic performance of linear and non-linear codes for erasure list-decodability for certain settings of parameters. Further details on this can be found in [17] or [14, Chap. 10].

3 Application to Concatenated Codes

3.1 Main code construction

We now use the codes guaranteed by Lemma 5 (specifically Part (ii) of the lemma) as inner codes in a suitable concatenation scheme to construct binary code families of rate $\Omega(\varepsilon^2/\lg(1/\varepsilon))$ that can be efficiently list decoded using lists of size $O(1/\varepsilon)$ when up to a fraction $(1-\varepsilon)$ of its symbols are erased. This follows from the following theorem (Theorem 10), which is the main result of this section. The proof uses the simple, yet powerful, tool of code concatenation [12]. Given an $[N, K, D]_{2^m}$ code C_1, and an $[n, m, d]_2$ code C_2, the concatenated code $C_1 \oplus C_2$ is an $[Nn, Km, \geq Dd]_2$ binary linear code whose codewords are obtained by encoding a message x first using C_1, and then encoding each of the symbols of $C_1(x)$ using C_2 (note that each symbol of $C_1(x)$ is an element of $\mathrm{GF}(2^m)$ and can hence be interpreted as an m-bit vector, and m is exactly the dimension of C_2).

Theorem 10 *For every $\varepsilon > 0$, there exists a family of binary linear codes of rate $\Omega(\varepsilon^2/\log(1/\varepsilon))$ such that a code of block length in the family can be constructed in $2^{\varepsilon^{-O(1)}} + \mathrm{poly}(N, 1/\varepsilon)$ time, and can be list decoded in polynomial time using lists of size $O(1/\varepsilon)$ when up to a fraction $(1 - \varepsilon)$ of its symbols are erased.*

The above result follows immediately from the following lemma.

Lemma 11 *There exist absolute constants b, d such that for all large enough integers K and all small enough $\varepsilon > 0$, there exists a binary linear code C_K that has the following properties:*

(i) C_K has dimension K and block length $N \leq \frac{bK\log(1/\varepsilon)}{\varepsilon^2}$.

(ii) The generator matrix of C_K can be constructed in $2^{\varepsilon^{-O(1)}} + \mathrm{poly}(N, 1/\varepsilon)$ time.

(iii) C_K *is* $((1-\varepsilon)N, \frac{d}{\varepsilon})$*-erasure list-decodable (and since* C_K *is linear there is an* $O(N^3)$ *time list decoding algorithm to decode up to* $(1-\varepsilon)N$ *erasures using lists of size* $O(1/\varepsilon)$*).*

Proof: The code C_K is constructed by concatenating an outer code C_{out} over $\mathrm{GF}(2^m)$ of block length n_0, dimension $k_0 = K/m$ and minimum distance d_0, with an inner code C_{in} as guaranteed by Part (ii) of Lemma 5. By using a construction in [2], we can pick parameters so that $k_0 = K/m = \Omega(\varepsilon n_0)$, $m = O(1/\varepsilon)$ and $d_0 = (1 - O(\varepsilon))n_0$ (for convenience, we hide constants using big-Oh notation, but we stress that these are absolute constants that do not depend on ε).[3] We note that this choice of C_{out} is not linear over $\mathrm{GF}(2^m)$ (though it *is* additive), but a (succinct) description of C_{out} can be constructed in time $\mathrm{poly}(n, 1/\varepsilon)$. Moreover, after concatenation with an inner binary linear code, the overall concatenated code will be a binary linear code.

The inner code C_{in} will be a code as guaranteed by Lemma 5 of dimension m and block length $n_1 = O(\frac{m \lg(1/\varepsilon)}{\varepsilon})$ that is $((1-\sigma)n_1, B/\sigma)$-erasure list-decodable for every $\sigma \geq \varepsilon/2$. We can construct such a code by a brute-force search in $2^{O(mn_1)} = 2^{\varepsilon^{-O(1)}}$ time. Once we have descriptions of C_{out} and C_{in}, one can construct the generator matrix of the concatenated code C_K in $\mathrm{poly}(N, 1/\varepsilon)$ time where $N = n_0 n_1$ is the block length of C_K. Now $N = n_0 \cdot n_1 = O(\frac{K}{m\varepsilon}) \cdot O(\frac{m \lg(1/\varepsilon)}{\varepsilon}) = O(\frac{K \lg(1/\varepsilon)}{\varepsilon^2})$. This proves Parts (i) and (ii) of the statement of the lemma.

We now prove that C_K is $((1-\varepsilon)N, O(1/\varepsilon))$-erasure list-decodable. Let y be a received word with $(1-\varepsilon)N$ erasures, and let $c'_1, c'_2, \ldots, c'_M$ be the codewords in C_K "consistent" with y, and let c_j be the codeword of C_{out} corresponding to c'_j, for $1 \leq j \leq M$. We wish to prove that $M = O(1/\varepsilon)$. Let y_i be the portion of y corresponding to the i'th outer codeword position, for $1 \leq i \leq n_0$. Let the number of symbols in y_i be $\sigma_i n_1$. We have $\sum_i \sigma_i = \varepsilon n_0$. Define $Q = \{i : \sigma_i \geq \varepsilon/2\}$. Clearly we have

$$\sum_{i \in Q} \sigma_i \geq \frac{\varepsilon n_0}{2} \ . \tag{6}$$

Now define "weights" $w_{i,\beta}$ for $1 \leq i \leq n_0$ and $\beta \in \mathrm{GF}(2^m)$ as follows. If $i \notin Q$, set $w_{i,\beta} = 0$ for all $\beta \in \mathrm{GF}(2^m)$. If $i \in Q$, then $\sigma_i \geq \varepsilon/2$ and hence by construction C_{in} is $((1-\sigma_i)n_1, B/\sigma_i)$-erasure list-decodable. In other words, if $J_i = \{\beta : C_{\text{in}}(\beta) \text{ is consistent with } y_i\}$, then $|J_i| \leq B/\sigma_i$. We set (for $i \in Q$):

$$w_{i,\beta} = \begin{cases} \sigma_i & \text{if } \beta \in J_i \\ 0 & \text{if } \beta \notin J_i \end{cases}$$

Since $|J_i| \leq B/\sigma_i$, our choice of weights clearly satisfy

$$\sum_{i,\beta} w_{i,\beta}^2 \leq B \sum_{i \in Q} \sigma_i \ . \tag{7}$$

We now use a combinatorial result that gives an upper bound on the number of codewords of C_{out} that satisfy a certain weighted condition depending on the distance d_0 of C_{out}. This result is a "weighted" version of the classical "Johnson bounds", and appears, for example, in [20] and [14, Section 3.4]. Specifically, this result states that, for any $\varepsilon' > 0$, the number of codewords $(\alpha_1, \alpha_2, \ldots, \alpha_{n_0}) \in C_{\text{out}}$ that satisfy

$$\sum_{i=1}^{n_0} w_{i,\alpha_i} \geq \Big((n_0 - d_0(1-\varepsilon')) \sum_{i,\beta} w_{i,\beta}^2 \Big)^{1/2} \tag{8}$$

[3] Actually, we can use certain families of algebraic-geometric codes and even have $m = O(\log(1/\varepsilon))$. But we prefer to use the codes from [2] as they are easier to construct and have lower construction complexity, and we do not want to give the impression that we need complicated AG-codes for our construction. We will highlight the improvement in performance attainable using AG-codes in Section 3.2.2.

is at most $1/\varepsilon'$.

Now for each c_j, $1 \le j \le M$, and each $i \in Q$, we have $c_{j,i} \in J_i$, where $c_{j,i} \in \mathrm{GF}(2^m)$ denotes the i'th symbol of c_j. Now, $w_{i,c_{j,i}} = \sigma_i$ for every $i \in Q$ and $w_{i,c_{j,i}} = 0$ for $i \notin Q$. Thus, we have

$$\sum_{i=1}^{n_0} w_{i,c_{j,i}} = \sum_{i \in Q} \sigma_i \ . \tag{9}$$

Combining Equations (6), (7) and (9), we have that the codeword c_j satisfies Condition (8) as long as

$$\frac{\varepsilon n_0}{2} \ge (n_0 - d_0(1-\varepsilon'))B \ ,$$

which can be satisfied provided $d_0 \ge n_0(1 - \frac{\varepsilon}{2B})(1-\varepsilon')^{-1}$. Picking $\varepsilon' = \frac{\varepsilon}{4B}$, we only need $d_0 = n_0(1 - O(\varepsilon))$ which is satisfied for our choice of C_{out}. Hence the number of codewords consistent with the received word y is at most $1/\varepsilon' = O(1/\varepsilon)$, thus proving Part (iii) of the lemma as well. $\Box$ *(Lemma 11)*

Remark: The previous results from [19] (obtained by appealing to Lemma 2 for code constructions from [2, 3]) along the lines of the above theorem achieved a block length of $N = \min\{O(\frac{K^2}{\varepsilon^2}), O(\frac{K}{\varepsilon^3})\}$. Thus, our result achieves the best features of both these bounds and gets $N = \tilde{O}(K/\varepsilon^2)$ (hiding the $\lg(1/\varepsilon)$ factor).

3.2 Some improvements

3.2.1 Obtaining near-linear encoding and near-quadratic decoding times

The codes constructed in Lemma 11 being linear can be encoded in $O(N^2)$ time and can be list decoded from a fraction $(1-\varepsilon)$ of erasures in $O(N^3)$ time by solving a linear system. By using Reed-Solomon codes as outer code, the encoding can be accomplished in $O(N \log^{O(1)} N)$ time. This is because the outer encoding can be performed using $O(N \log N)$ field operations by employing FFT based methods, and then the inner encoding just takes $O(\log^2 N)$ time for each of the N outer codeword symbols. For decoding, the inner codes can be decoded in $O(\log^3 N)$ time by solving a linear system, and then the Reed-Solomon code can be decoded from a set of weights that satisfy Condition (8) in near-quadratic time using fast implementations [25, 24] of the Reed-Solomon decoding algorithm in [18]. There is, however, a potential problem in using Reed-Solomon codes as outer codes. In such a case, the inner code needs to be a linear code of dimension $\Omega(\log N)$, and hence a brute force search for the inner code will take time which has a quasi-polynomial dependence on N. Therefore, if we seek a deterministic construction, then the construction time no longer appears to be polynomial in N.

Nevertheless, there is a way to obtain a polynomial time construction of the necessary inner code. A random linear code will have the erasure list-decodability property required by the inner code with high probability (cf. Lemma 5 and the remark following it). This randomized construction can be derandomized using the method of conditional expectations in time significantly better than that required for brute-force search over all possible codes. We only sketch the details below — the reader can find a discussion of the method of conditional expectations, for example, in [4, Chap. 15].

To pick an $[n,k]_2$ linear code, we pick the n rows of the $n \times k$ generator matrix G in sequence, each time going through all possible choices in $\mathbb{F}_2^k$ and picking the one which minimizes a certain conditional expectation. The relevant conditional expectation in question concerns the expected number of $t \times k$ sub-matrices of G which have rank at most $k - \log(n/t) - c$, summed over all t, $\varepsilon n \le t \le n$ (here c is a large enough *absolute* constant). Lemma 6 implies that finding a generator matrix where there are no such sub-matrices is equivalent to the construction of an $[n, k]_2$ linear code that satisfies the requirement of Lemma 5, Part (ii). To be able to compute the relevant conditional expectations easily, we will introduce indicator random variables whose expectation provides a pessimistic estimation of the true expectation, and compute their expectations instead.

We will introduce an indicator random variable $I(T, S)$ for each $T \subseteq [n]$ with $|T| = t \ge \varepsilon n$ and each $S \subseteq \mathbb{F}_2^k$ of linearly independent vectors with $|S| = \log(n/t) + c$. We will define $I(T, S) = 1$ if the span of the rows of the generator matrix G that are indexed by T is contained in the null space of the span of S, and $I(T, S) = 0$ otherwise. Thus if $I(T, S) = 1$, then the $|T| \times k$ sub-matrix of G corresponding to rows in T has rank at most $k - |S|$, and provides a "counterexample" to the code constructed having the desired erasure list-decodability property. The event "$I(T, S) = 1$" is therefore one that we would like to avoid, for every choice of T, S. It is easy to see that for each fixed T, S, the conditional expectation of $I(T, S)$ after the first few rows of G have been fixed (taken over the random choice of the remaining rows) can be computed exactly, since it is simply the probability that each of the rows in T that have not been fixed yet lie in the null space of S.

Now, define the random variable $X = \sum_{T,S} I(T, S)$. By linearity of expectation, the conditional expectation of X can be computed exactly for each choice of a subset of rows of G. The initial expectation of X, say $\mathbf{E}$, is exponentially small in n provided $k = \Theta(\varepsilon n / \log(1/\varepsilon))$ (this follows from the proof of Part (ii) of Lemma 5).

Hence the above derandomization procedure will find a matrix G whose X value is at most $\mathbf{E}$, and since for each fixed G, the random variable X takes on an integer value, we must have $X(G) = 0$. This implies in turn that G has the required rank property for its sub-matrices, as desired.

The runtime of this derandomization procedure is dominated by the time to consider all possible candidates for the sets T and S above. Therefore, it takes

$$2^{O(H(\varepsilon)n)} \cdot 2^{O(k \log(1/\varepsilon))} = 2^{O(\varepsilon \log(1/\varepsilon) n)} = 2^{O(k \log^2(1/\varepsilon))}$$

time (since $k = \Theta(\varepsilon n / \log(1/\varepsilon))$.)

Applying this to the context of concatenation with an outer Reed-Solomon code where the dimension $k = O(\log N)$, we conclude that the generator matrix of a linear inner code that has the required properties can be computed in $N^{O(\log^2(1/\varepsilon))}$ time. This is dominant component in the construction of the concatenated code, and therefore the overall code can be constructed in $N^{O(\log^2(1/\varepsilon))}$ time.

To summarize, we can prove the result of Lemma 11 with codes that have a near-linear time encoder, a quadratic time list decoder, and an $N^{O(\log^2(1/\varepsilon))}$ time deterministic construction.

3.2.2 Improving construction time using AG-codes

Despite the impressive encoding and decoding times, the drawback of the above construction with Reed-Solomon codes as the outer code is the rather high exponent of N in the construction time. In particular, the code family ceases to be *uniformly constructive*, since the construction time is no longer of the form $O(f(\varepsilon)N^a)$ where a is an absolute constant independent of ε.

Instead of Reed-Solomon codes, we can use AG-codes of distance $(1 - O(\varepsilon))$ as the outer code. Such codes of rate $\Omega(\varepsilon)$ exist over an alphabet size of $O(1/\varepsilon^2)$ (these are codes which meet the Drinfeld-Vlădut bound, cf. [27, 13]). Now, the dimension k of the inner binary code is only $O(\log(1/\varepsilon))$. Applying the same derandomization procedure discussed above, we can now deterministically find a binary linear inner code which has the required erasure list-decodability properties in $2^{O(\log^3(1/\varepsilon))}$ time.

Since the overall code is linear, the claims of quadratic encoding time and cubic decoding time still holds. Thus, we can obtain the same properties as the codes from Lemma 11 with an improved construction time of $2^{O(\log^3(1/\varepsilon))} + \mathrm{poly}(N, 1/\varepsilon)$. We formally record this improvement below.

Theorem 12 *There exist absolute constants b, d such that for all large enough integers K and all sufficiently small $\varepsilon > 0$, there exists a binary* linear *code C_K that has the following properties:*

(i) C_K has dimension K and block length $N \le \frac{bK\lg(1/\varepsilon)}{\varepsilon^2}$.

(ii) The generator matrix of C_K can be constructed in $2^{O(\log^3(1/\varepsilon))} + \mathrm{poly}(N, 1/\varepsilon)$ time.

(iii) C_K is $((1-\varepsilon)N, \frac{d}{\varepsilon})$-erasure list-decodable. In other words, the $N \times K$ generator matrix G of C_K has the property that every $(\varepsilon N \times K)$ sub-matrix has rank at least $(k - \lfloor \lg(d/\varepsilon) \rfloor)$. Also, since since C_K is linear, there is an $O(N^3)$ time decoding algorithm to list decode it from up to $(1-\varepsilon)N$ erasures.

3.3 The ε^2 "rate barrier" for binary linear codes

We now present a connection between good erasure list-decodable binary linear codes and certain bipartite Ramsey graphs. This indicates that improving the rate in Theorem 12 to, say ε^a for some $a < 2$, without worsening the other parameters significantly, is likely to be very difficult. This connection was pointed out to us by Noga Alon [1].

For a Boolean matrix, a sub-matrix is said to be *monochromatic* if either all its entries equal 0 or all of them equal 1. The bipartite Ramsey problem asks for a construction of $n \times n$ Boolean matrices which have no $p \times p$ monochromatic sub-matrices (where p is a certain function of n), for infinitely many values of n. Such a matrix has an obvious connection to $n \times n$ bipartite graphs which have no complete bipartite subgraph $K_{p,p}$ or its complement as an induced subgraph. Hence the terminology "bipartite Ramsey problem" is used to refer to this problem. A straightforward application of the probabilistic method shows that, for all large enough n, there exist such matrices which achieve $p = O(\log n)$. However, an explicit or polynomial time construction of such a matrix is a much harder task. The best known polynomial time constructions of $n \times n$ 0-1 matrices only rule out the existence of $p \times p$ monochromatic sub-matrices for p about $\sqrt{n}$. A polynomial time construction with p much smaller than $\sqrt{n}$ is a folklore open problem. The following result shows a connection between certain erasure list-decodable codes and bipartite Ramsey graphs. It shows that achieving a rate of $\Omega(\varepsilon^a)$ for some $a < 2$ for our linear code construction from the previous sections, will, under some conditions, imply an improvement to the "$\sqrt{n}$ bound" for the bipartite Ramsey problem.

Proposition 13 (Alon) *Assume that for every large enough integer k and every $\varepsilon > 0$, there exists an $[n,k]_2$ linear code C of rate $k/n = \varepsilon^a$, $a > 1$, such that (a) the generator matrix of C can be constructed in $O(n^c f(\varepsilon))$ time where c is an absolute constant and f is an arbitrary real-valued function, and (b) C is $((1-\varepsilon)n, 1/\varepsilon^{O(1)})$-erasure list-decodable. Then, for any $\gamma > 0$, for infinitely many n, there exists an $O(n^c f(n^{-(1/a-\gamma)}))$ time construction of an $n \times n$ matrix over $\mathbb{F}_2$ which has no monochromatic $p \times p$ sub-matrix, i.e., no $p \times p$ sub-matrix that consists of all 0's or all 1's, for $p = n^{1-1/a+\gamma}$.*

Proof: For a large enough integer n and small enough $\gamma > 0$, pick $\varepsilon = n^{-(1/a-\gamma)}$. Then by the hypothesis, there exists an $[n,k]_2$ linear code C with the claimed properties for $k = \varepsilon^a n = n^{a\gamma}$. Let A be the $n \times k$ generator matrix of C. Define $p = \varepsilon n = n^{1+\gamma-1/a}$. By Lemma 6, the assumed erasure list-decodability property of C implies that the rank of each $p \times k$ sub-matrix of A is at least $(k - O(\log 1/\varepsilon)) = k - O(\log n)$. Define j to be the smallest integer for which the number of subsets of $\{1,2,\ldots,k\}$ of size exactly j is at least n. Since $k = n^{a\gamma}$, we have $j \le \frac{2}{a\gamma}$ (for large enough n). .

Construct an $n \times n$ matrix B by having its columns be all the linear combinations (over $\mathbb{F}_2$) of exactly j distinct columns of A. (This might create more than n columns since the number of such combinations could exceed n; in this case we just pick a subset of n linear combinations arbitrarily.)

We will prove that B has no $p \times p$ monochromatic sub-matrix consisting of all 0's or all 1's. Suppose not, and let B have a $p \times p$ sub-matrix consisting of only 0's (the case of all 1's can be dealt with similarly). Let I be the set of rows involved in this sub-matrix, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be the restrictions of the columns of A to the rows in I. Let M be the $p \times k$ sub-matrix of A indexed by the rows in I.

By assumption about the $p \times p$ all 0s sub-matrix, we have at least p linear combinations, each of exactly j of the vectors $\mathbf{v}_i$, all giving the all 0's vector. This implies that the null space of M (i.e. the space $\{\mathbf{x} \in \mathbb{F}_2^k : M\mathbf{x} = \mathbf{0}\}$) has at least p non-zero vectors of Hamming weight exactly j. It is an easy exercise to show that the dimension of such a space must be at least $p^{1/j}$. Hence we have

$$\mathrm{rank}(M) \le k - p^{1/j} = k - n^{(1+\gamma-1/a)/j} \le k - n^{(a\gamma(1+\gamma)-\gamma)/2} \ , \tag{10}$$

since $j \le 2/(a\gamma)$. On the other hand, by hypothesis we have that every $p \times k$ sub-matrix of A has rank at least $(k - O(\log n))$ and hence we must have $\mathrm{rank}(M) \ge k - O(\log n)$. This contradicts Equation (10) for large enough n.

Hence the matrix B has no $p \times p$ monochromatic sub-matrix as claimed. The claimed construction time for B follows easily from the choice of parameters, and the proof is complete. □

Corollary 14 *Suppose we could prove the result of Theorem 12 with a rate of $\Omega(\varepsilon^a)$ for some $1 \le a < 2$. Then there exists a quasi-polynomial time construction of $n \times n$ matrices over $\{0,1\}$ which have no monochromatic $n^b \times n^b$ sub-matrix for some $b < 1/2$.*

Proof: Pick b so that $1 - 1/a < b < 1/2$ (this is possible since $a < 2$), and let $\gamma = b - (1 - 1/a)$. For $\varepsilon = n^{-(1/a-\gamma)} = n^{b-1}$, the construction time of the code guaranteed by Theorem 12 is $2^{O(\log^3 n)}$, and hence by Proposition 13, we have a $2^{O(\log^3 n)}$ time construction of an $n \times n$ 0-1 matrix with no monochromatic $p \times p$ sub-matrix where $p = n^{1-1/a+\gamma} = n^b$. □

As remarked earlier, a polynomial time construction of bipartite Ramsey graphs with p much smaller than $\sqrt{n}$ is a folklore open problem. It is not known how to achieve p significantly smaller than $\sqrt{n}$ even if one allows quasi-polynomial time to construct the matrix. Therefore, Corollary 14 indicates that improving the rate in our construction from Theorem 12 to sub-quadratic in ε is likely to be quite difficult.[4]

There is, however, some hope to beat the "ε^2 barrier" without having to confront some major open problem in constructive Ramsey theory. For our coding applications we have always been thinking of ε as a small positive *constant* (independent of n), and hence we allow for a construction that runs in time exponential in $1/\varepsilon$ and/or a list size that is exponential in $1/\varepsilon$. It is an interesting

[4]On the other hand, optimistically speaking this provides further motivation to try and improve the ε^2 rate in Theorem 12!

open question whether this can be exploited to improve the rate beyond ε^2. Also, the results of the next section get a rate better than ε^2 by using a slightly larger alphabet size. They also critically exploit non-linearity. It will also be interesting to get a better than ε^2 rate for *binary* codes using some non-linear construction. Such a construction would escape the confines of the relation to bipartite Ramsey graphs, and thus may not be very hard to obtain.

4 Concluding Remarks

In this paper, we studied erasure list-decodable *binary* codes. If one allows larger alphabets, known constructions of large distance algebraic-geometric codes imply code families of ELDR $(1-\varepsilon)$ and rate $\Omega(\varepsilon)$ over an alphabet of size $O(1/\varepsilon^2)$, even for a list size of 1. Explicit constructions of linear codes of rate better than $\Omega(\varepsilon^2)$ with ELDR $(1-\varepsilon)$ even for a slightly smaller alphabet do not appear to be known. However, in recent work, Guruswami and Indyk [17] present constructions of *non-linear* codes over large alphabets (of size *independent* of ε) that can be efficiently list decoded from a fraction $(1-\varepsilon)$ of erasures and which have rate better than $\Omega(\varepsilon^2)$ (in fact they have rate approaching the optimal bound of $\Omega(\varepsilon)$ as the alphabet size grows).

It is also an interesting question whether a code construction similar to Theorem 12 can be obtained in time which is polynomial in *both* N and $1/\varepsilon$.

Acknowledgments

I would like to thank Piotr Indyk, Dan Spielman and Madhu Sudan for several useful discussions. I am grateful to Noga Alon for pointing out the connection to Ramsey graphs outlined in Section 3.3 and for kindly permitting me to include the results of Proposition 13 and Corollary 14 in this paper. I would like to thank Ralf Koetter for bringing to my attention the literature on Generalized Hamming weights.

References

[1] N. Alon. Personal Communication, October 1999.

[2] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Trans. on Information Theory*, **38** (1992), pp. 509-516.

[3] N. Alon, O. Goldreich, J. Håstad and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, **3** (1992), pp. 289-304.

[4] N. Alon and J. Spencer. *The Probabilistic Method.* John Wiley and Sons, Inc., 1992.

[5] A. Ashikhmin, A. Barg, and S. Litsyn. New upper bounds on generalized weights. *IEEE Transactions on Information Theory*, **45**(4) (1999), pp. 1258-1263.

[6] V. M. Blinovsky. 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.

[7] B. Bollobás. *Combinatorics.* Cambridge University Press, Cambridge, U.K, 1986.

[8] G. Cohen, S. Litsyn, and G. Zémor. Upper bounds on generalized distances. *IEEE Transactions on Information Theory*, **40** (1994), pp. 2090-2092.

[9] P. Elias. Coding for two noisy channels. *Information Theory*, Third London Symposium, September 1955, Buttersworth's Scientific Publications, pp. 61-76.

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

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

[12] G. D. Forney. *Concatenated Codes*. MIT Press, Cambridge, MA, 1996.

[13] A. Garcia and H. Stichtenoth. A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlădut bound. *Inventiones Mathematicae*, 121:211–222, 1995.

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

[15] V. Guruswami, J. Håstad, M. Sudan and D. Zuckerman. Combinatorial Bounds for List Decoding. *IEEE Transactions on Information Theory*, May 2002, to appear. Perliminary version in the *Proceedings of the 38th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, October 2000.

[16] V. Guruswami and P. Indyk. Expander-based Constructions of Efficiently Decodable Codes. *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 658-667, October 2001.

[17] V. Guruswami and P. Indyk. Near-optimal linear-time codes for unique decoding and new list decodable codes over smaller alphabets. *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, May 2002, to appear.

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

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

[20] V. Guruswami and M. Sudan. Extensions to the Johnson Bound. Manuscript, 2001. Available from `http://theory.lcs.mit.edu/~madhu/npapers.html`.

[21] T. Helleseth, T. Klϕve, V. I. Levenshtein, and O. Ytrehus. Bounds on minimum support weights. *IEEE Trans. Inform. Theory*, **41** (2) (1995), pp. 432-440.

[22] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier/North-Holland, Amsterdam, 1981.

[23] R. J. McEliece, E. R. Rodemich, H. Rumsey Jr., and L. R. Welch. New upper bounds on the rate of a code via the Delsarte-Macwilliams inequalities. *IEEE Transactions on Information Theory*, **23** (1997), pp. 157–166.

[24] R. R. Nielsen and T. Hϕholdt. Decoding Reed-Solomon codes beyond half the minimum distance. *Proceedings of the International conference on Coding Theory and Cryptography*, 1998.

[25] R. Roth and G. Ruckenstein. Efficient decoding of Reed-Solomon codes beyond half the minimum distance. *IEEE Transactions on Information Theory*, **46**(1), pp. 246-257, January 2000.

[26] M. A. Tsfasman and S. G. Vlădut. Goemetric approach to higher weights. *IEEE Transactions on Information Theory*, **41** (1995), pp. 1564-1588.

[27] 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.

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

[29] V. Wei. Generalized Hamming weights for linear codes. *IEEE Trans. Inform. Theory*, **37**(5) (1991), pp. 1412-1418.

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

[31] 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.