Closed-Form Approximation of Maximum Free Distance for Binary Block Codes

J. Akhtman, R. Maunder, N. Bonello and L. Hanzo
School of ECS., Univ. of Southampton, SO17 1BJ, UK.
Email: lh@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

***Abstract*—We devise an analytically simple as well as invertible approximate expression, which describes the relation between the maximum free distance of a binary code and the corresponding maximum attainable code-rate. For example, for a half-rate, length-128 binary code the known bounds limit the maximum attainable free distance to $16 < d(n = 128, r = 0.5) < 32$, while our solution yields $d(n = 128, r = 0.5) \approx 22$. The results provided may be utilized for the design and characterization of efficient coding schemes.**

I. Introduction

One of the fundamental open problems in coding theory is constituted by the issue of determining the highest cardinality $|\mathcal{C}| = 2^k$ attainable by a binary code $\mathcal{C}$ of length n, having a rate of $r = k/n$ and a free distance of d [1], where the free distance d is defined as the minimum Hamming distance between any two codewords in the codebook $\mathcal{C}$. In addition to its theoretical significance, the problem considered appears in numerous important applications, including the design of efficient coding schemes and their characterization in terms of the achievable probability of error. Although the complete solution of the rate-versus-free-distance problem does not exist at the time of writing, several theoretical lower and upper bounds on the desired relation may be found in the literature [1]–[5]. The best known asymptotic ($n \to \infty$) as well as finite-n-related lower and upper bounds are summarized in Table I. More specifically, the tightest known asymptotic ($n \to \infty$) lower bound was derived by Gilbert [3], while the corresponding upper bounds were devised by Hamming [2] and McEliece *et. al.* (MRRW) [5]. The important asymptotic lower and upper bounds are depicted in Figure 1. Furthermore, the best known finite-n bounds are constituted by the Gilbert lower bound, as well as the Hamming and Plotkin upper bounds [4]. The finite-n lower and upper bounds for the specific case of having $n = 7$ are depicted in Figure 2.

Unfortunately, however, most of the theoretical bounds are notoriously difficult to use in practice. On the one hand, as may be inferred from Figures 1 and 2, the asymptotic bounds bear little relevance to a wide range of finite-n scenarios, routinely encountered in practical applications. On the other hand, the theoretical bounds corresponding to the finite-n cases involve excessively complex numerical computations. Against this background, *the novel contribution of this paper is constituted by the formulation of an analytically simple as well as invertible expression $r(n, \delta)$, where we define a normalized free distance $\delta = d/n$, which would comply with all known theoretical bounds in both finite-n as well as in asymptotic ($n \to \infty$) contexts, while providing a practical tool for the design and characterization of efficient binary codes.*

The binary entropy function $H(q)$ in Table I is defined as $H(q) = q\log_2(q) + (1-q)\log_2(1-q)$.

II. Rate versus free distance trade-off

Firstly, let us consider three special cases, where the exact value of the maximum free distance d is known.

a) For a unity-rate binary code of length $n = 1, 2, \ldots$, we have $d = 1$.
b) In the case of a block length of $n = 2^k - 1$, $k = 1, 2, \ldots$, we may consider an optimum rate-$\left[r{=}k/(2^k-1)\right]$ code, having a constant Hamming distance of $d = 2^{k-1}$ between any pair of codewords.
c) For any block length $n = 1, 2, \ldots$, we may consider an optimum rate-$(r{=}1/n)$ n-repetition code conveying a single bit of information and exhibiting $d = n$.

Secondly, we would like to point out the following list of important empirical observations.

i) As confirmed by Figure 1, a simple quadratic function
$$r(\delta) = (2\delta - 1)^2 \tag{1}$$
satisfies all known asymptotic bounds, namely the upper MRRW [5] and Hamming [2] bounds, as well as the lower Gilbert-Varshamov [3] bounds summarized in Table I.
ii) As exemplified by the specific case of $n = 7$, detailed in Table II and portrayed in Figure 2, the actual achievable values $r(\delta)$ constitute a discrete function, which cannot have an exact monotonic analytical description.
iii) As may be inferred from comparing Figures 1 and 2, the asymptotic bounds of Figure 1 bear little relevance to the important practical codes designed for example for interactive, real-time speech and video systems having $1 \leq n \ll 1000$.
iv) As further suggested by the specific example of having $n = 7$, both the finite-n Gilbert and Hamming bounds are relatively loose, while the Plotkin bound is tight for $\delta > \lceil n/2 \rceil / n$.
v) The Plotkin upper bound coincides with the actual achievable maximum rate r in the special cases of (b) and (c) considered above, which further substantiates the assumption that the Plotkin bound constitutes the

TABLE I
KNOWN BOUNDS ON A CODE RATE.

	finite n	asymptotic $n \to \infty$	notes
Varshamov-Gilbert [3]	$r \geq 1 - \frac{1}{n}\log_2 \sum_{i=0}^{d-1} \binom{n}{i}$	$r \geq 1 - H(\delta)$	tightest known lower bound
Hamming [1]	$r \leq 1 - \frac{1}{n}\log_2 \sum_{i=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{i}$	$r \leq 1 - H(\delta/2)$	tight upper bound for very high rate codes
MRRW [5]		$r \leq H(1/2 - \sqrt{\delta(1-\delta)})$	tightest known asymptotic upper bound for medium and low-rate codes
Plotkin [4]	$r \leq \frac{1}{n}\left[1 - \log_2(2 - \frac{1}{\delta})\right]$		very tight upper bound for $\delta > 1/2$

tightest possible analytical bound in the $\delta > \lceil n/2 \rceil / n$ range.

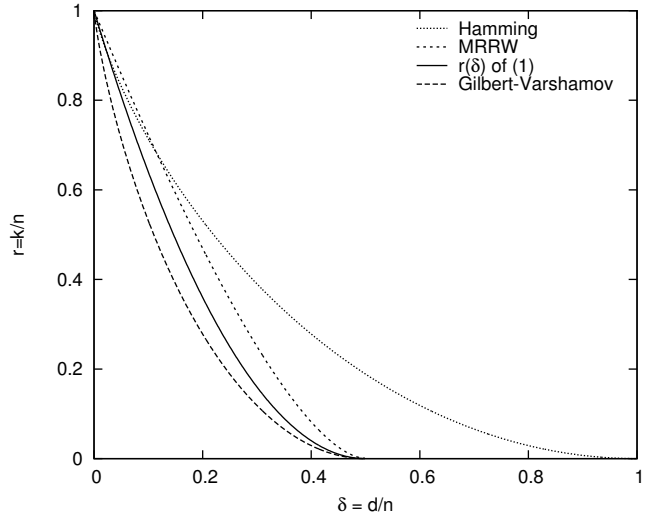

Fig. 1. Rate versus normalized free distance for known asymptotic bounds.

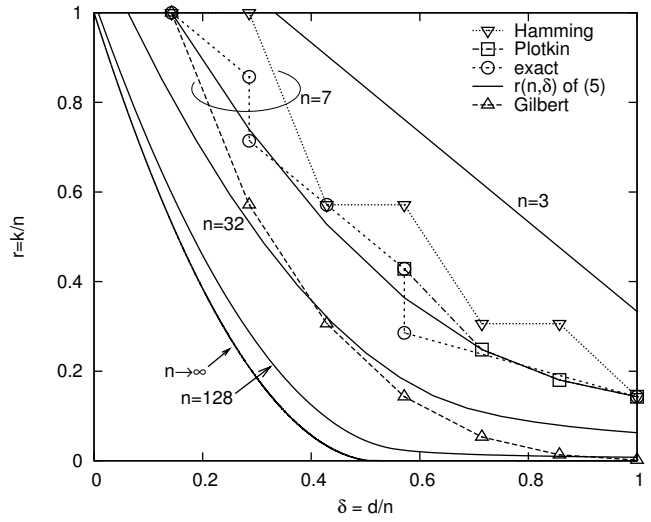

Fig. 2. Rate versus normalized free distance for finite length codes.

Taking into consideration observations (i)-(v), we hypothesize a solution exhibiting the following properties:

- Asymptotic quadratic approximation of (1)

$$\lim_{n\to\infty} r(n,\delta) = (2\delta - 1)^2. \tag{2}$$

- Unity-rate special case (a)

$$r(n, 1/n) = 1. \tag{3}$$

- Plotkin bound [4] and special cases (b) and (c)

$$r\left(n, \delta > \frac{\lceil n/2 \rceil}{n}\right) \approx \frac{1}{n}[1 - \log_2(2 - 1/\delta)]. \tag{4}$$

Specifically, we propose a solution in the form of a smooth two-segment function $r(n,\delta)$ expressed as

$$r(n,\delta) = \begin{cases} a\delta^2 + b\delta + c & \text{if } \delta < \lceil n/2 + \xi \rceil / n \\ \frac{1}{n}[1 - \log_2(2 - 1/\delta)] & \text{otherwise,} \end{cases} \tag{5}$$

where the free parameters a, b, c and ξ are chosen to ensure that the quadratic constituent in Equation (5) complies with the constraints (2) and (3), while the constraint (4) is automatically obeyed by the corresponding logarithmic constituent of (5).

Furthermore, the requirement of smoothness in the expression of (5) imposes the following additional constraints on the quadratic constituent in (5):

- Continuity at the transition point of $\delta_2 = \lceil n/2 + \xi \rceil / n$

$$r_2 = a\delta_2^2 + b\delta_2 + c = [1 - \log_2(2 - 1/\delta_2)]/n. \tag{6}$$

- Continuity of the first derivative at the transition point δ_2, which may be attained by imposing continuity of the discrete function of (5) in the next consecutive point $\delta_3 = (\lceil n/2 + \xi \rceil + 1)/n$, yielding

$$r_3 = a\delta_3^2 + b\delta_3 + c = [1 - \log_2(2 - 1/\delta_3)]/n. \tag{7}$$

By combining the constraints of (6) and (7) with (3), we arrive at a system of three equations, which uniquely determines the values of the parameters a, b and c. Specifically, we have

$$\begin{cases} r_1 = a\delta_1^2 + b\delta_1 + c \\ r_2 = a\delta_2^2 + b\delta_2 + c \\ r_3 = a\delta_3^2 + b\delta_3 + c, \end{cases} \tag{8}$$

where in addition to the parameters defined in (6) and (7), we have $r_1 = 1$ and $\delta_1 = 1/n$. The general solution of the system

TABLE II
BEST KNOWN BOUNDS ON THE FREE DISTANCE OF BINARY CODES OF LENGTH $n = 7$.

d	k	Gilbert	Hamming	Plotkin	type
1	7	7	7	-	uncoded
2	6	4	7	-	single parity bit
3	4	2	4	-	Hamming
4	3	1	4	3	SP
5	-	-	2	1	
6	-	-	2	1	
7	1	-	1	1	repetition

of equations in (8) is given by

$$a = \frac{r_3\left(\delta_2-\delta_1\right)+r_2\left(\delta_1-\delta_3\right)+r_1\left(\delta_3-\delta_2\right)}{\left(r_1-r_2\right)\left(r_1-r_3\right)\left(r_2-r_3\right)},$$
$$b = \frac{\left(\delta_2-\delta_3\right)r_1^2+r_3^2\left(\delta_1-\delta_2\right)+r_2^2\left(\delta_3-\delta_1\right)}{\left(r_1-r_2\right)\left(r_1-r_3\right)\left(r_2-r_3\right)},$$
$$c = \frac{\left(r_3\delta_1-x_1\delta_3\right)r_2^2+\left(r_1^2\delta_3-r_3^2\delta_1\right)r_2+r_1r_3\left(r_3-r_1\right)\delta_2}{\left(r_1-r_2\right)\left(r_1-r_3\right)\left(r_2-r_3\right)}. \tag{9}$$

Observe that despite it seemingly complex appearance, Equation (9) contains simple closed-form expressions, which may be readily calculated for any given value of n. Furthermore, it may be readily demonstrated that constraint (2) is satisfied if

$$\lim_{n\to\infty}\xi = \infty \tag{10}$$

and

$$\lim_{n\to\infty}\frac{n/2+\xi}{n} = \frac{1}{2} \;\Rightarrow\; \lim_{n\to\infty}\frac{\xi}{n} = 0. \tag{11}$$

Our analysis has shown that any sensible choice of the parameter ξ, where $\xi(n)$ is a monotonically increasing function satisfying the conditions (10) and (11) as well as $0 \le \xi(1) \le 1$ yields similar results. Specifically, in this study we assume having $\xi = \log_2(n)$.

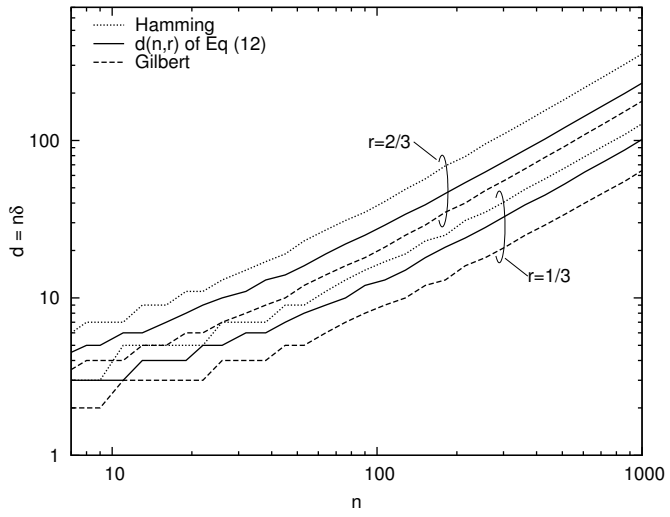

Fig. 3. Maximum free distance versus code-length for short binary codes.

The resultant expression $r(n,\delta)$ of Equation (5) is compared to the existing theoretical bounds in Figures 1 and 2 for the asymptotic case ($n \to \infty$) and the finite-n cases of $n = 3, 7, 32, 128$, respectively.

Expression (5) may be deemed analytically simple, since it has a closed form and is composed of elementary functions. Moreover, (5) is readily invertible, yielding

$$\delta(n,r) = \begin{cases} \dfrac{-b-\sqrt{b^2-4a(c-r)}}{2a} & \text{if } r > \frac{1}{n}\log_2(n+1) \\ \dfrac{2^{rn-1}}{2^{rn}-1} & \text{otherwise,} \end{cases} \tag{12}$$

where the coefficients a, b and c may be readily calculated using (6)–(9). In the asymptotic case of having $n \to \infty$, which in practice may be safely employed for all scenarios having $n \gg 100$, we may simply use the inverse of (1), yielding $\delta(r) = (1+\sqrt{r})/2$.

Figure 3 portrays the comparison between the expression of Equation (12) and the best known theoretical upper and lower bounds of Table I for the specific cases of rate-$(1/3)$ as well as rate-$(2/3)$ binary codes. Observe, that the devised approximate expression of Equation (12) coincides with both the lower and the upper bounds for certain values of the code-length n, which is indicative of the tightness of the derived approximation to the desired value of the maximum free distance.

III. CONCLUSION

We formulated an analytically simple as well as invertible expression $r(n,\delta)$, which approximates the optimum trade-off between the maximum rate and the corresponding maximum free distance attainable by binary codes of length n. The resultant closed-form analytical expression complies with all known theoretical bounds in both finite-n as well as in asymptotic ($n \to \infty$) contexts.

For instance, for the rate-$1/2$ binary code of length 128, the maximum attainable free distance d is bounded by the best known upper and lower theoretical bounds, where we have $16 < d(n = 128, r = 0.5) < 32$, which leaves a substantial ambiguity concerning the realistically attainable value of d. In this paper, we have demonstrated that the maximum attainable free distance may be more accurately approximated, yielding $d(n = 128, r = 0.5) = 22$ in this specific example. Likewise, for the case of a rate-$1/2$ binary code of length $n = 1024$, we have $117 < d(n = 128, r = 0.5) = 153 < 231$. Ultimately, the proposed method may be utilized for the design and characterization of efficient binary codes.

REFERENCES

[1] T. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley-Interscience, 2006.
[2] R. Hamming, "Error detecting and error correcting codes," *Bell System Technical Journal*, vol. 26, no. 2, pp. 147–160, 1950.
[3] E. Gilbert, "A comparison of signaling alphabets," *Bell Systems Tech. J.*, vol. 31, pp. 504–522, 1952.
[4] M. Plotkin, "Binary codes with specified minimum distance," *IEEE Transactions on Information Theory*, vol. 6, no. 4, pp. 445–450, September 1960.
[5] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 157–166, March 1977.