Efficiently Decodable Codes Meeting Gilbert-Varshamov Bound for Low Rates

Venkatesan Guruswami* Piotr Indyk†

Abstract

We demonstrate a probabilistic construction of binary linear codes meeting the GV bound (with overwhelming probability) for rates up to about 10^{-4} together with polynomial time algorithms to perform encoding and decoding up to half the distance. The only previous result of this type (for rates up to about 0.02) suffered from sub-exponential time decoding [3].

1 Introduction

Constructing of codes with best rate vs. distance trade-off together with efficient encoding and decoding procedures is the central problem of coding theory. However, even without requirement for efficient decoding algorithms, we do not know how to construct optimal codes explicitly. A random linear code, although not-known to achieve optimal rate/distance tradeoff, provides very good bounds with overwhelming probability. In fact, this is the best known rate vs. distance trade-off for binary codes called the Gilbert-Varshamov (GV) bound. At the same time, since there is no general way to certify that a random code is good or to actually decode a good code, the random code construction is only of theoretical interest.

While a random linear code has little structure, Thommesen [2] proved that one can meet the GV bound by picking a random code from a more structured ensemble of binary linear codes. Specifically, he proved that a concatenated code, with outer Reed-Solomon code of suitable parameters and binary linear inner codes picked independently at random, meets the GV bound with high probability. Still, unique decoding the resulting codes up to the optimal radius (i.e., up to half the GV bound) remained an open problem. The only previous result of this type was shown in [3], where a decoding algorithm with running time $\exp(\sqrt{n})$ (n is the code block-length) for rates up to 0.02, is given.

In this paper, we give a polynomial-time decoding algorithm for low rates. We note that the classical GMD approach to decoding concatenated codes can correct only up to half the product of the outer and inner distances, or half the so-called Zyablov bound, which is in general much smaller than half the GV bound. Instead, we use the list decoding algorithms known for Reed-Solomon codes to accomplish this task. To get the best bound on rate up to which we can decode up to half the GV bound, we use a recent algorithm [1] that uses soft information passed from the inner decodings to the Reed-Solomon list decoder.

2 Background on the Thommesen Construction

We refer to the paper by Thommesen [2] for details of the construction and proof and just state the main result from that paper in a form which will be useful to us. For $0 \le x \le 1$, define the function $\alpha(x) = 1 - H(1 - 2^{x-1})$ (where as usual $H(y)$ denotes the binary entropy function of y).

THEOREM 2.1. *[2] Let r_0 and R_0 be given such that $0 < r_0 \le 1$ and $0 < R_0 \le \alpha(r_0)/r_0$. For a large enough integer t, let $\mathrm{RS}_t(R_0)$ be the Reed-Solomon code over $\mathrm{GF}(2^t)$ of block length $N = 2^t$ and rate R_0. Consider an ensemble of concatenated codes with $\mathrm{RS}_t(R_0)$ as outer code with varying inner codes, that is codes where the codewords are of the form*

$$uG = [u_1G_1, u_2G_2, \ldots, u_NG_N], \quad u \in \mathrm{RS}_t(R_0)$$

where G_i, $1 \le i \le N$, are binary $t \times t/r_0$ matrices, each picked at random and independently of the others. Then, for every $\varepsilon > 0$, the probability that a random concatenated code from this ensemble (defined by a random choice of the G_i's) has relative distance at most $H^{-1}(1 - r_0R_0) - \varepsilon$ is exponentially small in the block length. This also implies that such a code has rate r_0R_0 with high probability. In other words, a random code from this ensemble meets the Gilbert-Varshamov bound with high probability.

*Department of Computer Science, University of Washington, Seattle.

†Laboratory for Computer Science, MIT, Cambridge, MA 01239.

3 The Decoding Algorithm and Analysis

Our decoding strategy is the one outlined in the following theorem, which is an adaptation of Theorem 3 of [1].

THEOREM 3.1. *[1] Consider a family of binary linear codes where each member of the family is a concatenated code with the outer code a Reed-Solomon code of relative distance Δ with block length equal to the size of the underlying field, say Q,*[1]*, and each inner code a (possibly different) binary linear code of dimension* $\log Q$ *such that a $(1-\gamma)$ fraction of the inner codes have relative distance at least δ. Then, codes from such a family can be encoded in polynomial time, as well as list decoding in polynomial time up to a fractional radius of*

$$\frac{1}{2}\cdot\left(1-\sqrt{1-2\delta}\right)-\sqrt{\delta(1-\Delta)}-\gamma\,. \tag{3.1}$$

The above theorem is proven by using a list decoding algorithm for Reed-Solomon codes that uses 'soft' information passed from the inner decodings.

4 Our construction

Our approach is to pick a concatenated code with outer Reed-Solomon code and varying inner codes picked at random. Such a code will lie on the Gilbert-Varshamov bound with high probability as per Theorem 2.1. We will also pick the outer and inner distances Δ and δ so that the error fraction that can be corrected as per Equation (3.1) will equal $1/4$ (which is the largest fraction of errors that can be unique decoded for binary codes). Our overall rate will be very close to $(1-\Delta)(1-H(\delta))$. Optimizing over the choice of δ,Δ to maximize the rate will give us our final bound. We now state the formal result that follows by combining the statements of Theorems 2.1 and 3.1.

THEOREM 4.1. *For every $\varepsilon > 0$ and $0 < \delta < 1/2$, let $r_0 = 1-H(\delta)-\varepsilon$ and let R_0 satisfy $0 < R_0 \le \alpha(r_0)/r_0$. For a large enough integer t, let $\mathrm{RS}_t(R_0)$ be the Reed-Solomon code over $\mathrm{GF}(2^t)$ of block length $N = 2^t$, rate R_0, and relative distance $(1-R_0)$. Consider an ensemble of concatenated codes with $\mathrm{RS}_t(R_0)$ as outer code with varying inner codes of dimension t and block length t/r_0 picked as in Theorem 2.1. Then:*

1. *A random concatenated code from this ensemble has rate r_0R_0 and relative distance at least $H^{-1}(1-r_0R_0)-\varepsilon$, and thus meets the GV bound, except with probability that is exponentially small in the block length.*
2. *There is a polynomial time decoding algorithm to decode every concatenated code from such an ensemble up to a fraction*

$$\frac{1}{2}\cdot\left(1-\sqrt{1-2\delta}\right)-\sqrt{\delta R_0}-\varepsilon \tag{4.2}$$

of errors.

Proof: Follows from Theorems 2.1 and 3.1 using the fact that a binary linear code defined by picking a random $t\times t/r_0$ matrix as its generator matrix where $r_0 = 1-H(\delta)-\varepsilon$ has relative distance at least δ with overwhelming probability. In particular, except with exponentially small probability (in N), more than a fraction $(1-\varepsilon)$ of the inner codes will have relative distance at least δ, and so the claim on decoding follows from Theorem 3.1. □

Pluggin in specific choice of the parameters, specifically δ, R_0, we conclude our main result:

THEOREM 4.2. (MAIN) *There is a probabilistic polynomial time procedure to construct codes whose rate vs. distance trade-off meets the Gilbert-Varshamov bound with high probability for all rates less than* 10^{-4}.

Proof: This follows by plugging into Theorem 4.1 the following choice of parameters (found by search using a simple program): $\delta = 0.421$, $r_0 = 0.01808$, and $R_0 \le 0.00623$. The choice of δ and R_0 can be seen to ensure that the decoding radius from Equation (4.2) is at least $1/4$, so that we are guaranteed to decode up to radius $1/4$ and in particular up to half the relative distance. It can also be checked that the choice of r_0, R_0 satisfy $r_0R_0 < \alpha(r_0)$, so that the condition for Thommesen's result is met and the codes will meet the GV bound w.h.p. Thus, we have efficiently decodable codes meeting the GV bound for rates up to $0.01808\times 0.00623 \approx 1.126\times 10^{-4}$. □

[1] This restriction on the block length is not necessary and the claim holds for any field size which is a polynomially growing function of the block length.

References

[1] Venkatesan Guruswami and Madhu Sudan. Decoding concatenated codes using soft information. *Proc. of the IEEE Conference of Computational Complexity*, Montreal, Canada, May 2002.

[2] Christian Thommesen. The existence of binary linear concatenated codes with Reed-Solomon outer codes which asymptotically meet the Gilbert-Varshamov bound. *IEEE Transactions on Information Theory*, 29(6):850–853, 1983.

[3] Victor V. Zyablov and Mark S. Pinsker. List cascading decoding, *Problems of Information Transmission*, 17(4):29–34, 1981 (in Russian), pages 236–240 (in English), 1982.