List Decoding from Erasures: Bounds and Code Constructions

Venkatesan Guruswami

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. Such bounds already exist in the literature, albeit under the label of Generalized Hamming Weights, and we make their connection to list decoding from erasures explicit. Our bounds 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 $O(1/\varepsilon)$ when an adversarially chosen $(1-\varepsilon)$ fraction of symbols are erased, for arbitrary $\varepsilon > 0$. This improves previous results in this vein, which achieved a rate of $\Omega(\varepsilon^3 \log(1/\varepsilon))$.**

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

I. 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], [29]. 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], [16], [19]).

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 [6], [11], [16], [30]. Though the case of erasures is algorithmically always easier to handle, it raises several interesting combinatorial questions which 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.

Venkatesan Guruswami's address is Department of Computer Science & Engineering, University of Washington, Seattle, WA 98195. Email: venkat@cs.washington.edu. Most of this work was done while the author was at the Massachusetts Institute of Technology supported by NSF CCR 9875511, and at University of California, Berkeley, supported by a Miller Research Fellowship.

A. *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

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

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 largest rate possible for a linear code family with list-of-L erasure decoding radius at least p, denoted $R_L(p)$, is defined as

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

B. 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-linear time from the obvious cubic time required to solve a linear system.

C. 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 $\text{supp}(D) = \{i : \exists\ \langle x_1, x_2, \ldots, x_n\rangle \in D,\ x_i \neq 0\}$.

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 [28] 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 [25] 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 [28]), 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.

Before we state the lemma, some remarks about our notation. 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. For $0 \leq x \leq 1$, $H(x) = -x\log x-(1-x)\log(1-x)$ denotes the binary entropy function of x.

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 $|\text{supp}(C')| > e$. Consider the received word $\mathbf{r}$ that has all positions in $\text{supp}(C')$ 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 $\text{supp}(C') > e$.

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 $\mathbf{c}_{|T} = \mathbf{r}_{|T}$ (here $\mathbf{x}_{|T}$ denotes the projection of $\mathbf{x}$ onto coordinates in T). 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 $\text{supp}(D) \cap T = \emptyset$, and hence $|\text{supp}(D)| \leq e$. 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 readers feel the need, they should have no difficulty in translating them into the language of generalized Hamming weights.

D. Main Results

D.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]. Therefore, 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], [30], [6], [16].

D.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 [3]). 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 [20] 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.[2] The best construction of such high distance codes achieves a rate of $\Omega(\varepsilon^3)$ (see [3]). 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-linear time.

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

[2]More generally, any binary code of relative distance δ has a small (linear in block length) number of codewords for up to a fraction 2δ of erasures, and this is the best possible in the sense that for every $\gamma > 0$, there exist linear codes C_γ of relative distance δ, and a received word y with a fraction $(2+\gamma)\delta$ of erasures, such that C_γ has exponentially many codewords consistent with y.

II. Combinatorial Bounds on Erasure List-decodability

A. Lower Bounds on $R_L(p)$

The connection to generalized Hamming weights from Lemma 1, together with existing lower bounds in the literature for the rate as a function of generalized Hamming weights [28], [22], [25], gives the following when stated in our notation. In particular, the result below follows from Theorem 8 in [22].

Theorem 2: For every integer $L \geq 1$ and every p, $0 \leq p \leq 1$, we have

$$R_L(p) \geq 1 - \frac{p}{r}\lg(2^r - 1) - \frac{H(p)}{r} \qquad (3)$$

where $r = \lceil \lg(L+1) \rceil$.

For our purposes, in Section III, we would like to use the codes guaranteed by the above result as inner codes in suitable concatenated schemes. The following lemma focuses on the parameter range that we will be interested in and also guarantees a "gradual" increase in list size as more and more codeword symbols are erased, up to a fraction $(1-\varepsilon)$ of erasures. Though such a result can be obtained by going through the proofs in the literature (eg. that of Theorem 2), the result 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. Also, details of the proof will come in handy when we try to "derandomize" this construction in Section III-B.

Lemma 3: There exists absolute constants $A, B > 0$ such that for every ε, $0 \leq \varepsilon \leq 1/2$, for all sufficiently large n, 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* $s \leq (1-\varepsilon)n$. Furthermore, a random $[n,k]$ binary linear code has this erasure list-decodability property with overwhelming (specifically, $1 - 2^{-\Omega(n)}$) probability.

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

Lemma 4: 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 $T \subseteq \{1, 2, \dots, n\}$ 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 $T \subseteq \{1, \dots, n\}$ with $|T| = n - s$, G_T has rank at least $k - \lfloor \lg L \rfloor$. ■

Proof of Lemma 3: 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.[3] We will prove that except with probability $2^{-\Omega(n)}$, such a random code is $(s, \frac{8n}{n-s})$-erasure list-decodable for every s, $n/2 \le s \le (1 - \varepsilon)n$. Indeed, let use fix an s between $n/2$ and $(-1\varepsilon)n$ and estimate the probability that C is not $(s, \frac{8n}{n-s})$-erasure list-decoable. By Lemma 4, 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 $\varepsilon \le \sigma \le 1/2$. For a fixed $(n - s) \times k$ submatrix of G and any $J \le k$, the probability that it has rank $(k - J)$ is at most $2^{kJ} 2^{-(n-s)J}$. This follows since for a fixed subspace S of $\mathbb{F}_{\not\models}^{\urcorner}$ of dimension $(k - J)$, the probability that all t rows of M lie in S is at most 2^{-Jt}, and furthermore the number of subspaces of $\mathbb{F}_{\not\models}^{\urcorner}$ 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 $\mathbb{F}_{\not\models}$. By a union bound the probability that a fixed $(n - s) \times k$ submatrix of G has rank at most $(k - J')$ is at most $k 2^{(k-n+s)J'}$.

Therefore the probability that *some* $(n - s) \times k$ submatrix of G has rank less than $k - \lg(8/\sigma)$ is at most

$$\binom{n}{n-s} \cdot k \cdot 2^{(k-n+s)\lg(8/\sigma)}$$

[3] We will assume for simplicity that C has dimension k, i.e. G has full column rank, since for k much smaller than n, as it is on our case, this happens with very high probability.

which is bounded from above by

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

This leads to an upper bound of

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

on the desired probability.

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

Now, applying the union bound again, the probability that for some s, $n/2 \le s \le (1 - \varepsilon)n$, 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 $n/2 \le s \le (1 - \varepsilon)n$. 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, $0 \le s \le (1 - \varepsilon)n$. This proves the claim of the lemma (with the choice $A = 8$ and $B = 16$ for the absolute constants). ■

B. *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 $R_L(p) \le 1 - p$. Indeed, let C be a code of block length n and rate r, and let $T = \{1, 2, \dots, (1 - p)n\}$. 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 $|S_y| \ge 2^{(r-(1-p))n}$. Since we want $|S_y| \le L \le \mathrm{poly}(n)$, we must have $r \le (1 - p)$. Hence $R_L(p) \le 1 - p$. 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 5 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 [25, Sec. V.B].

Theorem 5 ([8]) For every integer $L \ge 1$ and every p, $0 \le p \le 1 - 2^{-L}$, we have

$$R_L(p) \le \min\left\{1 - H(\lambda), 1 - \frac{p}{1 - 2^{-r}}\right\} , \qquad (4)$$

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

$0 \le \lambda \le 1/2$. For $p \ge 1 - 2^{-r}$ (this condition is satisfied if $p \ge 1 - \frac{1}{2L}$), we have $R_L(p) = 0$.

Corollary 6: For every integer L and every p, $0 < p < 1$, $R_L(p) < 1 - p$.

The results in [5] improve the bound of Theorem 5 for small values of r (specifically $r \le 3$). 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 5 are precisely the Elias-Bassalygo and Plotkin bounds on the rate of a code in terms of its minimum distance, which state that $R(\delta) \le 1 - H(\frac{1-\sqrt{1-2\delta}}{2})$ and $R(\delta) \le 1 - 2\delta$, respectively. For large L, the bound $R_L(p) \le 1 - p/(1 - 2^{-r})$ is better than the other bound $R_L(p) \le 1 - H(\rho)$ *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 5.

Corollary 7: Let $\varepsilon > 0$ and $\mathcal{C}^{\rm lin}$ be a binary linear code family of positive rate with $\mathsf{ELDR}_L(\mathcal{C}^{\rm lin}) \ge 1 - \varepsilon$. Then $L = \Omega(1/\varepsilon)$.

Note that the above bound is optimal up to constant factors, since Lemma 3 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 [18] 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 5 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 7, 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 [18] or [15, Chap. 10].

III. Application to Concatenated Codes

A. Main code construction

We now use the codes guaranteed by Lemma 3 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 8), which is the main result of this section. The proof uses the simple, yet powerful, tool of code concatenation [13]. Given an $[N, K, D]_{2^m}$ code C_1, and an $[n, m, d]_2$ code C_2, the concatenated code $C_1 \circ C_2$ is an $[Nn, Km, \ge Dd]_2$ 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 8: 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 N 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 9: 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 $N \le \frac{bK\log(1/\varepsilon)}{\varepsilon^2}$.
(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 $C_{\rm out}$ over $\mathrm{GF}(2^m)$ of block length n_0, dimension $k_0 = K/m$ and minimum distance d_0, with an inner code $C_{\rm in}$ as guaranteed by Part (ii) of Lemma 3. By using a construction in [3], 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).[4] We note that this choice of $C_{\rm out}$ is not linear over $\mathrm{GF}(2^m)$ (though it *is* additive), but a (succinct) description of $C_{\rm out}$ can be constructed in time $\mathrm{poly}(n_0, 1/\varepsilon)$. Moreover, after concatenation with an inner binary linear code, the overall concatenated code will be a binary linear code.

The inner code $C_{\rm in}$ will be a code as guaranteed by Lemma 3 of dimension m and block length

[4]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 [3] as they are more elementary, and we do not want to give the impression that we need complicated AG-codes for our construction.

$n_1 = O(\frac{m \lg(1/\varepsilon)}{\varepsilon})$ 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 $C_{\rm out}$ and $C_{\rm in}$, 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 $\mathbf{y}$ 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 $\mathbf{y}$, and let c_j be the codeword of $C_{\rm out}$ 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 $\mathbf{y}$ 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

$$\sum_{i \in Q} \sigma_i \geq \frac{\varepsilon n_0}{2} . \tag{5}$$

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 $C_{\rm in}$ is $((1-\sigma_i)n_1, B/\sigma_i)$-erasure list-decodable. In other words, if $J_i = \{\beta : C_{\rm in}(\beta) \text{ is consistent with } y_i\}$, 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

$$\sum_{i,\beta} w_{i,\beta}^2 \leq B \sum_{i \in Q} \sigma_i . \tag{6}$$

We now use a combinatorial result that gives an upper bound on the number of codewords of $C_{\rm out}$ that satisfy a certain weighted condition depending on the distance d_0 of $C_{\rm out}$. This result is a "weighted" version of the classical "Johnson bounds", and appears, for example, in [21] and [15, Section 3.4]. Specifically, this result states that, for any $\varepsilon' > 0$, the number of codewords $(\alpha_1, \alpha_2, \ldots, \alpha_{n_0}) \in C_{\rm out}$ that satisfy

$$\sum_{i=1}^{n_0} w_{i,\alpha_i} \geq \Big(\big(n_0 - d_0(1-\varepsilon')\big) \sum_{i,\beta} w_{i,\beta}^2 \Big)^{1/2} \tag{7}$$

is at most $1/\varepsilon'$.

Now for each c_j, $1 \leq j \leq M$, 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

$$\sum_{i=1}^{n_0} w_{i,c_{j,i}} = \sum_{i \in Q} \sigma_i . \tag{8}$$

Combining Equations (5), (6) and (8), we have that the codeword c_j satisfies Condition (7) as long as

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

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

Remark: The previous results from [20] 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).

B. Obtaining near-linear encoding and decoding times

The codes constructed in Lemma 9 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 (7) in near-linear time using fast implementations [12], [1] of the Reed-Solomon decoding algorithm in [19]. 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. By Lemma 3, a random linear code will have the erasure list-decodability property required by the inner code with high probability. 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 here — 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 4 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 3, 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 3).

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 at most

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

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 $N^{O(\varepsilon^{-1} \log(1/\varepsilon))}$ time. This is dominant component in the construction of the concatenated code, and therefore the overall code can be constructed in $N^{O(\varepsilon^{-1} \log(1/\varepsilon))}$ time.

To summarize, we can prove the result of Lemma 9 with codes that have a *near-linear time* encoder and list decoder, and an $N^{O(\varepsilon^{-1} \log(1/\varepsilon))}$ time deterministic construction.

C. *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 8 to, say ε^a for some $a < 2$, with a very efficient (say $\mathrm{poly}(N, 1/\varepsilon)$) construction time is likely to be difficult. This connection was pointed out to us by Noga Alon [2].

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 10 (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 4, 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 $\{1,2,\dots,k\}$ 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 submatrix 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 $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k}$ 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

$$\begin{aligned} \mathrm{rank}(M) &\le k - p^{1/j} \\ &= k - n^{(1+\gamma-1/a)/j} \\ &\le k - n^{(a\gamma(1+\gamma)-\gamma)/2}\,, \end{aligned} \tag{9}$$

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 (9) for large enough n.

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

Corollary 11: Suppose we could prove the result of Theorem 8 with a rate of $\Omega(\varepsilon^a)$ for some $1 \le a < 2$ and a construction time that is polynomial in N and $1/\varepsilon$. Then there exists a polynomial time construction of $n \times n$ matrices over $\{0,1\}$ which have no monochromatic $n^b \times n^b$ sub-matrix for some $b < 1/2$.

As remarked earlier, a polynomial time construction of bipartite Ramsey graphs with p much smaller than $\sqrt{n}$ is a folklore open problem. 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 open question whether this can be exploited to improve the rate beyond ε^2. 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.

IV. 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 [18] 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).

Acknowledgments

The author would like to thank Piotr Indyk, Dan Spielman and Madhu Sudan for useful discussions. He is grateful to Noga Alon for pointing out the connection to Ramsey graphs outlined in Section III-C and for his kind permission to include the results of Proposition 10 and Corollary 11 in this paper. Finally, the author would like to thank Ralf Koetter and the anonymous referees for bringing to his attention the literature on Generalized Hamming weights.

References

[1] M. Alekhnovich. Linear Diophantine equations over polynomials and soft decoding of Reed-Solomon codes. *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 439-448, November 2002.
[2] N. Alon. Personal Communication, October 1999.
[3] 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 Transactions on Information Theory*, **38**:509–516, 1992.
[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):1258–1263, 1999.
[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**:2090–2092, 1994.
[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 Transactions on Information Theory*, **37**(1):5–12, 1991.
[12] G. L. Feng. Two Fast Algorithms in the Sudan Decoding Procedure. *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 545-554, 1999.
[13] G. D. Forney. *Concatenated Codes*. MIT Press, Cambridge, MA, 1996.
[14] 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.
[15] V. Guruswami. *List Decoding of Error-Correcting Codes*. Ph.D thesis, Massachusetts Institute of Technology, August 2001.
[16] V. Guruswami, J. Håstad, M. Sudan and D. Zuckerman. Combinatorial Bounds for List Decoding. *IEEE Transactions on Information Theory*, **48**(5):1021–1034, May 2002.
[17] 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.
[18] 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)*, pages 812-821, May 2002.
[19] V. Guruswami and M. Sudan. Improved Decoding of Reed-Solomon and Algebraic-geometric codes. *IEEE Transactions on Information Theory*, **45**:1757–1767, 1999.
[20] V. Guruswami and M. Sudan. List decoding algorithms for certain concatenated codes. *Proceedings of the 32nd ACM Symposium on the Theory of Computing (STOC)*, pages 181-190, May 2000.
[21] V. Guruswami and M. Sudan. Extensions to the Johnson Bound. Manuscript, 2001. Available from `http://theory.lcs.mit.edu/~madhu/npapers.html`.
[22] T. Helleseth, T. Kløve, V. I. Levenshtein, and Ø. Ytrehus. Bounds on minimum support weights. *IEEE Transactions on Information Theory*, **41**(2):432–440, 1995.
[23] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier/North-Holland, Amsterdam, 1981.
[24] 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**:157–166, 1997.
[25] M. A. Tsfasman and S. G. Vlădut. Geometric approach to higher weights. *IEEE Transactions on Information Theory*, **41**:1564–1588, 1995.
[26] 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.
[27] J. H. van Lint. *Introduction to Coding Theory*. Graduate Texts in Mathematics **86**, (Third Edition) Springer-Verlag, Berlin, 1999.
[28] V. Wei. Generalized Hamming weights for linear codes. *IEEE Transactions on Information Theory*, **37**(5):1412–1418, 1991.
[29] J. M. Wozencraft. List Decoding. *Quarterly Progress Report*, Research Laboratory of Electronics, MIT, Vol. 48 (1958), pp. 90-95.
[30] 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.