Numerical results on the asymptotic rate of binary codes

A. Barg and David B. Jaffe

ABSTRACT. We compute upper bounds on the maximal size of a binary linear code of length $n = 1000$, dimension k, and distance d. For each value of d, the bound is found by solving the Delsarte linear programming problem. Relying on the results of the calculations, we discuss the known bounds on the size of codes and some recent conjectures made about them. The most important conclusion is that Delsarte's linear programming method is unlikely to yield major improvements of the known general upper bounds on the size of codes.

1. Introduction: bounds on codes

A code C is a subset of the binary Hamming space H_2^n. The minimum distance between a pair of distinct points in C is called the distance of C, denoted $d(C)$. One of the main problems of coding theory is to find the maximal size $A(n,d)$ of a code with given distance d. This problem is solved exactly only for some small values of n and d. The general results known are in the form of upper and lower bounds.

The aim of this paper is to explore the limits of Delsarte's linear programming (or polynomial) method of deriving upper bounds on $A(n,d)$. After explaining the method and citing the best known general bounds, we present the results of calculations for $n = 1000$. Based on them, we speculate that the bounds currently known are likely to be close to the limits of Delsarte's method.

The best known lower bound (the Varshamov-Gilbert or VG bound) can be stated in the following form: *Let M be the maximal number such that*

$$(M-1)\sum_{i=0}^{d-1}\binom{n}{i} < 2^n.$$

Then there exists a code C of length n with $d(C) = d$ and $|C| = M$.

On the other hand, for any code,

$$|C| \le \begin{cases} 2^n / \sum_{i=0}^{(d-1)/2}\binom{n}{i} & d \text{ odd} \\ 2^{n-1} / \sum_{i=0}^{(d-2)/2}\binom{n-1}{i} & d \text{ even;} \end{cases} \tag{1}$$

this is the Hamming bound.

The second author was partially supported by NSF grant DMS-9801581.

The best known upper bounds are derived by Delsarte's linear programming method [**4**], [**5**]. We have

$$A(n,d) \le 1 + \max \sum_{i=d}^{n} A_i \tag{2}$$

$$A_i \ge 0 \quad (d \le i \le n); \qquad \sum_{i=1}^{n} A_i K_k(n,i) \ge -\binom{n}{k} \quad (0 \le k \le n).$$

Here $K_k(n,x)$ is a Krawtchouk polynomial,

$$K_k(n,x) = \sum_{i=0}^{n} (-1)^i \binom{x}{i}\binom{n-x}{k-i}.$$

Problem (2) in the dual form can be formulated as follows:

$$A(n,d) \le D(d) := \inf_f \{ f(0) : \sum_{i=0}^{n} f(i)\tbinom{n}{i} = 2^n;\ f(i) \le 0, i = d, d+1, \ldots, n\}, \tag{3}$$

where the infimum is taken over all polynomials $f \in \mathbb{R}[x]$ with nonnegative Fourier-Krawtchouk coefficients.

This method is often useful to derive bounds for specific values of n and d. General upper bounds on $A(n,d)$, i.e., those valid for any n, are derived by exposing a polynomial feasible with respect to the conditions in (3). The best known bounds are due to McEliece, Rodemich, Rumsey, and Welch (MRRW) [**11**] and Levenshtein [**9**]. Denote by $x_t(n)$ the first zero of $K_t(n,x)$. It is known that

$$x_t(n-1) \le x_t(n) \le x_{t-1}(n-1),$$

$t = 1, \ldots, n-1$; we also put $x_t(n) := n+1$; $x_{n+1}(n) := 0$. The first of the two MRRW bounds has the form

$$A(n,d) \le \binom{n}{t} \frac{(n+1)^2}{2a(t+1)}, \tag{4}$$

where $d \ge x_{t+1}(n)$ and $a \in [x_{t+1}(n), x_t(n)]$ is the root of $K_t(n,x) = -K_{t+1}(n,x)$.

Levenshtein's bound is of the following form (see also[**10**]):

$$A(n,d) \le L_n(d), \tag{5}$$

where

$$L_n(d) = \begin{cases} L_{k,n}(d) & \text{if } x_k(n-1)+1 < d \le x_{k-1}(n-2)+1, \\ 2L_{k,n-1}(d) & \text{if } x_k(n-2)+1 < d \le x_k(n-1)+1 \end{cases}$$

and

$$L_{k,n}(z) = \sum_{i=1}^{k} \binom{n}{i} - \binom{n}{k} \frac{K_{k-1}(n-1, z-1)}{K_k(n,z)}.$$

This bound for many finite values of (n,d) is better than (4).

Both (4) and (5) admit improvements based on the inequality

$$A(n,d) \le \min_w \frac{2^n}{\binom{n}{w}} A(n,d,w), \tag{6}$$

where $A(n,d,w)$ is the maximal size of a constant weight code with distance d and weight of all vectors w, and the minimum is taken over all integer w between 0 and $n/2$. Bounds

on $A(n,d,w)$ in [**11**] and [**9**] involve the extremal zero $y_t(n,w)$ of a dual Hahn polynomial of degree t (this family of polynomials is orthogonal on $(0,1,\dots,w)$ with weight $\binom{w}{i}\binom{n-w}{i}/\binom{n}{w}$). In particular, the result in [**11**] has the form

$$A(n,d,w) \le \binom{n}{t} \frac{(n^2-(2t-1)n-2t)^2(w-t)(n-w-t)}{y_{t+1}(n,w)(t+1)(n-t+1)(n-2t-1)(n-2t)(n-2t+1)}. \tag{7}$$

where t is such that $d \ge 2y_t(n,w)$. Inequalities (6)-(7) together form the second MRRW bound on $A(n,d)$. It is often better than (4), though notably more difficult to compute. Therefore, more specific comparison is easier for the asymptotic versions of these bounds. We do not formulate Levenshtein's bounds on $A(n,d,w)$ since they are not used below.

Let us also cite the asymptotical behavior of these bounds. For this purpose, let

$$R(\delta) = \limsup_{n\to\infty} \frac{1}{n} \log_2 A(n,d),$$

where the limit is computed over all sequences of codes with $\lim(d/n) \ge \delta$. Then by the VG bound

$$R(\delta) \gtrsim R_{vg} := 1 - H(\delta), \tag{8}$$

where H is the binary entropy function.

The asymptotic form of the MRRW bounds is as follows:

$$R(\delta) \lesssim R_{m1} := H\Big(\frac{1}{2} - \sqrt{\delta(1-\delta)}\Big). \tag{9}$$

$$R(\delta) \lesssim R_{m2} := \min_{0\le u\le 1-2\delta} 1 + g(u^2) - g(u^2+2\delta u+2\delta), \tag{10}$$

where $g(x) = H((1-\sqrt{1-x})/2)$. It is known [**11**] that $R_{m2} \le R_{m1}$ with equality for all $\delta \in [0.273, 1/2]$. Levenshtein's bounds for large n also converge to R_{m1} and R_{m2}, respectively.

Tightening the gap between the lower and upper bounds is the main problem of asymptotic coding theory. It is also far from solution. Until recently it was not even known whether Delsarte's method can lead to the proof of the (asymptotic) tightness of the VG bound. In 1998 A. Samorodnitsky [**12**] proved that

$$(1/n)\log_2 D(\delta n) \gtrsim (R_{vg} + R_{m1})/2, \tag{11}$$

resolving this question in the negative, and conjectured that asymptotically this inequality is tight. This conjecture, if true, would imply that $R(\delta) \lesssim (R_{vg}+R_{m1})/2$. The present study is to some extent motivated by this result and conjecture.

2. $\mathbf{n = 1000}$

2.1. Results. Here we present the results of the calculations. For $n = 1000$ and a given d we solved the linear programming problem (2). Two essential limitations of the calculations are as follows. Due to time constraints we were not able to deal with d odd or $d \le 43$. Further, we only looked for the maximal k such that the existence of a linear $[n, 2^k, d]$ code does not contradict the inequalities in Delsarte's problem (2). This assumption simplified the calculations considerably; see also the next section. While we cannot make any claims on the maximum of (2) with linearity restriction lifted, the results for

linear codes are so close to the upper bounds cited that this question seems irrelevant. The results are most easily visualized by plotting the curves in coordinates (δ, R), see Fig. 1, 2.

We also give a short table of the bounds. Some of the formulas computed in the table were stated above as bounds on the size of the code rather than on $R(\delta)$. In this case we compute the binary log of the answer and normalize it by n. The line typeset in boldface gives the results of our calculations of the maximum in (2).

d/n	0.05	0.01	0.15	0.2	0.25	0.3	0.35	0.4	0.45
(4)	0.881	0.750	0.624	0.502	0.391	0.287	0.196	0.115	0.057
(5)	0.878	0.748	0.621	0.501	0.387	0.283	0.191	0.115	0.0505
(1)	0.839	0.722	0.624	0.539	0.464	0.397	0.338	0.285	0.237
(2)	**0.839**	**0.712**	**0.597**	**0.488**	**0.380**	**0.280**	**0.18**8	**0.10**9	**0.047**
(9)	0.858	0.722	0.592	0.469	0.355	0.250	0.158	0.082	0.025
(10)	0.825	0.693	0.573	0.461	0.354	0.250	0.158	0.082	0.025
(11)	0.786	0.626	0.491	0.374	0.272	0.184	0.112	0.055	0.016
(8)	0.714	0.531	0.390	0.278	0.189	0.199	0.066	0.029	0.007

Here in the upper half we have collected the bounds for $n = 1000$: (top to bottom) the nonasymptotic MRRWI, Levenshtein's, Hamming, and a solution of the LP problem. In the lower half we show the asymptotic 1st and 2nd MRRW bounds, Samorodnistky's conjectured answer, and the GV bound. For small d/n the agreement of the computed solution for the LP problem and the Hamming bound (1) is very good:

d/n	0.044	0.046	0.048	0.05	0.052	.054	0.56
Hamming	0.8554	0.8500	0.8446	0.8393	0.8340	0.8287	0.8236
LP $(n = 1000)$	0.855	0.849	0.844	0.839	0.833	0.828	0.823

2.2. Comments. The results for $n = 1000$ follow the known bounds rather closely, both for large δ and in the range where (7) provides improvements over (4). Therefore, though the bounds most likely are not the best possible for finite (n, d) it is likely that the asymptotic MRRW bounds give the exact answer in the asymptotic Delsarte problem. The results do not, therefore, support Samorodnitsky's conjecture on the exponential behavior of $D(\delta n)$.

This leaves us in a rather uncomfortable situation with respect to finding $R(\delta)$. Namely, Delsarte's method is by far the most powerful method known in deriving upper bounds on this function; on the other hand it seems unlikely that the upper bounds (9)-(10) can be significantly improved within the frame of this method. So at present there seem to be no ideas around that would lead to tightening the gap between (8) and (10).

Another important parameter of a code is its distance distribution. Let us cite the values of an optimal assignment of variables $A_d,\ldots,A_n$ in the LP problem for $d = 350$. Again the easiest way to present the results is to plot them. We take the set of binomial probabilities $\mathcal{A}_w = \binom{n}{w}2^{k-n}, d \le w \le n$, as a reference distance distribution. The reason for this is that $(\mathcal{A}_w)$ is the average distance distribution of a code chosen in H_2^n with uniform probability. Sequences of codes whose distance distribution is asymptotically binomial are known to exist and have a number of interesting properties. Therefore, we plot the numbers $(1/n)\log_2 A_w$ and $(1/n)\log_2 \mathcal{A}_w$ with w/n on the x-axis. This plot is shown in Fig. 3. The results are in good accord with a theorem in [**1**] that asserts that codes meeting the MRRW bounds, if they exist, must have distance spectrum that converges to the binomial distribution.

3. Details on computations

In this section we explain how the data for $n = 1000$ was derived. In short, the program `Split` was used to mechanize the computations. For example, in `Split`, the following commands

```
type [1000,502,194_2];
via lp [current] = ;
type [1000,501,194_2];
!via lp [current] = ;
```

instruct the program to *attempt* to show that:

- There is no even binary linear code with parameters $n = 1000$, $k = 502$, $d = 194$, and hence no binary linear code (even or not) with these parameters. This is accomplished by linear programming.
- By linear programming one *cannot* show that there is no even binary linear code with parameters $n = 1000$, $k = 501$, $d = 194$.

The program performs these linear programming calculations using the simplex method, with floating point arithmetic (at a sufficiently high level of precision), and then verifies the conclusions using exact arithmetic. Thus, barring a programming error, the calculations are guaranteed to be correct.[1]

`Split` uses its own implementation of the simplex method, because commercial linear programming solvers do not support the high level of precision needed for the calculations, which is roughly ten to twenty times double precision. While the simplex method is still an excellent method for solving linear programs, the implementation in `Split` is primitive by comparison with the implementations in the commercial solvers (such as `CPLEX`), apart from their inability to calculate at high precision. Consequently, the calculations were far slower than they would be otherwise. As d decreased, the calculations became more and more time-consuming. For example, a single calculation for $d = 400$ took about an hour, whereas a single calculation for $d = 50$ took about 200 hours (using one 500 MhZ Alpha 21264 processor). This explains why the calculations terminate at $d = 44$.

[1] Owing to time and memory restrictions, not all the calculations of the second type were verified. But in all cases where they were verified, no inconsistency was observed.

In principle exactly the same method could be used to bound unrestricted binary codes, in which case we would fix $n = 1000$ and some d, and try to bound the number M of codewords. However, owing to the particular implementation of the simplex method in `Split`, it is easier to answer questions of the form "Is $M \geq C$?" or "Is $M \leq C$?" for fixed C, than to find M directly. Since in the linear case one need only use C values which are powers of 2, the calculations are easier.

More information about `Split` may be found in [**6**], [**7**], [**8**], [**2**], and [**3**].

Color plots of the data for $n = 1000$ may be found on the second author's webpage, `http://www.math.unl.edu/~djaffe`.

References

[1] A. Ashikhmin, A. Barg, and S. Litsyn, *Bounds on the distance distribution of codes and designs*, (2000), preprint.

[2] I. Bouyoukliev, D. B. Jaffe, and V. Vavrek, *The smallest length of eight-dimensional binary linear codes with prescribed minimum distance*, IEEE Transactions on Information Theory, to appear.

[3] I. Bouyoukliev and D. B. Jaffe, *Optimal binary linear codes of dimension at most seven*, Discrete Mathematics, to appear.

[4] P. Delsarte, *Bounds for unrestricted codes, by linear programming*, Philips Res. Rep. **27** (1972), 272–289.

[5] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Research Repts Suppl. **10** (1973), 1–97.

[6] D. B. Jaffe, *A brief tour of split linear programming* (Proc. AAECC 12, ed. T. Mora, H. Mattson), Lecture Notes in Computer Science **1255** (1997), 164–173.

[7] ———, *Binary linear codes: new results on nonexistence*, preprint (ongoing work), Version 0.6beta (3/21/2000), accessible over the World Wide Web via `http://www.math.unl.edu/~djaffe/codes/code.ps.gz` or `code.dvi.gz`; see `.../~djaffe/binary/codeform.html` for an online database, which is more frequently updated.

[8] ———, *Optimal binary linear codes of length* ≤ 30, Discrete Mathematics, to appear.

[9] V. I. Levenshtein, *On choosing polynomials to obtain bounds in packing problems*, Proc. 7th All-Union Conf. Coding Theory and Information Transmission, Part 2 (Moscow, Vilnius), 1978, (In Russian), pp. 103–108.

[10] ———, *Universal bounds for codes and designs*, Handbook of Coding Theory (V. Pless and W. C. Huffman, eds.), vol. 1, Elsevier Science, Amsterdam, 1998, pp. 499–648.

[11] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory **23** (1977), no. 2, 157–166.

[12] A Samorodnitsky, *On the optimum of Delsarte's linear program*, (1998).

BELL LABS, LUCENT TECHNOLOGIES, 600 MOUNTAIN AVE., RM. 2C-375, MURRAY HILL, NJ 07974

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF NEBRASKA, LINCOLN, NE 68588-0323

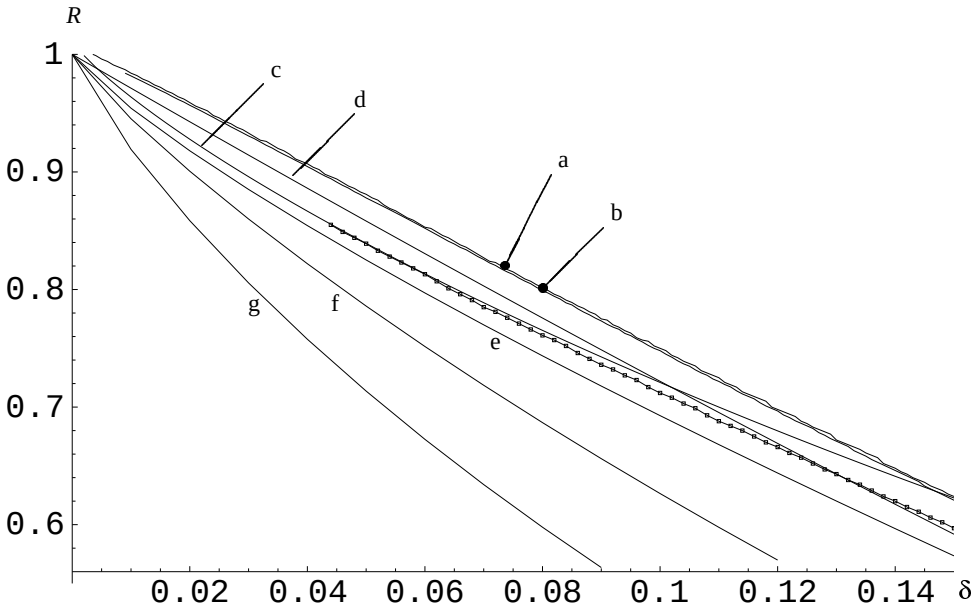

FIGURE 1. Bounds for low distances. The curve with marked dots represents the data for $n = 1000$ computed by (2). Nonasymptotic bounds ($n = 1000$): (a) MRRW bound (4), (b) Levenshtein's bound (5), (c) Hamming bound (1); Asymptotic upper bounds: (d) (9), (e) (10), (f) conjectured logarithmic asymptotics of $D(\delta n)$ (11), (g) the Varshamov-Gilbert bound.

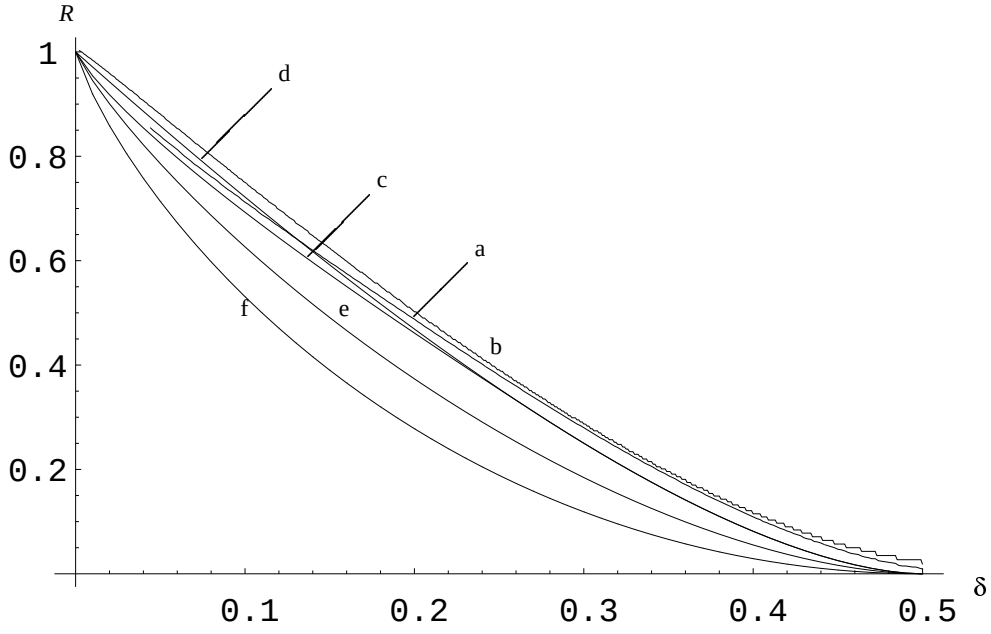

FIGURE 2. Bounds on codes: (a) data for $n = 1000$ computed by (2), (b) MRRW bound (4) for $n = 1000$, (c) (10), (d) (9), (e) (11), (f) the Varshamov-Gilbert bound.

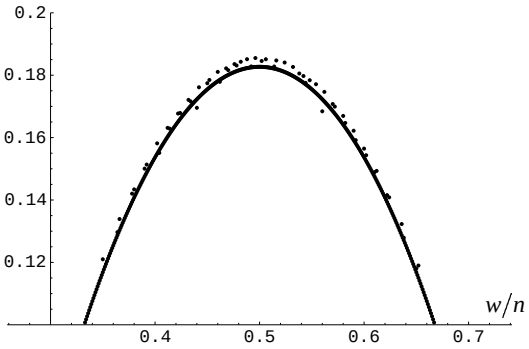

FIGURE 3. Optimal assignment of variables for $d = 320, k = 188$ (dots) and binomial distribution (solid curve).