Distance Distribution of Binary Codes and the Error Probability of Decoding

Alexander Barg, *Senior Member, IEEE,* and Andrew McGregor

***Abstract*—We address the problem of bounding below the probability of error under maximum-likelihood decoding of a binary code with a known distance distribution used on a binary-symmetric channel (BSC). An improved upper bound is given for the maximum attainable exponent of this probability (the reliability function of the channel). In particular, we prove that the "random coding exponent" is the true value of the channel reliability for codes rate R in some interval immediately below the critical rate of the channel. An analogous result is obtained for the Gaussian channel.**

***Index Terms*—Binary-symmetric channel (BSC), channel reliability, distance distribution, union bound.**

I. Introduction

We consider transmission with binary codes of length n over a binary-symmetric channel with crossover probability p. Let $X = \{0,1\}^n$ be the n-dimensional Hamming space. Let $C(n, M = 2^{Rn}) \subset X$ be a code of rate R and let $x_i \in C$ be the transmitted vector. Under this condition the probability that a vector y is received equals

$$P(y|x_i) = p^{|y+x_i|}(1-p)^{n-|y+x_i|}$$

where $|\cdot|$ is the Hamming weight. For a given set $S \subset X$, let $\mathbb{P}_i(S) = \sum_{y \in S} P(y|x_i)$.

Let $D(x)$ be the decision region of maximum-likelihood decoding for the codevector x. Given that x_i is transmitted, the error probability of maximum-likelihood decoding equals $P_e(x_i) = \mathbb{P}_i(X \backslash D(x_i))$. The (average) error probability of decoding for the code C equals

$$P_e(C,p) = \frac{1}{M}\sum_{i=1}^{M} P_e(x_i).$$

Computing this probability directly is prohibitively difficult in most nontrivial examples, therefore, there has been much interest in bounding it from both sides. In what follows, we focus on *lower bounds* on $P_e(C,p)$. Recent papers devoted to this problem include [2], [7], [9], [17], [20], [21], [25].

The problem that we are considering is given the distance distribution of the code to derive a lower bound on $P_e(C,p)$. Although there have been other attempts to bound P_e below, the approach via the distance distribution seems to offer a right combination of detailed analysis and tractability. Under this approach, one usually begins with computing the probability that the received vector y is closer to some code vector x_j than to x_i. We then restrict our attention to when x_j is some specific value w away from x_i. Say there are B_w^i such code vectors. One would like then to bound the probability $P_e(x_i)$ below by the sum of probabilities of the events $\{|y+x_j| < |y+x_i|\}$ for all the B_w^i vectors x_j; the problem is however that these events are not disjoint. A simple way of dealing with this problem was suggested in Kounias [18]; papers [2], [21] essentially rely on a simplified version of the Kounias bound. Another method is based on de Caen's inequality [11] and its refinements in [19], [9]. Lower bounds on the error probability using this method for codes on the binary-symmetric channel (BSC) and the additive white Gaussian noise (AWGN) channel are derived in [9], [17], [20]. A third method was suggested in Burnashev [6] and used in [7] to refine the result of [2] on the reliability of the AWGN channel. In this paper, we adapt this method to the BSC case and derive a new asymptotic lower bound on the error probability of binary codes. The modification is not entirely straightforward and is explained in detail below.

A. Error Exponents

Optimizing $P_e(C,p)$ over all codes of a given rate R has received much attention in information and coding theory. It is known that for the best possible codes this probability declines as an exponential function of the code length. Let us define the largest attainable exponent of the error probability

$$E(R,p) = \limsup_{n\to\infty} \frac{1}{n} \log \max_{C \subseteq \{0,1\}^n, R(C)=R} \frac{1}{P_e(C,p)}$$

also called the error exponent or the reliability of the channel. The problem of bounding the function $E(R,p)$ for the binary-symmetric and other communication channels was one of the central problems of information theory in its first decades. In particular, the standard textbooks [4], [10], [14], [28] all devote considerable attention to properties and bounds for channel reliability. There are a variety of methods for deriving upper and lower estimates of $E(R,p)$. The most successful approaches to lower bounds are averaging over a suitably chosen ensemble of codes (for instance, all binary codes or all linear codes) [14] and relying on the distance distribution of an average code in

Manuscript received March 25, 2004; revised April 21, 2005. The work of A. Barg was supported in part by the National Science Foundation under Grant CCR-031096 and by Minta Martin Aeronautical Research Fund of the University of Maryland. It was performed in part while he was at DIMACS, Rutgers University, Piscataway, NJ. The work of A. McGregor was supported in part by the National Science Foundation under Grants CCR-031096 and ITR 0205456. The results of this paper were presented in part at the International Workshop on Coding and Cryptography, Paris, France, March 2003, and at the 2003 IEEE International Symposium on Information Theory, Yokohama, Japan, June /July 2003.

A. Barg is with Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail abarg@umd.edu).

A. McGregor is with University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail andrewm@cis.upenn.edu).

Communicated by A. E. Ashikhmin, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2005.858977

a code ensemble [13], [24]. Recently, the distance distribution approach was the subject of several papers because of the renewed interest in performance estimates of specific code families (rather than ensemble average estimates).

The problem of upper bounds on the error exponent $E(R,p)$ also has a long history. Several important ideas in this problem were suggested in the paper [27]. The nature of the upper bounds is different for low values of R and for R close to capacity. For low code rates, paper [27] suggested to bound the error probability below by the probability of making an error to a closest neighbor of the transmitted codeword.

B. Notation and Previous Results

Since our main result is a new bound on the error exponent $E(R,p)$, in this section we overview the known bounds on this function. It should be noted that the method that follows applies to the analysis of any code sequence for which the distance distribution is known or can be estimated.

For notational convenience, we shall write d_{ij} for the Hamming distance between two codewords x_i and x_j. We shall write d_{iy} for the distance between a codeword x_i and an arbitrary word y. Let $B_w^i = |\{x \in C : d_{ix} = w\}|$ and let $B_w = \sum_i B_w^i/M$ be the local and average distance distributions of the code C of size M.

Let $h(x)$ be the binary entropy and $h^{-1}(x)$ its inverse function. Denote by $\delta_{\rm GV}(R) := h^{-1}(1-R)$ the relative Gilbert–Varshamov distance corresponding to R and by

$$D(x\|y) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y}$$

the information divergence between two binomial distributions (the base of logarithms is 2 throughout). Let

$$A(\omega) := \omega \log 2\sqrt{p(1-p)} \tag{1}$$

$\varphi(x) = h(^1/_2 - \sqrt{x(1-x)})$. Throughout $w = \omega n$, $l = \lambda n$, and $d = \delta n$. Let $[n] = \{1, 2, \dots, n\}$.

For a given p, define

$$\rho = \rho(p) = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}.$$

The function

$$E_{\rm sp}(R,p) = D(\delta_{\rm GV}(R)\|p)$$

is called the sphere-packing exponent; it gives an upper bound on $E(R,p)$ which is valid for all code rates $R \in [0, 1-h(p)]$ and tight for code rates $R \geq R_{\rm crit}$, where the value $R_{\rm crit} = 1 - h(\rho)$ is called the critical rate of the channel. For low rates, the best known results for a long time were given by the following theorem.

Theorem 1:

$$-A(\delta_{\rm GV}(R)) \leq E(R,p) \leq -A(\bar\delta). \tag{2}$$

Here the lower bound is Gallager's "expurgation exponent" [13] obtained for instance for a sequence of linear codes whose minimum distance meets the Gilbert–Varshamov bound. The upper bound in (2) is due to [22]. It is obtained by substituting the result of [23] into the "minimum-distance bound" of [27]. The function $\bar\delta = \delta_{\rm LP}(R)$ is the linear programming bound of [23] on the relative distance of codes of rate R defined as

$$\bar\delta := \min_{0\leq\alpha\leq\frac12} G(\alpha,\tau)$$

where

$$G(\alpha,\tau) = 2\frac{\alpha(1-\alpha) - \tau(1-\tau)}{1 + 2\sqrt{\tau(1-\tau)}}$$

and where τ satisfies $h(\tau) = h(\alpha) - 1 + R$. Note that Theorem 1 implies that $E(0,p) = -A(1/2)$.

Let

$$\tau_\nu(\xi) := \frac12\Big(1 - \sqrt{1 - 4\big(\sqrt{\nu(1-\nu) - \xi(1-\xi)} - \xi\big)^2}\Big).$$

Let $\bar R(\delta)$ be the inverse function of $\bar\delta(R)$

$$\bar R(\delta) = 1 + \min_{(^1/_2)(1-\sqrt{1-2\delta})\leq\alpha\leq ^1/_2} (h(\tau_\alpha(\delta/2)) - h(\alpha)).$$

Derivation of improved upper bounds on $E(R,p)$ is based on the following inequality for the error probability $P_e(x_i)$ conditioned on transmission of the codeword x_i. For every $j \neq i$ let

$$X_{ij} \subset \{y \in X : d_{jy} \leq d_{iy}\}$$

be an arbitrary subset. Let $C' \subset C$ be an arbitrary subcode of C such that $x_i \notin C'$. Then

$$P_e(x_i) \geq \sum_{x_j\in C'} \Big\{\mathbb{P}_i(X_{ij}) - \sum_{x_k \in C'\setminus\{x_j\}} \mathbb{P}_i(X_{ij}\cap X_{ik})\Big\}. \tag{3}$$

Let us take C' to be the set of codeword neighbors of x_i at distance w from it. We have, for any w

$$P_e(x_i) \geq B_w^i \mathbb{P}_i(X_{ij}) \Big[1 - (B_w - 1)\mathbb{P}_i(X_{ik}|X_{ij})\Big]_+\Big\}$$

where x_j, x_k are any codewords such this $d_{ij} = d_{kj} = w$, $d_{jk} = d$, where d is the code's minimum distance, and $[a]_+ = \max(a, 0)$. Summing both sides of the last inequality on i from 1 to M, we obtain the estimate of $P_e(C)$ in the form

$$P_e(C) \gtrsim \max_w \Big\{B_w \mathbb{P}_i(X_{ij}) \Big[1 - (B_w - 1)\mathbb{P}_i(X_{ik}|X_{ij})\Big]_+\Big\}. \tag{4}$$

Recall from [27] that a straight-line segment that connects a point on $E_{\rm sp}(R',p)$ with a point on any other upper bound on $E(R,p)$, $R < R'$ is also a valid upper bound on $E(R,p)$. This result is called the *straight-line principle*. It is usually applied in situations when there is a ∪-convex upper bound on $E(R,p)$ and results into the straight-line segment given by the common tangent to this bound and the curve $E_{\rm sp}(R,p)$.

THE RESULTS OF [21]. The upper bound in (2) was improved in [21] by relying on estimates of the distance distribution of the code. The proof in [21] is composed of two steps. The first part is bounding the distance distribution of codes by a new application of the linear programming method (similar ideas were independently developed in [1]). The second step is using (3) to derive a bound on the error exponent. The estimate of the distance distribution of codes of [21] has the following form.

Theorem 2: [21] For any family of codes of sufficiently large length and rate R, any $\alpha \in [0,1/2]$, and any τ that satisfies $0 \le h(\tau) \le h(\alpha) - 1 + R$, there exists a value $\omega, 0 \le \omega \le G(\alpha,\tau)$ such that $n^{-1}\log B_{\omega n} \ge \mu(R,\alpha,\omega) - o(1)$, where

$$\mu(R,\alpha,\omega) = R - 1 + h(\tau) + 2h(\alpha) - 2q(\alpha,\tau,\omega/2) - \omega - (1-\omega)h\Big(\frac{\alpha-\omega/2}{1-\omega}\Big) \quad (5)$$

and where

$$q(\alpha,\tau,\omega) = h(\tau) + \int_0^\omega dy \log\frac{P+\sqrt{P^2-4Qy^2}}{2Q} \quad (6)$$

where

$$P = \alpha(1-\alpha) - \tau(1-\tau) - y(1-2y)$$
$$Q = (\alpha-y)(1-\alpha-y)$$

is the exponent of the Hahn polynomial $H_{\tau n}^{\alpha n}(\omega n)$.

The bound on $E(R,p)$ in [21] has the following form.

Theorem 3:

$$E(R,p) \le \min_{\alpha,\tau}\ \max_{0\le\delta\le\bar\delta}\ \max_{\delta\le\omega\le G(\alpha,\tau)} N \quad (7)$$

where

$$N = \min\{-A(\delta), -\min(\mu(R,\alpha,\omega), -B(\omega,\delta)) - A(\omega)\} \quad (8)$$

$0 \le \tau \le h^{-1}(h(\alpha)-1+R), 0 \le \alpha \le 1/2$; $A(w)$ is defined in (1), and where

$$B(\omega,\lambda) = -\omega - (1-\omega)h(p) + \max_{\eta\in[\frac{\lambda p}{2},\min(\frac{\lambda}{4},p(1-\omega))]}\Bigg(\lambda h\Big(\frac{2\eta}{\lambda}\Big) + (\omega-\lambda/2)h\Big(\frac{\omega-2\eta}{2\omega-\lambda}\Big) + (1-\omega-\lambda/2)h\Big(\frac{p(1-\omega)-\eta}{1-\omega-\lambda/2}\Big)\Bigg). \quad (9)$$

Remark: In [21], optimization in (7) involves taking a maximum on α and τ. However, Theorem 2 is valid for any $\alpha \in [0,1/2], \tau \in [0, h^{-1}(h(\alpha)-1+R)]$, and therefore, a better bound is generally obtained by taking a minimum rather than a maximum. Throughout the rest of the paper we will assume that $h(\tau) = h(\alpha) - 1 + R$. This assumption simplifies the analysis somewhat and does not seem to affect the final results.

Analysis of the inequality (4) together with some additional ideas gives rise to Theorem 3 and its improvements. We begin with deriving a simplified form of the bound (7) for low rates R.

C. A Study of the Bound (7)

By omitting the term $A(\delta)$ in (8), the expression for N can be written as

$$N = \max\{-\mu(R,\alpha,\omega) - A(\omega), B(\omega,\delta) - A(\omega)\}.$$

As will be seen later, for low rates R, the first term under the maximum is the greater one. For this reason, we begin with the study of the first term for low rates. Since this term does not depend on δ, we have

$$\max_{0\le\delta\le\bar\delta}\ \max_{\delta\le\omega\le G(\alpha,\tau)}(-\mu - A(\omega)) \le \max_{0\le\omega\le G(\alpha,\tau)}(-\mu - A(\omega)).$$

Lemma 4: Let $p \ge 0.037, 0 \le R \le \varphi(\delta_1)$, where $\delta_1 = 2\rho(1-\rho)$. Then

$$\max_{0\le\omega\le G(\alpha,\tau)}(-\mu - A(\omega)) = -A(\bar\delta) - R + 1 - h(\bar\delta). \quad (10)$$

Proof: In the expression $-\mu(R,\alpha,\omega) - A(\omega)$ let us take α equal to the value that furnishes the minimum in the definition of $\bar\delta$. Under the assumptions of the lemma, $R \le 0.303$. In this case, it is known that $\alpha = 1/2$ and the expression $q(\alpha,\tau,\omega/2)$ simplifies as follows. The integral in (6) upon a substitution $\alpha = \frac{1}{2}, 2y = z$, takes the form

$$\int_0^{\omega/2}\log\frac{P+\sqrt{P^2-4Qy^2}}{2Q}dy = \frac{1}{2}\int_0^\omega \log\Big[\frac{(1-2\tau)^2}{2(1-z)^2} + \frac{\sqrt{(1-2\tau)^2((1-2\tau)^2-4z(1-z))}-2z(1-z)}{2(1-z)^2}\Big]dz = \int_0^\omega \log\frac{1-2\tau+\sqrt{(1-2\tau)^2-4z(1-z)}}{2(1-z)}dz.$$

Let

$$k(\tau,\omega) = h(\tau) + \int_0^\omega \log\frac{1-2\tau+\sqrt{(1-2\tau)^2-4z(1-z)}}{2(1-z)}dz.$$

It is known [16] that in the region $0 \le \omega \le (1/2) - \sqrt{\tau(1-\tau)}$, this function gives the exponent of the Krawtchouk polynomial $K_{\tau n}(\omega n)$, i.e.,

$$\log K_{\tau n}(\omega n) = n(k(\tau,\omega) + o(1)).$$

Therefore, we obtain the identity $q(1/2,\tau,\omega/2) = k(\tau,\omega)$. Substituting this in μ we obtain the following:

$$-\mu - A(\omega) = -2h(\tau) + 2k(\tau,\omega) - \omega\log\sqrt{4p(1-p)}.$$

Let $g(\omega) = \frac{\partial}{\partial\omega}(-\mu - A(\omega))$. From the equation $g(\omega) = 0$ we find that the maximizing argument ω satisfies

$$1 - 2\tau - 2\sqrt{u}(1-\omega) = -\sqrt{(1-2\tau)^2 - 4\omega(1-\omega)}$$

where $u = \sqrt{4p(1-p)}$. This equation has a real zero if

$$\omega \le \bar\omega := 1 - \frac{1-2\tau}{2\sqrt{u}}$$

and then the maximizing argument is

$$\omega_*(\tau) = \frac{\sqrt{u}}{1+\sqrt{u}}\Big(1 - \frac{2\tau}{1-\sqrt{u}}\Big).$$

Recall that $0 \le \omega \le G(^1/_2,\tau) = \frac{1}{2} - \sqrt{\tau(1-\tau)}$. We shall show that

$$\arg \max_{0\le\omega\le G(^1/_2,\tau)} (-\mu - A(\omega)) = G(1/2,\tau). \quad (11)$$

There are two cases.

i) Let $R = \phi(\delta_1)$. In this case, the stationary point ω_* is exactly at the right end of the interval, i.e., $\omega_*(\tau) = \frac{1}{2} - \sqrt{\tau(1-\tau)}$. To show this, compute

$$\delta_1 = 2\rho(1-\rho) = \frac{u}{1+u}$$
$$\tau = h^{-1}(R) = \frac{1}{2} - \sqrt{\delta_1(1-\delta_1)} = \frac{(1-\sqrt{u})^2}{2(1+u)}$$

and substituting this into $\omega_*, \bar\omega$ we find

$$\omega_*(\tau) = \frac{\sqrt{u}}{1+\sqrt{u}}\Big(1 - \frac{1-\sqrt{u}}{1+u}\Big) = \delta_1 = \bar\omega.$$

ii) Now consider code rates $0 \le R < \varphi(\delta_1)$. Observe that $\tau = h^{-1}(R)$ decreases as R decreases, and therefore $\bar\omega$ also decreases with R. On the other hand, $\omega_*(\tau)$ increases as τ falls, so in this case $\bar\omega < \omega_*$, and $g(\omega)$ has no zeros for $0 \le \omega \le G(^1/_2,\tau)$. It is positive throughout because $g(0) > 0$. This again proves (11).

Hence, $-\mu - A(\omega)$ increases on ω for all $\omega \in [0, G]$, attaining the maximum at the right end of this segment. Substituting $\omega = G(^1/_2,\tau)$ into this expression, we obtain the claim of the lemma. □

For $R \ge 0.305$, the minimum in the definition of $\bar\delta$ is given by some $\alpha < 1/2$. Fixing α equal to this value we observe that the function μ depends only on ω. Therefore, the behavior of the function $-\mu(R,\alpha,\omega) - A(\omega)$ can be studied numerically (for instance, using Mathematica). We observe that this function increases on ω for $\omega \le \bar\delta(R)$ as long as $R \le \bar R(\delta_1)$. For $R = \bar R(\delta_1)$, the maximum of $-\mu(R,\alpha,\omega) - A(\omega)$ on ω is attained for $\omega = \bar\delta = \delta_1$. Substituting $\omega = \bar\delta$ into μ, we again arrive at the expression (10).

To summarize, the bound (7) implies the following: let $R \le \bar R(\delta_1)$, then

$$E(R,p) \le \max\Big\{ -A(\bar\delta) - R + 1 - h(\bar\delta), \max_{\delta,\omega}(-B(\omega,\delta) - A(\omega))\Big\}. \quad (12)$$

Next we argue that for low code rates, the maximum in this expression is given by the term $-A(\bar\delta) - R + 1 - h(\bar\delta)$. This is difficult to verify analytically because of the complicated form of the term B; however, this can be verified numerically for any given value of the probability p. More precisely, there exists a value of the rate $R = R_0$, a function of p, such that for $0 \le R \le R_0$, the first term is (12) is greater than the second one.

As a result, we obtain the following proposition.

Proposition 5: Let $\bar R(\delta_1) \le R_0$. Then

$$E(R,p) \le -A(\bar\delta) - R + 1 - h(\bar\delta) \quad 0 \le R \le R_0 \quad (13)$$
$$E(R,p) \le \max_{0\le\delta\le\bar\delta} \max_{\delta\le\omega\le\bar\delta} (B(\omega,\delta) - A(\omega)) \quad R_0 \le R. \quad (14)$$

The example of $p = 0.01$ is shown in Fig. 1.

Some comments are in order. The first term on the right-hand side of (3) is the "reverse union bound" which suggests to estimate the error rate $P_e(x_i)$ by a sum of pairwise error probabilities. An interesting fact is that for large n and for certain values of R and p the union bound argument gives the correct value of the error exponent. From (13) we can see this and more, namely, that for large n and code rates below R_0, the error exponent is given by the sum of pairwise probabilities of incorrect decoding to a codeword at the minimum distance of the code C from the transmitted codeword. (Note that the relative minimum distance of C is bounded above by $\bar\delta$.) The improvement of (13) over the upper bound in (2) is in that it takes into account decoding errors to all $\exp(n(R-1+h(\bar\delta)))$ neighbors of the transmitted vector as opposed to just one such neighbor in (2). The main question addressed below is to determine the range of code rates where the union bound and (13) is true and to refine the inequality (3) for those rates where the union bound does not apply.

In general terms, the answer to this question for large n is given by (4). The bound $P_e(C) \gtrsim B_w \mathbb{P}_i(X_{ij})$ is valid as long as

$$B_w \mathbb{P}_i(X_{ik} \cap X_{ij}) \lesssim \mathbb{P}_i(X_{ij}). \quad (15)$$

In our analysis, we use the estimation method of [6], [7] which was originally developed for codes on the sphere in $\mathbb{R}^n$. In the following, we modify it for use in the Hamming space and improve the estimate (7). The analysis of the relation between the distance distribution and $P_e(C,p)$ for the Hamming space turns out to be more difficult than for $\mathbb{R}^n$. One of the issues to be addressed is the choice of decision regions in the estimation process. We suggest one choice which while still being tractable leads to improving the estimates.

The results of the present paper are twofold: first, we expand the applicability limits of the bound (13). Outside these limits, we will derive a bound on $E(R,p)$ which is better than the result obtained from Theorem 3.

II. A New Bound

A. Statement of the Result

Let us state a lower bound for the error probability of maximum-likelihood decoding of an arbitrary sequence of codes with a given distance distribution.

Theorem 6: Let $(C_i)_{i\ge1}$ be a sequence of codes with rate R, relative distance δ, and distance distribution satisfying $B_{\omega n} \ge 2^{n\beta(\omega)-o(n)}$, where $\beta(\omega) > 0$ for all $\delta \le \omega \le 1$. The error probability of maximum-likelihood decoding of these codes satisfies $P_e(C,p) \ge 2^{-En-o(n)}$, where

$$E = \min_{\delta\le\omega\le1} \max_{\delta\le\lambda\le\omega} \big[\max(-\beta(\omega) - A(\omega), B(\omega,\lambda) - A(\lambda))\big] \quad (16)$$

where A and B are defined as in (1) and (9), respectively.

Theorem 6 will be proved later in this section. We first discuss its application to the problem of bounding $E(R,p)$. Let us specify this theorem for the distance distribution defined by Theorem 2. Let $\alpha, \tau, G(\alpha,\tau)$ have the same meaning as in (7).

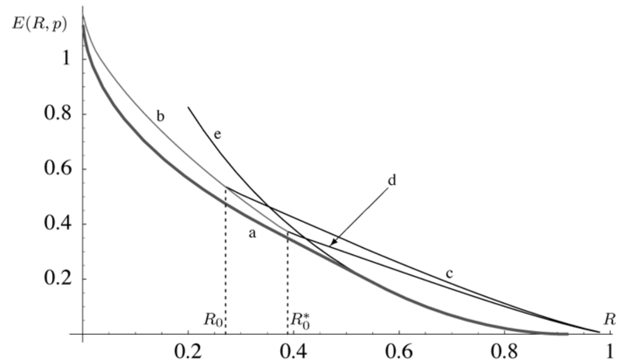

Fig. 1. Bounds on the error exponent for the BSC with $p = 0.01$. Notation explained in the text.

Recall that by Theorem 2, for any family of codes of rate R and every $\alpha \in [0, 1/2]$ there exists an $\omega, 0 \le \omega \le G(\alpha, \tau)$ such that the average number of neighbors at distance ωn can be bounded as $B_{\omega n} \ge 2^{n\mu(R,\alpha,\omega)-o(n)}$. Let us substitute this distance distribution in (16) and perform optimization. By Lemma 4 and the argument after it, for low values of R we conclude that the function $E(R,p)$ is bounded above by (10). Let R_0^* be the value of the rate, a function of p, for which the maximum shifts from the first term in (16) to the second one. As in the previous section, we arrive at the following theorem.

Theorem 7: Let $\bar{R}(\delta_1) \le R_0^*$. Then

$$E(R,p) \le -A(\bar{\delta}) - R + 1 - h(\bar{\delta}) \quad 0 \le R \le R_0^* \tag{17}$$

$$E(R,p) \le \max_{0 \le \lambda \le \bar{\delta}} \max_{\lambda \le \omega \le \bar{\delta}} B(\omega, \lambda) - A(\lambda) \quad R_0^* \le R \tag{18}$$

where A and B are defined as in (1) and (9), respectively.

Example: (Explanation of Fig. 1) To show that (16) improves over (7), let $p = 0.01$. Then from (13)–(14) we obtain $R_0 \approx 0.271$. From (16) we find that the bound (13) is valid for $R \le R_0^* \approx 0.388$. Note also that $R_{\text{crit}} = 0.559$, $\bar{R}(\delta_1) = 0.537$. See Fig. 1 for a graph of the known error bounds including our new bounds. In the figure, curve *a* is a combination of the best lower bounds on the error exponent. Curve *b* is the union bound of (13), (17). Curve *c* is the upper bound (14) given by Theorem 3, Proposition 5. Curve *d* is the upper bound (18) given by Theorem 6. Curve *e* is the sphere-packing bound $E_{\text{sp}}(R,p)$.

The improvement of Theorem 6 over Theorem 3 is in the extended region where the union bound *a* is applicable and in a better bound for greater values of the rate R.

Note that $E_{\text{sp}}(R,p)$ is better than *b* from $R \approx 0.422$; the straight-line bound (not shown) further improves the results.

Another set of examples together with some implications of Theorems 6 and 7 will be given in Section III.

Remark: Experience leads us to believe that the maximums in the equation are achieved for $\omega = \lambda = \bar{\delta}$ which would give us the bound

$$E(R,p) \le \begin{cases} -A(\bar{\delta}) - R + 1 - h(\bar{\delta}), & 0 \le R \le R_0^* \\ B(\bar{\delta}, \bar{\delta}) - A(\bar{\delta}), & R_0^* \le R. \end{cases}$$

However, this has proved too difficult to verify analytically due to the cubic condition for η in the maximization term in the definition of $B(\omega, \lambda)$ and other computational problems.

B. Preview of the Proof

The basic idea of the estimation method is from [7] although we make some modifications due to the fact that the observation space is discrete. To prove this theorem, we start by choosing a collection of sets $\{Y_{ij}\}$, each corresponding to a pair of codewords (x_i, x_j), such that Y_{ij} is outside the decoding region of x_i and

$$Y_{ij} \cap Y_{ik} = \emptyset, \qquad \text{for all } k \ne j.$$

Then we can bound the error probability in terms of these sets using the following inequality:

$$P_e \ge \frac{1}{M} \sum_{i=1}^{M} \sum_{j: d_{ij} = w} \mathbb{P}_i(Y_{ij}) \qquad (w = 1, 2, \ldots, n).$$

One of the main questions in applying this inequality and further ideas of [7] is the choice of the sets Y_{ij}. We construct the Y_{ij}'s via sets $X_{ij} \subset \mathbb{F}_2^n$, where

$$X_{ij} = \left\{ y \in F^n : d_{iy} = d_{jy} = \frac{d_{ij}}{2} + p(n - d_{ij}) \right\}.$$

See Fig. 2 for an illustration of the bounding process. To create the Y_{ij}'s from the X_{ij}'s we randomly "prune" these sets so that the disjointness condition is satisfied. To accomplish this pruning we define a set of codewords $T_i = \{x_j : d_{ij} = w\}$ for each codeword x_i. Then, as in [7], for each x_i, we randomly index by s_{ij} all the codewords x_j that are a distance w from x_i. Define sets

$$T(i,j) = \{k \in T_i : s_{ik} < s_{ij}\}.$$

We then get our Y_{ij}'s as follows:

$$Y_{ij} = X_{ij} \setminus [\cup_{k \in T(i,j)} X_{ik}].$$

These Y_{ij} satisfy the disjointness condition: assume there exists $x \in Y_{im} \cap Y_{il}$. Then $x \in X_{im}$ and $x \notin \bigcup_{k \in T(i,m)} X_{ik}$ gives that $s_{il} > s_{im}$. However, we also have $x \in X_{il}$ and $x \notin \bigcup_{k \in T(i,l)} X_{ik}$ and this gives that $s_{im} > s_{il}$ which is a contradiction.

Instead of calculating $\mathbb{P}_i(Y_{ij})$ directly we apply a "reverse union bound" to get

$$\mathbb{P}_i(Y_{ij}) \ge \mathbb{P}_i(X_{ij})(1 - K_{ij}) \tag{19}$$

where $K_{ij} = \sum_{k \in T(i,j)} \mathbb{P}_i(X_{ik} | X_{ij})$. Note that this inequality is the bound (3) with our particular choice of X_{ij}, Y_{ij}. Using the last inequality, we perform a recursive procedure which shows the existence of a subcode $C' \subset C$ with large error probability (among the codewords of C'). This gives the claimed lower bound on $P_e(C,p)$.

C. A Proof of Theorem 6

The error probability for two codewords is given by the following well-known lemma.

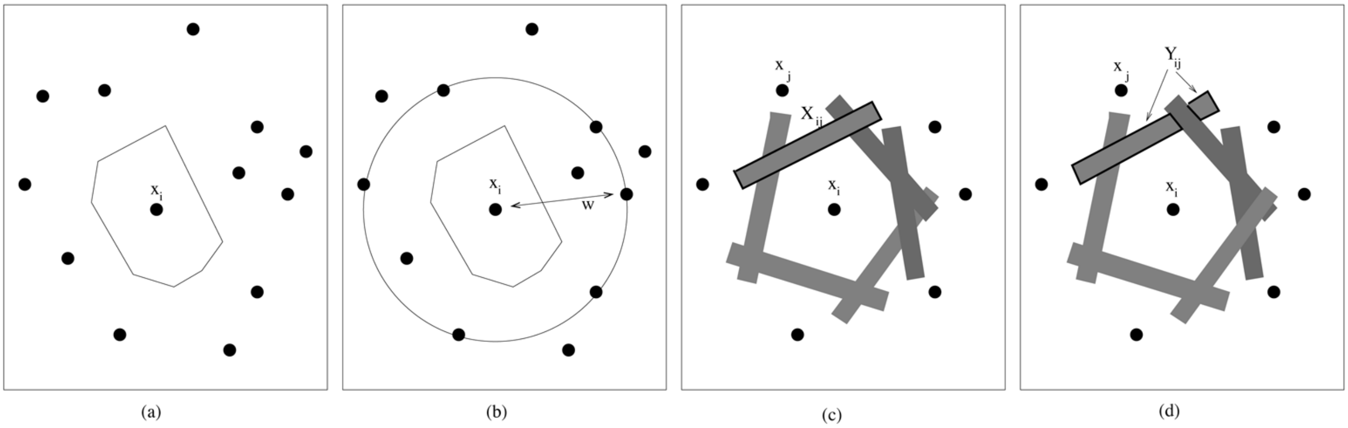

Fig. 2. The bounding process. (a) A codeword x_i, neighboring codewords, and the Voronoi region $D(x_i)$. (b) We restrict our attention to only those neighbors that are a distance w away. By only worrying that the received word y is closer to this subset of the neighbors, we upper-bound $\mathbb{P}_i(D(x_i))$. (c) For each neighbor x_j still under consideration, let X_{ij} be some set of words that are closer to x_j than they are to x_i. (d) We "prune" the X_{ij}'s to construct disjoint Y_{ij}'s with the required properties.

Lemma 8: For all codewords x_i and x_j that are a distance w apart

$$\lim_{n\to\infty}\frac{1}{n}\log P_i(X_{ij}) = A(\omega)$$

where $A(\omega)$ is defined in (1).

Lemma 9: For all codewords x_i, x_j, and x_k such that $d_{ij} = d_{ik} = w$ and $d_{jk} = l$ we have

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}_i\left(X_{ik}|X_{ij}\right) = B(\omega,\lambda)$$

where $B(\omega,\lambda)$ is defined in (9).

Proof: First consider

$$\mathbb{P}_i\left(X_{ik}\cap X_{ij}\right) = \sum_{m=0}^{\min(l/2,p(n-w))}\binom{l/2}{m}^2 \times\binom{w-l/2}{w/2-m}\binom{n-w-l/2}{p(n-w)-m} \times p^{w/2+p(n-w)}(1-p)^{n-w/2-p(n-w)}.$$

Then since

$$\log\mathbb{P}_i\left(X_{ik}|X_{ij}\right) = \log\mathbb{P}_i\left(X_{ik}\cap X_{ij}\right) - \log\mathbb{P}_i\left(X_{ij}\right),$$

substituting for $\mathbb{P}_i\left(X_{ij}\right)$ from the previous lemma and taking the appropriate limits gives the required result. □

The following properties of $B(\omega,\lambda)$ can be verified numerically.

Lemma 10: If $\omega\le\lambda\le 2\omega$, then $B(\omega,\lambda)\le B(\omega,\omega)$. If $\lambda\le\omega$, then $B(\lambda,\lambda)\le B(\omega,\lambda)$.

Recall that the indexing of pairs to create the sets $T(i,j)$ is done randomly. By linearity of expectation there exists an indexing such that

$$P_e \ge \frac{1}{M}\sum_{i=1}^{M}\sum_{j:d_{ij}=w}\mathbb{E}\left(\mathbb{P}_i\left(Y_{ij}\right)\right). \tag{20}$$

This equation will be the basis for our new bound on the error exponent but before deriving this bound we have two final preliminaries. First, we will refer to all codewords x_j that are a distance w from x_i as w-neighbors of x_i. (Recall that we defined B_w^i to be the number of codewords in the w-neighborhood of x_i.) Secondly, we shall say that a subset $S'\subseteq S$ of codewords is of *substantial* size (with respect to S) if its size has the same exponential order as the size of S. Note that for a family of codes $(C_i)_{i\ge1}$ where C_i has length n and rate R, we can consider $({C_i}')_{i\ge1}$, a family of codes where C_i' is a substantially sized subcode of C_i, when trying to bound the error exponent since

$$\lim_{n\to\infty}R(C_i') = \lim_{n\to\infty}R(C_i) = R$$

and

$$\limsup_{n\to\infty}\frac{1}{n}\log\frac{1}{P_e({C_i}',p)} \ge \limsup_{n\to\infty}\frac{1}{n}\log\frac{1}{P_e(C_i,p)}.$$

We now proceed with a case analysis dependent on the values of K_{ij}. Roughly speaking, when K_{ij} is typically less than a half, a union bound argument will be used to bound the error probability. When K_{ij} is typically larger than a half, a more complicated analysis will be required. Before we describe the two cases in our analysis we need the following two lemmas.

Lemma 11: [8] Suppose that there are L balls of K different colors. The number of balls of a color k is r_k. We are also given numbers n_k, $1\le k\le K$. Suppose that all balls are enumerated randomly by different integers from 1 up to L. Let τ be a random integer between 1 and L and let t_k be the number of balls of color k with numbers between 1 and τ. Then

$$\mathbb{P}\left(t_k\le n_k, k=1,\ldots K\right) \ge \min\left\{1, \frac{1}{4}\min_{1\le k\le K}\frac{n_k}{r_k}\right\}.$$

Recall that, for a given (i,j) pair, K_{ij} is a random variable. We then can prove the following lemma.

Lemma 12: Let $d_{ij} = \omega n$. With respect to the random indexing of all the (i,k) pairs (where x_k is any codeword such that $d_{ik} = \omega n$) we have

$$\mathbb{P}\left(K_{ij}\le\frac{1}{2}\right) \ge \min\left\{1, \min_{l\in\Lambda}\frac{2^{-nB(\omega,\lambda)-o(n)}}{\min\{B_w^i, B_l^j\}}\right\}$$

where

$$\Lambda = \{l\in[n] : |R_{w,l}| > N_{w,l}\}$$
$$R_{w,l} = \{x_k\in C : d_{ij} = d_{ik} = w, d_{jk} = l\}$$

and

$$N_{w,l} = \frac{2^{-nB(\omega,\lambda)}}{2(n+1)}.$$

Proof:

$$\begin{aligned}\mathbb{P}(K_{ij} \le 1/2) &= \mathbb{P}\left(\sum_{k\in T(i,j)} \mathbb{P}_i\left(X_{ik}|X_{ij}\right) \le 1/2\right)\\ &\cong \mathbb{P}\left(\sum_{l=0}^{n}\sum_{k\in T(i,j),d_{jk}=l} 2^{nB(\omega,\lambda)} \le 1/2\right)\\ &= \mathbb{P}\left(\sum_{l=0}^{n}|T(i,j)\cap R_{w,l}|2^{nB(\omega,\lambda)} \le 1/2\right)\\ &\ge \mathbb{P}\left(|T(i,j)\cap R_{w,l}| \le N_{w,l}\ \forall l\in\Lambda\right).\end{aligned}$$

Let there be a ball for each codeword in $\bigcup_l R_{w,l}$. Consider a ball from $R_{w,l}$ to have color l. Let $n_l = N_{w,l}$ and $\mu_l = |x_m \in R_{w,l} : s_{im} < s_{ij}|$. We have

$$\mathbb{P}\left(K_{ij} \le 1/2\right) \ge \mathbb{P}\left(\mu_l \le n_l\ \forall l \in \Lambda\right).$$

By the previous lemma we have

$$\mathbb{P}\left(\mu_l \le n_l\ \forall l\in\Lambda\right) \ge \frac{1}{4}\min_{l\in\Lambda}\frac{n_l}{|R_{w,l}|}$$

if the right-hand side is less than one. The lemma then follows from the fact that $|R_{w,l}| \le \min\{B_w^i, B_l^j\}$. □

In the analysis that leads to Theorem 6, we face a dichotomy of a relatively sparse w-neighborhood of the transmitted vector x_i when the union bound is asymptotically tight, and a cluttered neighborhood when is not. These two cases correspond to the first and the second terms in (16), respectively. When the union bound analysis is not applicable, we will rely crucially on the following lemma.

Lemma 13: If $K_{ij} > 1/2$ for some i, j such that $d_{ij} = \omega n$ then there exists a nonempty set Λ_{ij} such that for all $\lambda \in \Lambda_{ij}$

$$\min\{B_w^i, B_{\lambda n}^j\} > 2^{-nB(\omega,\lambda)-o(n)}.$$

Proof: Consider a pair of codewords x_i and x_j such that $K_{ij} > 1/2$. We deduce that $\mathbb{P}\left(K_{ij} \le 1/2\right) < 1$ since the event $\{K_{ij} > 1/2\}$ occurred. Therefore, by Lemma 12, there exists a λ such that

$$\frac{2^{-nB(\omega,\lambda)-o(n)}}{\min\{B_w^i, B_{\lambda n}^j\}} < 1.$$ □

Given a pair of codewords x_i, x_j with $K_{ij} \le 1/2$ we put $\Lambda_{ij} = \emptyset$; otherwise, we assume that Λ_{ij} contains all the values of $\lambda = l/n$ whose existence is established in the previous lemma. We now define, for all n possible values of $l = \lambda n$, the sets

$$G_{l,w} = \{x_j : \exists x_i \text{ such that } K_{ij} >^1/_2 \text{ and } l/n \in \Lambda_{ij}\}.$$

In words, for a given l, the set $G_{l,w} \subset C$ contains all the codewords x_j that have a w-neighbor x_i such that the set Λ_{ij} contains the value $\lambda = l/n$. Let $H_{l,w}$ be defined as the set of all $x_i \in C$ such that a substantial number of the w-neighbors x_j of x_i satisfy $K_{ij} >^1/_2$ and $l/n \in \Lambda_{ij}$. Note that the "substantial number" here is in relation to B_w^i.

We say $\lambda = l/n$ is a "nuisance level" for ω if $H_{l,w}$ and $G_{l,w}$ are both substantially sized subcodes of C. The two cases in the following analysis correspond to whether or not a nuisance level exists. The next theorem bounds the error probability in the case that it does not exist.

Theorem 14: Consider any code C of sufficiently large length n and rate R. Assume that for some ω and bounding function f we have $\frac{1}{n}\log B_{\omega n}^i \ge f(\omega)$ for all i. If there does not exist a nuisance level for ω then

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \le -f(\omega) - A(\omega) + o(1).$$

Proof: Let us define the sets

$$\begin{aligned}S_1 &= \{l : H_{l,w} \text{ is not a substantially sized subcode}\}\\ S_2 &= \{l : G_{l,w} \text{ is not a substantially sized subcode}\}.\end{aligned}$$

Since w does not have a nuisance level, $S_1 \cup S_2 = [n]$. Without loss of generality, we may assume that $G_{l,w} = \emptyset$ for all $l \in S_2$ since removing $\bigcup_{l\in S_2} G_{l,w}$ yields a substantially sized subcode. Hence, also $H_{l,w} = \emptyset$ for all $l \in S_2$. Now consider only transmitting the codewords in $C' = C \setminus \bigcup_{l\in[n]} H_{l,w}$ and note that this is a substantially sized number of codewords since neither $\bigcup_{l\in S_1} H_{l,w}$ nor $\bigcup_{l\in S_2} H_{l,w}$ are substantially sized. For each of these codewords we know that $\frac{1}{n}\log B_{\omega n}^i \ge f(\omega)$. Hence,

$$\begin{aligned}P_e(C,p) &\ge \frac{1}{M}\sum_{i=1}^{M}\sum_{j:d_{ij}=w}\mathbb{P}_i\left(Y_{ij}\right)\\ &\gtrsim \frac{1}{M}\sum_{x_i\in C'} B_w^i \min_{j:d_{ij}=w}\{\mathbb{P}_i\left(Y_{ij}\right)\}\\ &\ge \frac{1}{2}\min_{i,j:d_{ij}=w}\left(B_w^i\mathbb{P}_i\left(X_{ij}\right)\right)\\ &\ge 2^{n(A(\omega)+f(\omega))-o(n)}.\end{aligned}$$

The second inequality follows from the fact that for each $x_i \in C'$, a substantial number of w-neighbors x_j are such that $K_{ij} \le 1/2$, and the third one is implied by (19) since $\mathbb{P}_i\left(Y_{ij}\right) \ge \mathbb{P}_i\left(X_{ij}\right)/2$ whenever $K_{ij} \le 1/2$. □

We now bound the error probability (and ensure another property of the distance distribution) in the case that there exists a nuisance level.

Theorem 15: Consider any code C of sufficiently large length n and rate R and an $\omega \in [0,1]$. Let λ be a nuisance level for ω. The subset of codewords $x_j \in C$ such that

$$|\{x_k \in C : \mathrm{d}(x_j, x_k) = \lambda n\}| \ge 2^{-nB(\omega,\lambda)-o(n)}$$

forms a substantially sized subcode. Furthermore

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \le B(\omega,\lambda) - A(\omega) + o(1).$$

Proof: Since $G_{l,w}$ is substantially sized, it follows by Lemma 13 that a substantial number of codewords x_j have at least $2^{-nB(\omega,\lambda)-o(n)}$ neighbors at a relative distance λ. Now consider $x_i \in H_{l,w}$. By definition, there is a substantially sized subset $N(i)$ of the ωn-neighbors of x_i such that $\lambda \in \Lambda_{ij}$ for all $x_j \in N(i)$. Hence, appealing to Lemma 12, for each $x_j \in N(i)$

$$\mathbb{P}\left(K_{ij} \le 1/2\right) \ge \frac{2^{-nB(\omega,\lambda)-o(n)}}{B_w^i}.$$

Now

$$\begin{aligned}\mathbb{E}\left(\mathbb{P}_i\left(Y_{ij}\right)\right) &= \mathbb{E}\left(I_{K_{ij}\le\frac{1}{2}}\mathbb{P}_i\left(Y_{ij}\right)\right) + \mathbb{E}\left(I_{K_{ij}>\frac{1}{2}}\mathbb{P}_i\left(Y_{ij}\right)\right)\\ &\ge \mathbb{E}\left(I_{K_{ij}\le\frac{1}{2}}\mathbb{P}_i\left(Y_{ij}\right)\right)\\ &\ge \frac{2^{nA(\omega)}}{2}\mathbb{P}\left(K_{ij}\le\frac{1}{2}\right)\end{aligned}$$

and so, by the above discussion and (20), we get

$$\begin{aligned}P_e &\ge \frac{1}{M}\sum_{x_i\in H_{l,w}}\sum_{j\in N(i)}\mathbb{E}\left(\mathbb{P}_i\left(Y_{ij}\right)\right)\\ &\ge \frac{1}{M}\sum_{x_i\in H_{l,w}} 2^{nA(\omega)}2^{-nB(\omega,\lambda)-o(n)}\\ &= 2^{n(A(\omega)-B(\omega,\lambda))-o(n)}.\end{aligned}$$

□

Proof of Theorem 6: Let C be the code from the statement of the theorem. Let

$$F = \frac{1}{n}\log\frac{1}{P_e(C,p)}.$$

As discussed in [2], [7], for any $w = \omega n, \delta \le \omega \le 1$, the code C contains a subcode C' of size $M' \ge M/n^2$ such that for all codewords x_i in this subcode

$$\frac{1}{n}\log B_{\omega n}^i > \beta(\omega) - o(n).$$

Since the subcode is substantially sized we may now consider this subcode as our new code.

For a fixed ω, construct Y_{ij}, X_{ij}, and K_{ij} for all (i,j) pairs with $d_{ij} = \omega n$. By Theorems 14 and 15 we get

$$F \le \begin{cases}\beta(\omega)-A(\omega), & \text{if no nuisance level exists for } \omega\\ B(\omega,\lambda_1)-A(\omega), & \text{if a nuisance level } \lambda_1 \text{ exists for } \omega.\end{cases}$$

Hence, we get

$$F \le \max\{-\beta(\omega), B(\omega,\lambda_1)\} - A(\omega).$$

Now if $\lambda_1 \ge \omega$ then $B(\omega,\lambda_1) \le B(\omega,\omega)$ and so we get

$$F \le \max\{-\beta(\omega), B(\omega,\omega)\} - A(\omega). \tag{21}$$

If $\lambda_1 < \omega$, then we use the fact from Theorem 15 that for a substantial number of codewords x_i, $B_{\lambda_1 n}^i \ge 2^{-nB(\omega,\lambda_1)}$. We now construct new Y_{ij}, X_{ij}, and K_{ij} for all (i,j) pairs with $d_{ij} = \lambda_1 n$. Hence, by Theorems 14 and 15 we get

$$F \le \begin{cases}B(\omega,\lambda_1)-A(\lambda_1), & \text{if no nuisance level exists for } \lambda_1\\ B(\lambda_1,\lambda_2)-A(\lambda_1), & \text{if a nuisance level } \lambda_2 \text{ exists for } \lambda_1.\end{cases}$$

Hence, we get

$$F \le \max\{B(\omega,\lambda_1), B(\lambda_1,\lambda_2)\} - A(\lambda_1).$$

If $\lambda_2 \ge \lambda_1$ then $B(\lambda_1,\lambda_2) \le B(\lambda_1,\lambda_1) \le B(\omega,\lambda_1)$ then

$$F \le B(\omega,\lambda_1) - A(\lambda_1).$$

If $\lambda_2 < \lambda_1$, then we use the fact that for a substantial number of codewords x_i, $B_{\lambda_2 n}^i \ge 2^{-nB(\lambda_1,\lambda_2)}$ and continue as before.

We continue in this manner and get a sequence $\omega > \lambda_1 > \lambda_2 \ldots$ such that at step i we get the bound

$$F \le \max\{B(\lambda_{i-1},\lambda_i), B(\lambda_i,\lambda_{i+1})\} - A(\lambda_i).$$

This process terminates after at most n steps since there are only n possible values for the nuisance level. At the last step, $i = f$, the nuisance level λ_{f+1}, if it even exists, is not less than λ_f itself and, therefore, we have

$$\begin{aligned}F &\le \max\{B(\lambda_{f-1},\lambda_f), B(\lambda_f,\lambda_{f+1})\} - A(\lambda_f)\\ &\le \max\{B(\lambda_{f-1},\lambda_f), B(\lambda_f,\lambda_f)\} - A(\lambda_f)\\ &\le B(\omega,\lambda_f) - A(\lambda_f).\end{aligned}$$

Now for our code either this equation or (21) is valid, and so we have shown that for every $\omega, \delta \le \omega \le 1$ there exists $\lambda \le \omega$ such that

$$F \le \max(-\beta(\omega) - A(\omega), B(\omega,\lambda) - A(\lambda)).$$

This completes the proof. □

III. More on the Bound of Theorem (7)

In this section, we take a closer look at the bound (17) with the aim to show that it provides a new segment of code rates where the BSC channel reliability is known exactly. We rely on the notation of Section I-B. Let $R_x = 1 - h(2\rho(1-\rho))$. Recall that the best known lower bound on $E(R,p)$ below the critical rate is given by

$$E_x(R,p) = -A(\delta_{\rm GV}(R)), \qquad 0 \le R \le R_x \tag{22}$$

$$E_0(R,p) = D(\rho\|p) + R_{\rm crit} - R, \quad R_x < R \le R_{\rm crit}. \tag{23}$$

For $R > R_{\rm crit}$, the reliability function $E(R,p) = E_{\rm sp}(R,p)$. Note that both E_x and $E_{\rm sp}(R,p)$ can be viewed as instances of the union bound and that both are tangent on $E_0(R,p)$. Let us make one simple observation showing that the bound (17) has the same property.

The following lemma is verified by direct calculation.

Lemma 16: Let $\delta_1 = 2\rho(1-\rho)$ and let $R_1 = \bar{R}(\delta_1)$. Then

$$-A(\delta_1) - R_1 + 1 - h(\delta_1) = E_0(R_1,p).$$

Proof: Indeed, (23) can be rewritten as

$$E_0(R,p) = 1 - R + \log(1 + 2\sqrt{p(1-p)}).$$

The equality in the statement is equivalent to the relation

$$h(\delta_1) + \delta_1\log 2\sqrt{p(1-p)} = \log(1 + 2\sqrt{p(1-p)})$$

which is an easily verifiable identity. □

Next we can prove the main result of this section.

Theorem 17: Let $p, 0.046 \le p < 1/2$ be the channel transition probability. Then the channel reliability $E(R,p)$ equals the random coding exponent $E_0(R,p)$ for $R_1 \le R \le R_{\rm crit}$.

Proof: We check numerically that $R_1 < R_0^*$ for $p \ge 0.046$. Thus, by Theorem 7, for these values of p we have $E(R_1,p) = E_0(R_1,p)$. The full claim follows from the straight-line bound of Shannon, Gallager, and Berlekamp [27]. □

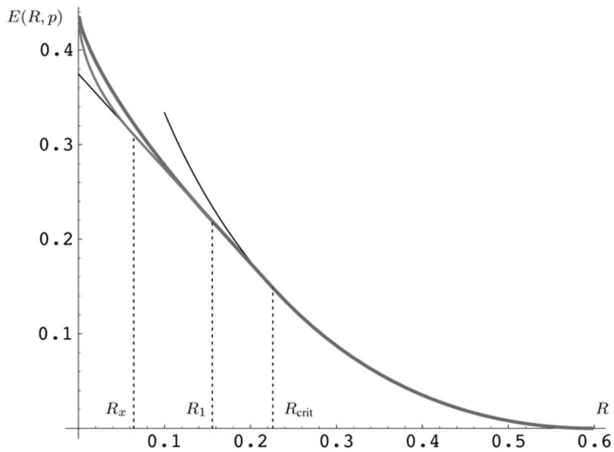

Fig. 3. Bounds on the error exponent for the BSC with $p = 0.08$. In the interval $R_1 \le R \le R_{\mathrm{crit}}$ the random coding bound $E_0(R,p)$ is tight.

Remark: We have seen in Lemma 4 that for $p \ge 0.037$, it suffices to rely on the simple form of the function $\bar{R}(x)$, namely, $R(x) = \varphi(x)$. Thus, the only numerical calculation involved in the proof of this theorem relates to the function $B(\omega, \delta)$.

The random coding exponent $E_0(R,p)$ gives the best known lower bound on $E(R,p)$ for $R_x \le R \le R_{\mathrm{crit}}$. The fraction of this segment in which Theorem 17 shows it to be tight is given by

$$\frac{R_{\mathrm{crit}} - R_1}{R_{\mathrm{crit}} - R_x}.$$

This fraction equals about $1/3$ for $p = 0.05$ and tends to one as $p \to {}^1/_2$.

We give an example of the new picture for the $E(R,p)$ function in Fig. 3. Previously the reliability of the BSC was known exactly only for $R \ge R_{\mathrm{crit}}$ [12].

IV. Random Linear Codes

The inequality of Theorem 6 can be used for a code with an arbitrary distance distribution. In this section, we are interested in the estimate of the error exponent for a random linear code C. Here by a random code we mean a binary code whose weight distribution behaves as the binomial distribution: $\mathcal{B}_{\omega n} \cong \exp[n(R + h(\omega) - 1)]$. The reason for calling this code random is that the weight distribution of a randomly chosen linear code with high probability converges to the binomial distribution (e.g., [3]).

The error exponent $\tilde{E}(R,p)$ for random linear codes for low rates is bounded below by the expurgation exponent: $\tilde{E}(R,p) \ge -A(\delta_{\mathrm{GV}}(R))$. For $R_x \le R_0 \le R_{\mathrm{crit}}$, the exponent $\tilde{E}(R,p) \ge E_0(R,p)$. Moreover, it is known that the error probability $P_e(C)$ averaged over the ensemble of all binary codes meets this bound with equality [15]. The proof of this result in [15] is accomplished by computing the ensemble average probability of error under list decoding into lists of size 2, where by error we mean the event that the transmitted codeword is not in the resulting list. It turns out that under this definition the error occurs in an exponentially smaller fraction of cases than the error of maximum-likelihood decoding. In other words, in all the cases of error under maximum-likelihood decoding (i.e., decoding into a size-1 list) except for an exponentially small fraction of them, there is exactly one codeword which is at least as close to the received word as is the transmitted word. This shows that for exponential asymptotics of the error probability of random codes the union bound is tight. An analogous result can also be proved for the ensemble of binary linear codes.

Here we compute a lower bound on the decoding error probability of a code with weight distribution $\mathcal{B}_{\omega n}$. A closed-form expression again seems beyond reach, however, computational evidence with the bound (16) suggests that in a certain segment of code rates $0 \le R \le R^{**}$, the error exponent of maximum-likelihood decoding of the code C is bounded above as follows:

$$\tilde{E}(R,p) \le -A(\delta_{\mathrm{GV}}(R)).$$

In other words, the expurgation exponent is tight for a random linear code in the region of low code rates.

V. The Gaussian Channel

Given the results for the BSC of Section III, it is natural to assume that qualitatively similar results hold for the reliability function of the Gaussian channel. Here, we consider briefly this problem and show that the random coding exponent is tight for a certain interval of rates immediately below the critical rate. As in the binary case, the length of this segment depends on the level of the channel noise.

Let a be the signal-to-noise ratio in the channel. Denote by $E(R,a)$ the channel reliability function defined analogously to the BSC case. It is known to be bounded below by the random coding bound $E_0(R,a)$ [26] which has the form

$$E_0(R,a) = \frac{a}{4}(1 - \cos\theta_x) + R_x - R$$

and is the best known lower bound for $R_x \le R \le R_{\mathrm{crit}}$, where

$$\begin{aligned} R_x &= \frac{1}{2}\ln\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{a^2}{4}}\right) \\ \theta_x &= \cos^{-1}\sqrt{1 - e^{-2R_x}} \\ R_{\mathrm{crit}} &= \frac{1}{2}\ln\left(\frac{1}{2} + \frac{a}{4} + \frac{1}{2}\sqrt{1 + \frac{a^2}{4}}\right). \end{aligned}$$

Let C be a code on $S^{n-1}(\mathbb{R})$ (the unit sphere in $\mathbb{R}^n$). Let $\theta(x_i, x_j)$ be the angle between the vectors that correspond to the codewords x_i, x_j. Denote by $B(\theta)$ the distribution of angular distances in the code C. The exponent of the union bound on the error probability $P_e(C,a)$ has the form

$$E_U = \frac{a}{4}(1 - \cos\theta) - \frac{1}{n}\ln B(\theta).$$

Used together with an estimate of the distance distribution of a code of rate R obtained in [2] this bound takes the form

$$E_U(R,a) = \frac{a}{4}(1 - \cos\bar{\theta}) - \ln\sin\bar{\theta} - R$$

where $\bar{\theta} = \bar{\theta}(R)$ is the root of the equation $R = \psi(\theta)$ and

$$\psi(x) = -\frac{1-\sin x}{2\sin x}\ln\frac{1-\sin x}{1+\sin x} - \ln\frac{2\sin x}{1+\sin x}$$

(which represents the Kabatiansky–Levenshtein bound on spherical codes). The strongest known condition for the union bound to be valid asymptotically as a lower bound on $P_e(C, a)$ was announced in [5]. According to it, $E(R, a) \le E_U(R, a)$ for all rates $R \le R^*$, where R^* is the root of

$$R + \ln\sin\bar{\theta}(R) = \frac{a}{8}(1-\cos\bar{\theta}). \tag{24}$$

Other conditions were obtained in [2], [7], [9].

Next we state a result analogous to Lemma 16. Its proof is immediate by comparing the expressions for E_U and E_0.

Lemma 18: Let $R_1 = \psi(\theta_x)$, then $E_0(R_1, a) = E_U(R_1, a)$.

We conclude that $E_0(R_1, a)$ is the correct value of $E(R_1, a)$ if $R_1 \le R^*$. The last inequality holds for $0 < a \le 5.7$. Coupled with the straight-line principle of [27] this gives the following result.

Theorem 19: Let $0 < a \le 5.7$ be the signal-to-noise ratio in the channel. Then

$$E(R, a) = E_0(R, a) \qquad (R_1 \le R \le R_c).$$

Example: Let $a = 2$. Then $R_x = 0.094$, $R_1 = 0.199$, $R^* = 0.263$, $R_{\text{crit}} = 0.267$.

If instead of (24) we rely on conditions with a published proof, we would still be able to make a tightness claim of E_0 but for a smaller segment of the signal-to-noise ratio values.

Final Note: Recently, a generalized de Caen inequality was used to derive lower estimates of error probability of a code via its distance distribution [9]. In particular, [9] gives a condition for the union bound to be valid asymptotically as a lower bound on P_e in the BSC case. Although the condition is stated as an optimization problem ([9, Proposition 5.3]), computational evidence suggests that its solution is given by (15). Thus, the methods of this paper and of [9], although different in nature, seem to lead to the same general estimates. Note that [9] does not contain results on the BSC reliability function.

References

[1] A. Ashikhmin and A. Barg, "Binomial moments of the distance distribution: Bounds and applications," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 438–452, Mar. 1999.

[2] A. Ashikhmin, A. Barg, and S. Litsyn, "A new upper bound on the reliability function of the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1945–1961, Sep. 2000.

[3] A. Barg and G. D. Forney Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.

[4] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[5] M. V. Burnashev, "On relation between code geometry and decoding error probability," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 133.

[6] ——, "A new lower bound for the α-mean error of parameter transmission over the white Gaussian channel," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 1, pp. 23–34, Jan. 1984.

[7] ——, "On the relation between the code spectrum and the decoding error probability," *Probl. Inf. Transm.*, vol. 36, no. 4, pp. 3–24, 2000.

[8] M. V. Burnashev and Y. A. Kutoyants, "On minimal α-mean error parameter transmission over a Poisson channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 2505–2515, Feb. 2001.

[9] A. Cohen and N. Merhav, "Lower bounds on the error probability of block codes based on improvements of de Caen's inequality," *IEEE Trans. Inf. Theory*, no. 2, pp. 290–310, Feb. 2004.

[10] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Channels*. Budapest, Hungary: Akadémiai Kiadó, 1981.

[11] D. de Caen, "A lower bound on the probability of a union," *Discr. Math.*, vol. 169, no. 1–3, pp. 217–220, 1997.

[12] P. Elias, "Coding for noisy channels," in *IRE Conv. Rec.*, Mar. 1955, pp. 37–46.

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

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

[15] ——, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 244–246, Mar. 1973.

[16] G. Kalai and N. Linial, "On the distance distribution of codes," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1467–1472, Sep. 1995.

[17] O. Keren and S. Litsyn, "A lower bound on the probability of error on a BSC channel," in *Proc. 21st IEEE Conv. Electrical and Electronic Engineers in Israel*, 2000, pp. 217–220.

[18] E. G. Kounias, "Bounds for the probability of a union, with applications," *Ann. Math. Statist.*, vol. 39, pp. 2154–2158, 1968.

[19] H. Kuai, F. Alajaji, and G. Takahara, "A lower bound on the probability of a finite union of events," *Discr. Math.*, vol. 215, no. 1–3, pp. 147–158, 2000.

[20] ——, "Tight error bounds for nonuniform signalling over AWGN channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2712–2718, Nov. 2000.

[21] S. Litsyn, "New upper bounds on error exponents," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 385–398, Mar. 1999.

[22] R. J. McEliece and J. K. Omura, "An improved upper bound on the block coding error exponent for binary-input discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 5, pp. 611–613, Sep. 1977.

[23] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, "New upper bound on the rate of a code via the Delsarte-MacWilliams inequalities," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 2, pp. 157–166, Mar. 1977.

[24] G. S. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1284–1292, Jul. 1994.

[25] G. E. Séguin, "A lower bound on the error probability for signals in white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3168–3175, Nov. 1998.

[26] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.

[27] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for codes on discrete memoryless channels, II," *Inf. Contr.*, vol. 10, pp. 522–552, 1967.

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