Correspondence

Random Codes: Minimum Distances and Error Exponents

Alexander Barg, *Senior Member, IEEE,* and
G. David Forney, Jr., *Fellow, IEEE*

***Abstract*—Minimum distances, distance distributions, and error exponents on a binary-symmetric channel (BSC) are given for typical codes from Shannon's random code ensemble and for typical codes from a random linear code ensemble. A typical random code of length N and rate R is shown to have minimum distance $N\delta_{\rm GV}(2R)$, where $\delta_{\rm GV}(R)$ is the Gilbert–Varshamov (GV) relative distance at rate R, whereas a typical linear code (TLC) has minimum distance $N\delta_{\rm GV}(R)$. Consequently, a TLC has a better error exponent on a BSC at low rates, namely, the expurgated error exponent.**

***Index Terms*—Distance distributions, exponential error bounds, minimum distance, random codes, random linear codes, typical linear codes, typical random codes.**

I. Introduction

The performance of random codes is one of the earliest topics in information theory, dating back to Shannon's random code ensemble (RCE). Our interest in this topic has been reawakened recently by the development of "random-like" capacity-approaching codes.

In this correspondence, we consider binary codes from Shannon's random code ensemble and also from a random linear code ensemble (LCE), used over a binary-symmetric channel (BSC). We derive the minimum distance, distance distribution, and error exponent of a typical random code (TRC) from the RCE, and of a typical linear code (TLC) from the LCE, and also averages over these ensembles. Most of these results were previously known and have appeared in the literature in some form or another.

Many of our results are stated in terms of the relative Gilbert–Varshamov (GV) distance $\delta_{\rm GV}(R)$, which for $0 \le R \le 1$ is defined as the root $\delta \le \frac{1}{2}$ of the equation $\mathcal{H}(\delta) = 1 - R$, where $\mathcal{H}(\delta)$ is the binary entropy function. Thus, $\delta_{\rm GV}(0) = \frac{1}{2}$, and $\delta_{\rm GV}(R)$ decreases monotonically to 0 as $R \to 1$. When $R > 1$, we define $\delta_{\rm GV}(R) = 0$. It is well known that for $0 \le R \le 1$, as $N \to \infty$ there exists a binary code of length N, rate R, and minimum distance $N\delta_{\rm GV}(R)$. Moreover, $\delta_{\rm GV}(R)$ is the best such asymptotic lower bound known.

It is known that, with probability approaching 1 as $N \to \infty$, the minimum distance of a binary linear code of length N and rate R drawn from a standard random LCE is at least $N\delta_{\rm GV}(R)$. In other words, the minimum distance of a TLC meets the GV bound. The earliest result of this type that we know of appears in [5, Sec. 2.1].

We show that, with probability approaching 1 as $N \to \infty$, the minimum distance of a binary code of length N and rate R from Shannon's RCE is approximately $N\delta_{\rm GV}(2R)$. Thus, a TRC has poorer minimum distance than a TLC for $0 < R < 1$. Moreover, the minimum distance of a TRC is positive only when $R < \frac{1}{2}$.

Manuscript received October 4, 2001; revised March 16, 2002. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

A. Barg is with Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA and with the Institute of Information Transmission Problems (IPPI), Moscow, Russia (e-mail: abarg@research.bell-labs.com).

G. D. Forney, Jr. is with the Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: forneyd@attbi.com).

Communicated by R. Koetter, Associate Editor for Coding Theory.
Publisher Item Identifier 10.1109/TIT.2002.800480.

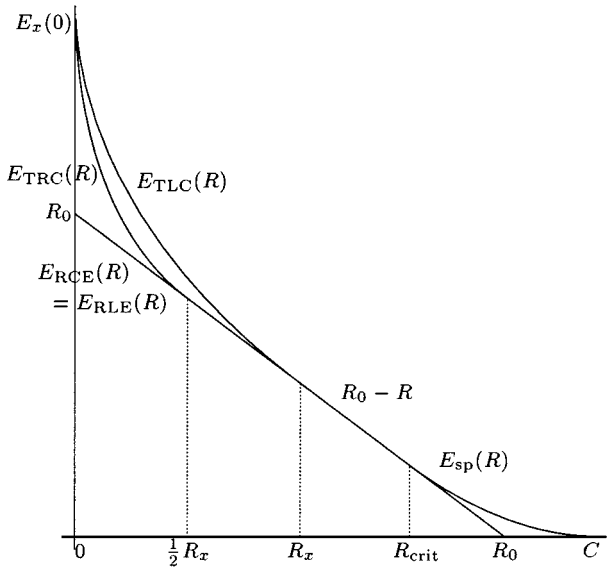

Fig. 1. Error exponents $E_{\rm RCE}(R)$, $E_{\rm TRC}(R)$, and $E_{\rm TLC}(R)$ for a BSC with $p = 0.007$.

It is well known that on a BSC with crossover probability p, the channel capacity is $C = 1 - \mathcal{H}(p)$. The random coding exponent $E_r(R)$ is positive for $0 \le R < C$ and given by [5], [6]

$$E_r(R) = \begin{cases} R_0 - R, & 0 \le R \le R_{\rm crit} \\ E_{\rm sp}(R), & R_{\rm crit} \le R < C \end{cases}$$

where the pairwise exponent R_0 (sometimes called the "cutoff rate"), the critical rate $R_{\rm crit}$, and the sphere-packing exponent $E_{\rm sp}(R)$ will be defined later. Moreover, Gallager [7] has shown that the random coding exponent is the true error exponent for the RCE on the BSC (and, more generally, on any discrete memoryless channel).

The random coding exponent may be improved at low rates by a process called "expurgation," which yields an "expurgated exponent" $E_x(R)$ that exceeds $E_r(R)$ for $0 \le R < R_x$, where $E_x(R)$ and R_x will also be defined later [6]. Many have conjectured (e.g., [9]) that the combination of the expurgated and the random coding exponents is the best possible error exponent—i.e., that the expurgated/random coding exponent is the "reliability function" of the BSC.

We argue that the relative minimum distance $\delta_{\rm GV}(R)$ and complete distance distribution $\mathcal{N}_{\rm TLC}(d)$, $d = 1, 2, \ldots N$ of a TLC imply that on a BSC its true error exponent $E_{\rm TLC}(R)$ is equal to the expurgated/random coding exponent; i.e., TLCs achieve the best lower bound on error exponent at all rates, without expurgation.

Similarly, we find the complete distance distribution $\mathcal{N}_{\rm TRC}(d)$ of a TRC, and from this distribution derive the true error exponent $E_{\rm TRC}(R)$ of a TRC on a BSC. (All of the error exponents that we give are the true exponents, not merely bounds.) $E_{\rm TRC}(R)$ lies between the random coding exponent $E_r(R)$ and the expurgated exponent $E_x(R)$; it is equal to $E_r(R)$ for $R \ge R_x/2$, and is equal to $E_x(0)$ at $R = 0$.

Fig. 1 shows $E_r(R)$, $E_x(R)$ and $E_{\rm TRC}(R)$ for a BSC with crossover probability $p = 0.007$, which is representative of the general case.

Finally, we derive the typical number of channel errors and typical distance from the transmitted codeword to the decoded codeword in typical decoding error events. We observe that errors are typically *not* made to minimum-distance codewords when $R > R_x$, even though pairwise analysis and the union bound give the true error exponent when $R \leq R_{\rm crit}$.

II. Random Code Ensembles and Typical Codes

In this section, we introduce the appropriate ensembles of random codes for a BSC. We then discuss the properties of the typical random code (TRC) from the random code ensemble (RCE), and the typical linear code (TLC) from the linear code ensemble (LCE). In particular, we note that whereas the distance distributions of the RCE or LCE and the TRC or TLC are exponentially identical for distances above the GV distance, they differ radically for small distances.

A. Preliminaries

We will be concerned with code parameters (such as the error probability and distance distribution) that behave as exponential functions of the code length N as $N \to \infty$. For any two functions $a(N)$, $b(N)$ we write $a(N) \doteq b(N)$ if

$$\lim_{N\to\infty} N^{-1}\log(a(N)/b(N)) = 0.$$

We say that the exponent of $a(N)$ is $|E_a|$ if $a(N) \doteq 2^{NE_a}$.

As usual, for any two nonnegative functions the expression $f(N) = \Omega(g(N))$ means that there exist positive constants N_0 and c such that $f(N) \geq cg(N)$ for every $N \geq N_0$.

From the viewpoint of large-deviation theory, it is natural to express exponents in terms of the binary Kullback–Leibler (KL) divergence $D(p\|q)$, defined as

$$D(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}.$$

All logarithms will be to the base 2. The KL divergence $D(p\|q)$ is a strictly convex function of p and q that has a minimum of 0 when and only when $p = q$. Note that

$$D\left(p\left\|\frac{1}{2}\right.\right) = 1 - \mathcal{H}(p)$$

where $\mathcal{H}(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function.

B. Random Binary Codes

A binary code $\mathcal{C}$ of length N and rate R bits per symbol is a (multi)set of $M = 2^{NR}$ binary N-tuples $\boldsymbol{x}_i$, $0 \leq i \leq M-1$.

Because of the symmetry of a BSC, the appropriate RCE is the equiprobable ensemble in which each of the M bits in each of the N codewords is chosen independently at random with equal probability of being a 0 or a 1. Equivalently, each of the 2^{NM} possible binary codes of length N and rate R in the RCE is assigned probability 2^{-NM}.

In this case, the probability that a given random codeword $\boldsymbol{x}_i$ of length N will be at Hamming distance $d = N\delta$ from an arbitrary binary N-tuple $\boldsymbol{b}$ is independent of $\boldsymbol{b}$ and equals

$$\Pr\{d_H(\boldsymbol{x}_i, \boldsymbol{b}) = d\} = \binom{N}{d}\left(\frac{1}{2}\right)^d\left(\frac{1}{2}\right)^{N-d} \doteq 2^{-N(1-\mathcal{H}(\delta))} = 2^{-ND(\delta\|\frac{1}{2})}.$$

Under this RCE, two distances $d_H(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $d_H(\boldsymbol{x}_{i'}, \boldsymbol{x}_{j'})$ are independent random variables unless $\{i, j\} = \{i', j'\}$ or $\{i, j\} = \{j', i'\}$.

Consider the number of unordered pairs of codewords $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ with $i \neq j$ in $\mathcal{C}$ at distance d apart

$$S_{\mathcal{C}}(d) = \sum_{i=0}^{M-1}\sum_{j=0}^{i-1}\Phi\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\}$$

where $\Phi\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\}$ is the indicator of the event in the brackets; i.e., $\Phi\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\}$ is equal to 1 if $d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d$ and to 0 otherwise. Then $S_{\mathcal{C}}(d)$ is a sum of $\binom{M}{2}$ pairwise-independent, identically distributed random variables $\Phi\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\}$, each with mean

$$\mathsf{E}\Phi = \Pr\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\}$$

and variance

$$\mathrm{Var}[\Phi] = \mathsf{E}\Phi^2 - (\mathsf{E}\Phi)^2 = \mathsf{E}\Phi - (\mathsf{E}\Phi)^2 < \mathsf{E}\Phi$$

where we have observed that $\Phi^2 = \Phi$ since Φ is a $\{0, 1\}$-valued function. Thus, we have

$$\mathsf{E}S_{\mathcal{C}}(d) = \binom{M}{2}\mathsf{E}\Phi \doteq 2^{N(2R-1+\mathcal{H}(\delta))}$$

and, since the variance of a sum of uncorrelated random variables is equal to the sum of their variances

$$\mathrm{Var}(S_{\mathcal{C}}(d)) = \binom{M}{2}\mathrm{Var}(\Phi) < \mathsf{E}S_{\mathcal{C}}(d).$$

The minimum distance $d_{\min}(\mathcal{C})$ of the code is equal to the minimum d such that $S_{\mathcal{C}}(d) \neq 0$. The following theorem shows that the relative minimum distance of a random code is highly likely to be near $\delta_{\rm GV}(2R)$ as $N \to \infty$, and moreover, that the exponent of the distance distribution is highly likely to be near that of the average distribution wherever the exponent is positive.

Theorem 2.1: For $0 \leq R < \frac{1}{2}$ and any $\varepsilon > 0$, the probability that a code of length N and rate R from the RCE has relative minimum distance less than $\delta_{\rm GV}(2R) - \varepsilon$ goes to zero exponentially as $N \to \infty$. For $0 \leq R < 1$, if $d = N\delta$ is such that

$$\delta_{\rm GV}(2R) + \varepsilon \leq \delta \leq 1 - \delta_{\rm GV}(2R) - \varepsilon$$

then the probability that the number of codeword pairs at distance d satisfies $S_{\mathcal{C}}(d) \doteq 2^{N(2R-1+\mathcal{H}(\delta))}$ goes to one as $N \to \infty$.

Proof: For a given value of the code rate R, choose d so that $\frac{d}{N} \to \delta \leq \delta_{\rm GV}(2R) - \varepsilon$. Then

$$\Pr\{S_{\mathcal{C}}(d) \geq 1\} \leq \mathsf{E}S_{\mathcal{C}}(d) \doteq 2^{-N(1-\mathcal{H}(\delta)-2R)} \to 0.$$

In other words, with probability differing from 1 by an exponentially falling quantity there will be no pairs at distance d. Conversely, if $\delta_{\rm GV}(2R) + \varepsilon < \delta < 1 - \delta_{\rm GV}(2R) - \varepsilon$, then $1 - \mathcal{H}(\delta) < 2R$, and the average number of pairs $\mathsf{E}S_{\mathcal{C}}(d)$ at distance d is exponentially large. By the Chebyshev inequality, for any $\alpha > 0$, we have

$$\Pr\left\{|S_{\mathcal{C}}(d) - \mathsf{E}S_{\mathcal{C}}(d)| \geq \binom{M}{2}\alpha\right\} \leq \frac{\mathsf{E}\Phi}{\binom{M}{2}\alpha^2}.$$

If we choose $\alpha \doteq 2^{-N(1-\mathcal{H}(\delta)+\Delta)} < \mathsf{E}\Phi$ for any $\Delta > 0$, then we obtain

$$\Pr\left\{|S_{\mathcal{C}}(d) - \mathsf{E}S_{\mathcal{C}}(d)| > \binom{M}{2}\alpha\right\} \leq \frac{2\mathsf{E}\Phi}{M(M-1)\alpha^2} \doteq 2^{-N(2R-1+\mathcal{H}(\delta)-2\Delta)}. \quad (2.1)$$

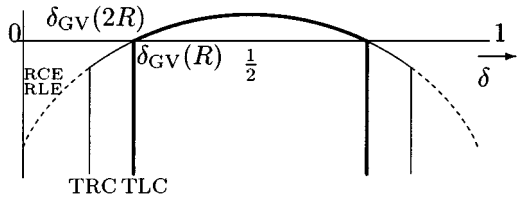

Fig. 2. Exponents of average distance distributions $\mathcal{N}_{\rm RCE}(d)$ and $\mathcal{N}_{\rm LCE}(d)$ for the RCE and the random LCE, and of distance distributions $\mathcal{N}_{\rm TRC}(d)$ and $\mathcal{N}_{\rm TLC}(d)$ for a TRC and a TLC.

The exponent on the right-hand side can be made positive by choosing Δ small enough. This establishes the fact that $S_{\mathcal{C}}(d) \doteq 2^{N(2R-1+\mathcal{H}(\delta))}$ for the chosen value of d with probability tending to one as $N \to \infty$. □

We can express this result succinctly as follows.

- With probability $1 - 2^{-\Omega(N)}$ as $N \to \infty$, the relative minimum distance of a code drawn from the RCE will be approximately $\delta_{\rm GV}(2R)$ for $0 \le R \le \frac{1}{2}$ and 0 for $\frac{1}{2} \le R \le 1$.

Define the *distance distribution* of $\mathcal{C}$ as

$$\mathcal{N}_{\mathcal{C}}(d) = \frac{2}{M} S_{\mathcal{C}}(d) \qquad (d = 0, 1, \ldots, N).$$

Thus, $\mathcal{N}_{\mathcal{C}}(d)$ is the average over the M codewords $\boldsymbol{x}_i$ of the number of other codewords $\boldsymbol{x}_j$, $j \ne i$, at Hamming distance d from $\boldsymbol{x}_i$.

We have shown that the complete average distance distribution over the RCE is

$$\begin{aligned}\mathcal{N}_{\rm RCE}(d) &= \frac{2}{M}\,\mathsf{E}S_{\mathcal{C}}(d) \\ &\doteq 2^{N(R-1+\mathcal{H}(\delta))} \qquad (d = 0, 1, \ldots, N). \end{aligned} \tag{2.2}$$

Since the probability in (2.1) tends to zero exponentially and since there are only $N+1$ different values of the distance d, Theorem 2.1 implies that for almost all codes in the RCE $\mathcal{N}_{\mathcal{C}}(d) \doteq \mathcal{N}_{\rm RCE}(d)$ for all d such that $\delta_{\rm GV}(2R) + \varepsilon \le \delta \le 1 - \delta_{\rm GV}(2R) - \varepsilon$. However, for $\delta \le \delta_{\rm GV}(2R) - \varepsilon$ or $\delta \ge 1 - \delta_{\rm GV}(2R) + \varepsilon$, the TRC has $\mathcal{N}_{\rm TRC}(d) = 0$. In short, the TRC has distance distribution

$$\mathcal{N}_{\rm TRC}(d) \begin{cases} \doteq 2^{N(R-1+\mathcal{H}(\delta))}, & \left|\frac{1}{2} - \delta\right| \le \frac{1}{2} - \delta_{\rm GV}(2R) - \varepsilon \\ = 0, & \left|\frac{1}{2} - \delta\right| \ge \frac{1}{2} - \delta_{\rm GV}(2R) + \varepsilon. \end{cases}$$

The exponents of $\mathcal{N}_{\rm RCE}(d)$ and $\mathcal{N}_{\rm TRC}(d)$ are plotted in Fig. 2 as functions of the continuous variable δ. The exponent $R - 1 + \mathcal{H}(\delta)$ of the RCE is nonnegative for $|\frac{1}{2} - \delta| \le \frac{1}{2} - \delta_{\rm GV}(2R)$, continuous, strictly concave and symmetrical about $\delta = \frac{1}{2}$. The exponent for the TRC is equal to the RCE exponent for $|\frac{1}{2} - \delta| \le \frac{1}{2} - \delta_{\rm GV}(2R) - \varepsilon$; however, for $\delta \downarrow \delta_{\rm GV}(2R)$ or $\delta \uparrow 1 - \delta_{\rm GV}(2R)$ it goes to $-\infty$.

C. Random Binary Linear Codes

A binary linear code $\mathcal{C}$ of length N and rate K/N is a (multi)set of $M = 2^K$ binary N-tuples that is generated by K N-tuples $\boldsymbol{g}_j$, $1 \le k \le K$; i.e., $\mathcal{C}$ is the set of all binary linear combinations

$$\boldsymbol{x}(\boldsymbol{u}) = \sum_k u_k \boldsymbol{g}_k$$

where $\boldsymbol{u}$ is an arbitrary binary K-tuple.

We will consider the random LCE in which each of the N bits in each of the K generators is chosen independently at random with equal probability of being a 0 or a 1. Equivalently, each of the 2^{NK} possible $K \times N$ matrices $G = \{\boldsymbol{g}_k, 1 \le k \le K\}$ is assigned probability 2^{-NK}. An exponentially small fraction of the resulting codes will have dimension less than K (and thus, $d = 0$), but this will not affect our development. We define the code rate as $R = K/N$ bits per symbol.

(In [5], Gallager considers a similar ensemble of linear codes defined by the $2^{N(N-K)}$ possible $N \times (N-K)$ parity-check matrices, and derives an asymptotic lower bound $\delta_{\rm GV}(R)$ on the relative minimum distance of a TLC.)

In a linear code, the distribution of distances $\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j), j \ne i\}$ from any given codeword $\boldsymbol{x}_i$ is independent of i. The average distance distribution of a linear code $\mathcal{C}$ therefore reduces to

$$\mathcal{N}_{\mathcal{C}}(d) = \sum_{j \ne i} \Phi\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = d\} \qquad (d = 1, 2, \ldots, N) \tag{2.3}$$

where $\boldsymbol{x}_i$ is an arbitrary codeword. Typically, $\boldsymbol{x}_i$ is taken as the all-zero codeword $\boldsymbol{0} = \boldsymbol{x}(\boldsymbol{0})$.

In this ensemble, if $\{\boldsymbol{u}_k\}$ is a set of $K' \le K$ linearly independent information K-tuples, then the K' corresponding codewords $\{\boldsymbol{x}(\boldsymbol{u}_k)\}$ are equally likely to be any of the $2^{NK'}$ possible sets of K' binary N-tuples. In other words, the probability distribution over any such set is the same as that in the RCE. In particular, the K' codewords $\{\boldsymbol{x}(\boldsymbol{u}_k)\}$ are statistically independent.

Therefore, if $(\boldsymbol{u}_j, \boldsymbol{u}_k)$ is any pair of distinct nonzero K-tuples, then the corresponding codewords $(\boldsymbol{x}(\boldsymbol{u}_j), \boldsymbol{x}(\boldsymbol{u}_k))$ are a pair of independent random binary N-tuples. It follows that two distinct distances $d_H(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $d_H(\boldsymbol{x}_i, \boldsymbol{x}_k)$ from a given codeword $\boldsymbol{x}_i$ (e.g., the all-zero codeword $\boldsymbol{0}$) are pairwise-independent and distributed as in the RCE. In particular

$$\Pr\{d_H(\boldsymbol{x}_i, \boldsymbol{x}_j) = N\delta\} \doteq 2^{-N(1-\mathcal{H}(\delta))}.$$

For any $d = N\delta$, the quantity $\mathcal{N}_{\mathcal{C}}(d)$ in (2.3) is thus a sum of $M - 1 \doteq 2^{NR}$ pairwise-independent, identically distributed random variables with mean $\mathsf{E}\Phi \doteq 2^{-N(1-\mathcal{H}(\delta))}$ and variance $\mathrm{Var}(\Phi) = \mathsf{E}\Phi - (\mathsf{E}\Phi)^2$. Its mean value is thus again equal to

$$\mathcal{N}_{\rm LCE}(d) \doteq 2^{N(R-1+\mathcal{H}(\delta))} \tag{2.4}$$

the same as for the RCE. Moreover, its variance for any d is upper-bounded by its mean.

We can now repeat the argument of Theorem 2.1 and arrive at analogous conclusions for the LCE. Namely, if the exponent $R - 1 + \mathcal{H}(\delta)$ of (2.4) is greater than 0, then with high probability there will be exponentially many codewords at distance $d = N\delta$ from a given codeword. On the other hand, if $1 - \mathcal{H}(\delta) > R$, then the average number of codewords at distance d from a given codeword will be exponentially small, and with probability approaching 1 there will be no such codewords at distance d. In short, the relative minimum distance of a code chosen at random from the LCE will be, with probability $1 - 2^{-\Omega(N)}$, approximately equal to the GV relative distance $\delta_{\rm GV}(R)$.

The typical random linear code from this ensemble has the distance distribution $\mathcal{N}_{\rm TLC}(d)$, $d = 1, 2, \ldots, N$, where

$$\mathcal{N}_{\rm TLC}(d) \begin{cases} \doteq 2^{N(R-1+\mathcal{H}(\delta))}, & \left|\frac{1}{2} - \delta\right| \le \frac{1}{2} - \delta_{\rm GV}(R) - \varepsilon \\ = 0, & \left|\frac{1}{2} - \delta\right| \ge \frac{1}{2} - \delta_{\rm GV}(R) + \varepsilon. \end{cases}$$

The exponent of $\mathcal{N}_{\rm TLC}(d)$ is also plotted in Fig. 2 as a function of a continuous variable δ. It is equal to the other three exponents for $\delta_{\rm GV}(R) + \varepsilon \le \delta \le 1 - \delta_{\rm GV}(R) - \varepsilon$; however, for $\delta \downarrow \delta_{\rm GV}(R)$ or $\delta \uparrow 1 - \delta_{\rm GV}(R)$ it goes to $-\infty$.

In summary, the typical minimum distance in the LCE is better than that in the RCE because it is the minimum of only $M - 1$ pairwise-independent distances, whereas in the RCE it is the minimum of $\binom{M}{2}$ pairwise-independent distances. In fact, the typical random linear code has a minimum distance equal to the GV distance, which many people (e.g., [9]) have conjectured to be the best possible asymptotic minimum distance of binary codes.

III. Error Exponents for the BSC

Now, using the distance distributions $\mathcal{N}_{\rm RCE}$, $\mathcal{N}_{\rm TRC}$, $\mathcal{N}_{\rm LCE}$, and $\mathcal{N}_{\rm TLC}$, we will find the true error exponents $E_{\rm RCE}(R)$, $E_{\rm TRC}(R)$, $E_{\rm LCE}(R)$, and $E_{\rm TLC}(R)$ for each of these ensembles and typical codes. We will find that the difference in distance distribution exponents illustrated in Fig. 2 is reflected in a difference in error exponents at low rates.

The error exponent of a family of codes $\mathcal{C}$ of rate R and increasing lengths N is defined as

$$E_{\mathcal{C}}(R) = \lim_{N\to\infty} -\frac{1}{N}\log \Pr(E)$$

where $\Pr(E)$ is the average probability of decoding error of a maximum-likelihood decoder, which is assumed to decrease exponentially with N, and where the limit is assumed to exist.

For brevity, we will take the following facts as given.

- Above a certain rate called $R_{\rm crit}$, the RCE and in fact all of these ensembles and codes achieve a certain error exponent called the sphere-packing exponent $E_{\rm sp}(R)$, which is known to be the best possible error exponent for rates $R \geq R_{\rm crit}$.
- For $0 \leq R \leq R_{\rm crit}$, a union bound analysis as below yields the true error exponent for the RCE, and in fact for all of these ensembles and codes.

The last fact is proved for the RCE in [7], and can be deduced for the TRC from [4]. Proofs of this fact for the LCE and the TLC are well known in the information theory community, although perhaps not published explicitly. This fact is also closely related to the Bassalygo–Elias bound on codes [2], and it explains why the Elias distance appears in Section III-A as the typical weight of incorrect codewords at $R = R_{\rm crit}$.

Given two codewords $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ at Hamming distance $d = d_H(\boldsymbol{x}_i, \boldsymbol{x}_j)$, if $\boldsymbol{x}_i$ is transmitted over a BSC with crossover probability p, the probability that the received word $\boldsymbol{y}$ will be at least as close to $\boldsymbol{x}_j$ as to $\boldsymbol{x}_i$ is

$$\Pr\{\boldsymbol{x}_i \to \boldsymbol{x}_j\} \doteq \binom{d}{\lceil d/2\rceil} p^{\lceil d/2\rceil}(1-p)^{\lfloor d/2\rfloor} \doteq 2^{-dD(\frac{1}{2}\|p)}$$

where

$$D\left(\frac{1}{2}\middle\|p\right) = \frac{1}{2}\log\frac{\frac{1}{2}}{p} + \frac{1}{2}\log\frac{\frac{1}{2}}{1-p} = -\log 2\sqrt{p(1-p)}\,.$$

The union bound estimate of $\Pr(E)$ is simply the sum of all these pairwise error probabilities; i.e., if a code has average distance distribution $\mathcal{N}(d)$, then the union bound estimate is

$$\Pr_{\rm UBE}(E) = \frac{1}{M}\sum_i \sum_{j\neq i} \Pr\{\boldsymbol{x}_i \to \boldsymbol{x}_j\}$$
$$\doteq \sum_{d=1}^{N} \mathcal{N}(d) 2^{-dD(\frac{1}{2}\|p)}.$$

For instance, let us substitute here the distance distribution (2.2) of the RCE. As $N\to\infty$, the sum will be dominated by the minimum of the $N+1$ exponents $D(\delta\|\frac{1}{2}) + \delta D(\frac{1}{2}\|p) - R$ for $\delta = d/N$, $0 \leq d \leq N$. Switching to a continuous variable δ, we observe that $D(\delta\|\frac{1}{2})$ is a strictly convex function of δ and $\delta D(\frac{1}{2}\|p)$ is linear; thus, the exponent has a unique minimum. Setting the derivative of the exponent to zero, we find that its minimum occurs at

$$\delta_{\rm crit}(p) = \frac{2\sqrt{p(1-p)}}{1+2\sqrt{p(1-p)}}.$$

Moreover,

$$D\left(\delta_{\rm crit}(p)\middle\|\frac{1}{2}\right) + \delta_{\rm crit}(p) D\left(\frac{1}{2}\middle\|p\right)$$
$$= -\log\frac{1}{2}\left(1+2\sqrt{p(1-p)}\right) = R_0$$

where R_0 is the pairwise error exponent of a BSC with crossover probability p. Thus, for $0 \leq R \leq R_{\rm crit}$, the true error exponent of the RCE is

$$E_{\rm RCE}(R) = R_0 - R$$

a straight line of slope -1 which is equal to R_0 at $R = 0$.

Since the LCE has the same average distance distribution as the RCE, it has the same error exponent.

On the other hand, for a TRC with minimum distance $\delta_{\rm GV}(2R)$, the exponent of the union bound $\Pr_{\rm UBE}(E)$ has the form

$$\min_{\delta\geq\delta_{\rm GV}(2R)}\left[D\left(\delta\middle\|\frac{1}{2}\right) + \delta D\left(\frac{1}{2}\middle\|p\right) - R\right].$$

This minimum occurs for $\delta = \delta_{\rm crit}(p)$ if $\delta_{\rm crit}(p) \geq \delta_{\rm GV}(2R)$; otherwise, because the exponent is monotone increasing in δ, it will be dominated by the term with $\delta = \delta_{\rm GV}(2R)$. Thus, we obtain the following theorem.

Theorem 3.1: For $0 \leq R \leq R_{\rm crit}$, the true error exponent of a TRC is

$$E_{\rm TRC}(R) = \begin{cases} E_y(R), & 0 \leq R \leq R_x/2 \\ R_0 - R, & R_x/2 \leq R \leq R_{\rm crit} \end{cases} \tag{3.1}$$

where R_x is the rate for which $\delta_{\rm crit}(p) = \delta_{\rm GV}(R_x)$, and

$$E_y(R) = D\left(\delta_{\rm GV}(2R)\middle\|\frac{1}{2}\right) + \delta_{\rm GV}(2R) D\left(\frac{1}{2}\middle\|p\right) - R.$$

Similarly, for a typical random linear code with minimum distance $\delta_{\rm GV}(R)$ we have the following.

Theorem 3.2: For $0 \leq R \leq R_{\rm crit}$, the true error exponent of a typical random linear code is

$$E_{\rm TLC}(R) = \begin{cases} E_x(R), & 0 \leq R \leq R_x \\ R_0 - R, & R_x \leq R \leq R_{\rm crit} \end{cases} \tag{3.2}$$

where R_x is defined in Theorem 3.1 and

$$E_x(R) = D\left(\delta_{\rm GV}(R)\middle\|\frac{1}{2}\right) + \delta_{\rm GV}(R) D\left(\frac{1}{2}\middle\|p\right) - R$$
$$= -\delta_{\rm GV}(R)\log 2\sqrt{p(1-p)}\,.$$

We conclude that

- with probability $1-2^{-\Omega(N)}$, the error exponent of a random code (resp., random linear code) is given by (3.1) (resp., (3.2)).

The function $E_x(R)$ is the usual "expurgated exponent" for a BSC. Since $\delta_{\rm GV}(0) = \frac{1}{2}$, at $R = 0$ we have

$$E_x(0) = E_y(0) = -\frac{1}{2}\log 2\sqrt{p(1-p)}\,.$$

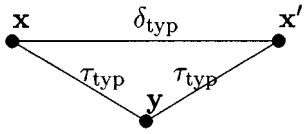

Fig. 3. Typical relative distances, given a decoding error, between transmitted (correct) codeword $\boldsymbol{x}$, received word $\boldsymbol{y}$, and decoded codeword $\boldsymbol{x}'$.

Note that $E_y(R)$ may be written in terms of the expurgated exponent $E_x(R)$ as follows:

$$E_y(R) = E_x(2R) + R, \qquad 0 \le R \le R_x/2.$$

Fig. 1 illustrates the error exponents $E_{\rm RCE}(R) = E_{\rm LCE}(R)$, $E_{\rm TRC}(R)$, and $E_{\rm TLC}(R)$. They coincide above the rate R_x. The error exponent of the TRC lies between the exponents of the RCE and the TLC for rates $0 < R < R_x/2$; it equals the TLC exponent at $R = 0$ and the RCE exponent at $R = R_x/2$.

A. Geometry of Typical Errors

The preceding calculation also yields the asymptotic values of (relative) distances between the transmitted codeword $\boldsymbol{x}$, incorrect codeword $\boldsymbol{x}'$, and the received vector $\boldsymbol{y}$ in a typical error event. Let $\delta_{\rm typ} = \delta_{\rm typ}(R)$ denote the typical value of the distance $(1/n)\, d_H(\boldsymbol{x}, \boldsymbol{x}')$ as N grows. Likewise, let $\tau_{\rm typ} = \tau_{\rm typ}(R)$ be the typical (relative) number of BSC channel errors. The relative distances $\delta_{\rm typ}$ and $\tau_{\rm typ}$ are illustrated in Fig. 3.

We observe that $\delta_{\rm typ} = \delta_{\rm crit}(p)$ for the RCE, the LCE, the TRC at rates $R_x/2 \le R \le R_{\rm crit}$, and the TRC at rates $R_x \le R \le R_{\rm crit}$, whereas $\delta_{\rm typ} = \delta_{\rm GV}(2R)$ for the TRC at rates $0 \le R \le R_x/2$ and $\delta_{\rm typ} = \delta_{\rm GV}(R)$ for the TLC at rates $0 \le R \le R_x$.

When $\delta_{\rm typ} = \delta_{\rm crit}(p)$, we have

$$\tau_{\rm typ} = \tau_{\rm crit}(p) = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}.$$

In terms of $\tau_{\rm crit}(p)$, the critical rate $R_{\rm crit}$ is the rate at which $\tau_{\rm crit}(p) = \delta_{\rm GV}(R_{\rm crit})$ [6]. For $R_{\rm crit} < R < C = 1 - \mathcal{H}(p)$, the decoding error probability is dominated by the probability that the number of channel errors will exceed $N\delta_{\rm GV}(R) < N\tau_{\rm crit}(p)$, in which case with overwhelming probability there will be an exponentially large number of incorrect codewords as close or closer to the received word as to the correct codeword. In this rate interval, the error exponent is therefore given by $E_{\rm sp}(R) = D(\delta_{\rm GV}(R)\|p)$, and the typical number of channel errors is $\tau_{\rm typ} = \delta_{\rm GV}(R)$. Since the correct word and the decoded word are equally likely to be in any direction from the received word and at distance $\tau_{\rm typ}$, the typical distance between the correct and the decoded word is

$$\delta_{\rm typ} = 2\tau_{\rm typ}(1 - \tau_{\rm typ}) = 2\delta_{\rm GV}(R)(1 - \delta_{\rm GV}(R))$$

which is sometimes called the Elias distance.

The typical relative distances $\tau_{\rm typ}$ and $\delta_{\rm typ}$ are illustrated in Fig. 4 as a function of R for both the TRC and the TLC, along with the GV relative distance $\delta_{\rm GV}(R)$, for a BSC with $p = 0.05$, which is representative of the general case.

B. Concluding Remarks

1) For all cases, for rates $0 \le R < R_{\rm crit}$, the decoding error probability is dominated by the probability that a single incorrect codeword is as close or closer to the received word as to the correct (transmitted) codeword. However, even for a TLC with relative minimum distance $\delta_{\rm GV}(R)$, errors are typically *not* made to minimum-distance codewords when $R > R_x$. Rather, for $R_x < R < R_{\rm crit}$, errors are typically made to codewords at distance $\delta_{\rm crit}(p) > \delta_{\rm GV}(R)$.
2) For $R_x \le R \le R_{\rm crit}$, the typical distance $\delta_{\rm typ}(R) = \delta_{\rm crit}(p)$ is not a function of R.
3) The typical random linear code not only achieves the GV distance, as has been noted previously [5], but also, as a result, it achieves the expurgated/random coding error exponent, which many conjecture to be the best possible. Moreover, it achieves this exponent without expurgation. In particular, the typical random linear code retains the usual symmetry properties of linear codes, which expurgated linear codes do not.
4) It is perhaps surprising that the typical random code is not as good as the typical random linear code. We have not been able to find this observation in the previous literature, although we believe that it was well known to researchers at MIT, IPPI (Moscow), and probably others.
5) It is perhaps also surprising that the TRC performs much better than the average performance over the RCE, at least at low rates. For rates $R < R_x/2$, the average performance over the RCE is dragged down by the performance of an exponentially small number of atypical bad codes.
6) It is straightforward to generalize these results to codes over an alphabet of arbitrary size q used over a q-ary symmetric channel. Moreover, the derivation of error exponents for an arbitrary discrete memoryless channel [3] is very similar to the argument of Section III. For that derivation, one employs codes chosen randomly with a uniform distribution from the set of all vectors of a fixed type (composition). In particular, the general GV bound and "distance distribution" of pairwise joint compositions possess the same properties as those discussed for the random ensemble of all binary codes.

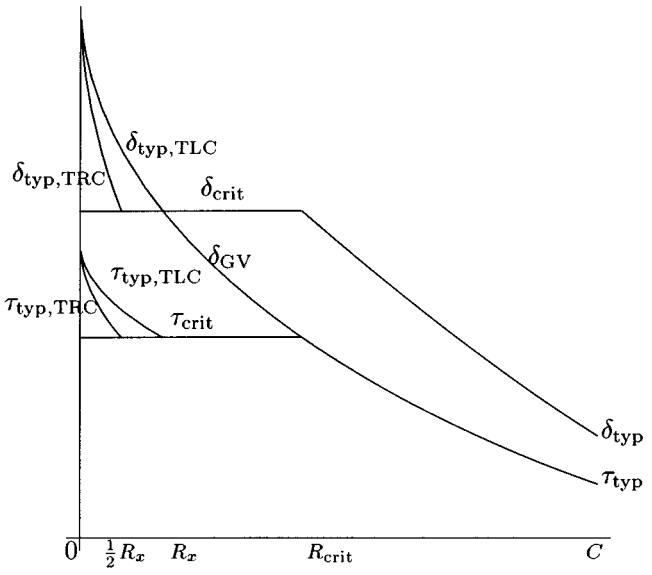

Fig. 4. Typical relative distances $\tau_{\rm typ}$ and $\delta_{\rm typ}$ (cf. Fig. 3) as functions of code rate R, for $p = 0.05$.

ACKNOWLEDGMENT

Section II-A was to some extent inspired by [1] which proves a lower bound $\delta_{\rm GV}(2R)$ on the relative distance of a TRC of rate R. The authors wish to thank M. Mézard for a stimulating lecture [8] on the random energy model (REM) of statistical physics, which closely resembles Shannon's RCE. They are also grateful to E. Arikan, R. Gallager, J. Massey, and A. Montanari for helpful comments.

REFERENCES

[1] L. A. Bassalygo, S. I. Gelfand, and M. S. Pinsker, "Simple methods for deriving lower bounds in the theory of codes" (in Russian), *Probl. Pered. Inform.*, vol. 27, no. 4, pp. 3–8, 1991. English translation in *Probl. Inform. Transm.*, vol. 27, pp. 277–281, 1991.

[2] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[3] I. Csiszár, "The method of types," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2505–2523, Oct. 1998.

[4] A. G. Dyachkov, "Bounds on the average error probability for a code ensemble with fixed composition" (in Russian), *Probl. Pered. Inform.*, vol. 16, no. 4, pp. 3–8, 1980. English translation in *Probl. Inform. Transm.*, vol. 16, pp. 255–259, 1980.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[6] ——, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[7] ——, "The random coding bound is tight for the average code," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 244–246, Mar. 1973.

[8] M. Mézard, "Aspects of spin glass theory," presented at the Workshop on Statistical Physics and Capacity-Approaching Codes, Trieste, Italy, May 2001.

[9] A. J. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

More Results on the Weight Enumerator of Product Codes

Ludo M. G. M. Tolhuizen, *Senior Member, IEEE*

***Abstract*—We consider the product code C_p of q-ary linear codes with minimum distances d_c and d_r. The words in C_p of weight less than $d_r d_c + \max(d_r\lceil \frac{d_c}{q}\rceil, d_c\lceil \frac{d_r}{q}\rceil)$ are characterized, and their number is expressed in the number of low-weight words of the constituent codes. For binary product codes, we give an upper bound on the number of words in C_p of weight less than $\min(d_r(d_c + \lceil \frac{d_c}{2}\rceil + 1), d_c(d_r + \lceil \frac{d_r}{2}\rceil + 1))$ that is met with equality if C_c and C_r are (extended) perfect codes.**

***Index Terms*—Product codes, weight enumerator.**

I. INTRODUCTION

If C_r and C_c are codes over the same alphabet, the product code $C_p = C_c \times C_r$ consists of all matrices with all rows in the row code C_r and all columns in the column code C_c. We will restrict ourselves to the case that both C_c and C_r are linear codes over the field $\mathbb{F}_q$, with parameters $[n_c, k_c, d_c]$ and $[n_r, k_r, d_r]$, respectively. Then C_p is an $[n_c n_r, k_c k_r, d_c d_r]$ code [1, Ch. 18, Sec. 2].

In this correspondence, we study weight enumerators of such product codes. For large codes, determination of the weight enumerator by enumeration of the codewords is computationally prohibitive. We are interested in deriving information on the weight enumerator of a product code from the weight enumerators of its—usually much smaller—constituent codes.

In [2] and [3], closed expressions are obtained for the weight enumerator of product codes for which the component codes belong to specific classes. In [4], [5], it was shown that the weight enumerator of C_p is not determined by the weight enumerators of C_c and C_r. Moreover, in this reference, the words in C_p of weight less than $d_r d_c + \max(d_r, d_c)$ were characterized, and the number of such codewords was expressed in the number of low-weight words of C_c and C_r.

In Section II, we extend the results from [4], [5]. We characterize the words of C_p of weight less than

$$w(d_r, d_c) = d_r d_c + \max\left(d_r \left\lceil \frac{d_c}{q} \right\rceil, d_c \left\lceil \frac{d_r}{q} \right\rceil\right)$$

where, as usual, $\lceil x \rceil$ is the smallest integer larger than or equal to x, and we express the number of such words in the number of low-weight words of the constituent codes. We also characterize the words of C_p of weight equal to $w(d_r, d_c)$ for the special case where $q = 2$ and d_r and d_c both are odd. In both examples in [4], [5], the number of product codewords of weight equal to $w(d_r, d_c)$ differ for different constituent codes with equal weight enumerator. As a consequence, the number of words in C_p of weight at least $w(d_r, d_c)$ in general cannot be derived from the weight distributions of the constituent codes.

In Section III, we move on to slightly higher weights in binary product codes. We obtain an upper bound on the number of product codewords of each weight w less than $v(d_r, d_c)$, where

$$v(d_r, d_c) = \min\left(d_r\left(d_c + \left\lceil \frac{d_c}{2} \right\rceil + 1\right), d_c\left(d_r + \left\lceil \frac{d_r}{2} \right\rceil + 1\right)\right).$$

The upper bound is met with equality if the constituent codes are perfect or extended perfect codes. In particular, we give explicit formulas for the number of words of weight less than 18 in the product of Hamming codes, and of weight less than 28 in the product of extended Hamming codes.

II. LOW-WEIGHT WORDS IN PRODUCT CODES

In this section, we characterize the low-weight words of a product code, and express their number in the number of low-weight words of the component codes.

In order to fix the notation, let C_c and C_r be linear codes over $\mathbb{F}_q$ with lengths n_c and n_r, and minimum distances d_c and d_r, respectively. We define

$$w(d_r, d_c) = d_r d_c + \max\left(d_r \left\lceil \frac{d_c}{q} \right\rceil, d_c \left\lceil \frac{d_r}{q} \right\rceil\right). \tag{1}$$

In this section, we will characterize the words in the product code $C_p = C_c \times C_r$ of weight less than $w(d_r, d_c)$.

For arbitrary vectors $\boldsymbol{a} \in \mathbb{F}_q^{n_r}$ and $\boldsymbol{b} \in \mathbb{F}_q^{n_c}$, $\boldsymbol{b} * \boldsymbol{a}$ is defined as the $n_c \times n_r$ matrix that has $b_i \boldsymbol{a}$ as ith row and $a_j \boldsymbol{b}$ as jth column, that is,

$$(\boldsymbol{b} * \boldsymbol{a})_{ij} = b_i a_j \text{ for all } (i, j) \in \{1, 2, \ldots, n_r\} \times \{1, 2, \ldots, n_c\}.$$

Clearly, for each $\boldsymbol{a} \in C_r$ and $\boldsymbol{b} \in C_c$, the matrix $\boldsymbol{b} * \boldsymbol{a}$ is in C_p. We call such words in C_p *obvious*. In [4], [5], we showed that all product codewords of weight less than $d_r d_c + \max(d_r, d_c)$ are obvious. In this section, we extend this result to higher weights.

The *support* of a vector $\boldsymbol{x}$, denoted by $\mathrm{supp}(\boldsymbol{x})$, is the index set of its nonzero positions, i.e.,

$$\mathrm{supp}(\boldsymbol{x}) = \{i | x_i \neq 0\}.$$

If C is a code of length n, and $I \subset \{1, 2, \ldots, n\}$, then $C(I)$ is the set of all words of C with their support inside I, that is,

$$\begin{aligned} C(I) &= \{\boldsymbol{c} \in C | \mathrm{supp}(\boldsymbol{c}) \subset I\} \\ &= \{\boldsymbol{c} \in C | c_j = 0 \text{ for all } j \in \{1, 2, \ldots, n\} \setminus I\}. \end{aligned}$$

Manuscript received December 21, 2000; revised April 16, 2002. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Cambridge, MA, August 1998.

The author is with Philips Research Laboratories, 5656 AA Eindhoven, The Netherlands (e-mail: ludo.tolhuizen@philips.com).

Communicated by J. Justesen, Associate Editor for Coding Theory.

Publisher Item Identifier 10.1109/TIT.2002.801476.