Distance distribution of binary codes and the error probability of decoding

Alexander Barg* 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. An improved upper bound is given for the maximum attainable exponent of this probability (the reliability function of the channel).

1 Introduction

1.1 Preliminaries

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 $\mathbb{P}_i(S) = \sum_{y \in S} P(y|x_i)$.

For a vector $x \in C$ define its Voronoi region $D(x)$ as follows:

$$D(x) = \{y \in X : \forall_{x' \in C \setminus x}\, \mathrm{d}(x,y) < \mathrm{d}(x',y)\}.$$

Given that x_i is transmitted the error probability of maximum likelihood decoding equals $P_e(x_i) = \mathbb{P}_i\left(X \setminus D(x_i)\right)$. The (average) error probability of decoding for the code C equals

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

Optimizing $P_e(C,p)$ over all codes of a given rate R has received much attention in information and coding theory [6], [10], [8], [7], [2], [4]. We will be interested in deriving lower bounds on the error probability of the best possible code of a given size used over a given channel.

Let $B_w, w = 0, 1, \ldots, n$ be the distance distribution of the code C. The problem that we are considering is given the distance distribution to derive a lower bound on $P_e(C,p)$. This problem can be reformulated with a geometric flavor: given the number B_w of the neighbors of a codeword x_i at a distance w, what can be said about the most unfavorable allocation of those neighbors in terms of the probability $\mathbb{P}_i(X \setminus D(x_i))$?

*DIMACS, Rutgers University, 96 Frelinghuysen Road, Piscataway, NJ 08854, abarg@dimacs.rutgers.edu.
†University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104, andrewm@cis.upenn.edu.

One thing is easy to compute, namely the probability that the received vector y is closer to some code vector x_j than to x_i. 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 vectors x_j; the problem is however that these events are not disjoint. We therefore face the following questions. First, is it true nevertheless that asymptotically for large n and for certain values of R and p the probability $P_e(x_i)$ essentially equals the sum of pairwise error probabilities? Next, if not then how such an estimate can be refined to obtain a valid bound on $P_e(x_i)$?

In this paper we will provide some answers to both questions improving over the previously known lower asymptotic bounds for the error probability of the best possible codes of a fixed rate R.

1.2 Notation and previous results

Let

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

be the largest attainable exponent of the error probability, also called the reliability of the channel.

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

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

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 high rates. For low rates the best known results for a long time were given by the following theorem.

Theorem 1.

$$-A(\delta_{\rm GV}(R)) \le E(R,p) \le -A(\delta_{\rm LP}(R)). \tag{2}$$

Here the lower bound is Gallager's "expurgation exponent" [6] and the upper bound is due to [10], [8]. The function $\delta_{\rm LP}(R)$ is the so-called JPL bound [9] on the relative distance of codes of rate R defined as

$$\delta_{\rm LP}(R) := \min_{0\le\tau\le\alpha\le\frac{1}{2}} G(\alpha,\tau)$$

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

The upper bound in (2) is based on the upper bound on the minimum distance of the code. The upper bound in (2) was improved in [7] by relying on estimates of the distance distribution of the code. The proof in [7] is composed of the 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 estimate of the distance distribution of codes of [7] has the following form.

Theorem 2. *[7] For any family of codes of sufficiently large length and rate R and any $\alpha \in [0, 1/2]$ 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*

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

$\tau = h^{-1}(h(\alpha) - 1 + R)$, and where

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

where $P = \alpha(1-\alpha) - \tau(1-\tau) - y(1-2y), Q = (\tau - y)(1 - \tau - y)$, is the exponent of the Hahn polynomial $H_{\tau n}^{\alpha n}(\omega n)$.

The second part of the proof in [7] is devoted to estimating the error probability of a code given its distance distribution. The same approach was used in [2] to derive analogous results for spherical codes and the Gaussian channel. The bound on the reliability of the Gaussian channel of [2] was improved in [4]. The improvement was obtained by using a more accurate method [3] of deriving bounds on $P_e(C)$ for a code with a known distance distribution than the one used in [7], [2].

The result of [7] has the following form (reformulated slightly from its original version):

Theorem 3.

$$E(R, p) \le \min_{0 \le \alpha \le 1/2} \max_{0 \le \delta \le \delta_{\mathrm{LP}}(R)} \max_{\delta \le \omega \le G(\alpha, \tau)} N \tag{3}$$

where

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

$A(w)$ is defined in (1),

$$B(\omega, \lambda) = -\omega - (1-\omega)h(p) + \max_{\eta \in [\frac{\lambda p}{2}, \min(\frac{\lambda}{4}, p(1-\omega))]} \left(\lambda h\left(\frac{2\eta}{\lambda}\right) + (\omega - \lambda/2)h\left(\frac{\omega - 2\eta}{2\omega - \lambda}\right) + (1 - \omega - \lambda/2)h\left(\frac{p(1-\omega) - \eta}{1 - \omega - \lambda/2}\right)\right). \tag{4}$$

Below we use the method of [3]-[4] to improve the estimate (3). 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.

As it turns out, for low rates the estimate (3) simplifies to

$$E(R,p) \quad \leq \quad -A(\delta_{\rm LP}(R)) - R + 1 - h(\delta_{\rm LP}(R)). \tag{5}$$

The improvement of (5) over the upper bound in (2) is in that it takes into account decoding errors to all $\exp(n(R-1+h(\delta_{\rm LP}(R))))$ neighbors of the transmitted vector as opposed to just one such neighbor in (2). The results of the present paper are twofold: first, we expand the applicability limits of the bound (5). Outside these limits we will derive a bound on $E(R,p)$ which is better than the result obtained from Theorem 3.

Recall from [10] that a straight-line segment that connects a point on $E_{\rm sp}(R',p)$ with a point on any other upper bound on $E(R,p)$ is also a valid upper bound on $E(R,p)$. In particular, the common tangent to (5) and $E_{\rm sp}(R,p)$ also gives an upper bound on $E(R,p)$.

2 A Study of the Bound (3)

Let us derive a simplified form of the bound (3)

Proposition 4. *For some R_0, a function of p, we have*

$$E(R,p) \leq \begin{cases} -A(\delta_{\rm LP}(R)) - R + 1 - h(\delta_{\rm LP}(R)) & 0 \leq R \leq R_0 \\ \max\limits_{0\leq\delta\leq\delta_{\rm LP}(R)} \max\limits_{\delta\leq\omega\leq\delta_{\rm LP}(R)} \min(-A(\delta), B(\omega,\delta) - A(\omega)) & R_0 \leq R. \end{cases} \tag{6}$$

Proof. (outline) First let us prove that the term $-\mu - A(\omega)$ in N can be brought to the form (5). For this let us take α equal to the value that furnishes the minimum in the definition of $\delta_{\rm LP}$. It is known that $\alpha = 1/2$ for $R \leq 0.305$. We then have $q(1/2,\tau,\omega/2) = k(\tau,\omega)$, where $k(\tau,\omega)$ is the exponent of the Krawtchouk polynomial $K_{\tau n}(\omega n)$. Substituting this in μ and taking the derivative on ω of $-\mu - \omega\log Z$ we find this function to be a growing function of ω. Hence the maximum on ω is obtained for $\omega = \delta_{\rm LP}(R)$. Substituting this together with $k(\tau,\omega)$ and performing simplifications we obtain the claim.

For $R \geq 0.305$ the minimum in the definition of $\delta_{\rm LP}(R)$ is given by some $\alpha < 1/2$. Fixing α equal to this value we observe that the function μ depends only on ω. Therefore, it is possible to check numerically (for instance, using Mathematica) that μ increases on ω. Substituting $\omega = \delta_{\rm LP}(R)$ together with the value of $q(\alpha,\tau,\omega)$ into μ we again arrive at the bound (5).

It remains to show that for low code rates the maximum in N is achieved by the first of the two terms. 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. The example of $p = 0.01$ is shown in Fig. 1. □

3 A New Bound

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 code word x_i and an arbitrary word y. Throughout $w = \omega n$, $l = \lambda n$ and $d = \delta n$. Let $\alpha, \tau, G(\alpha,\tau)$ have the same meaning as in (3).

Theorem 5.

$$E(R,p) \leq \min_{0\leq\alpha\leq 1/2} \max_{0\leq\lambda\leq G(\alpha,\tau)} \max_{\lambda\leq\omega\leq G(\alpha,\tau)} \left[\max(-\mu(R,\alpha,\omega) - A(\omega), B(\omega,\lambda) - A(\lambda))\right] \quad (7)$$

where A and B are defined as in Equations (1) *and* (4) *respectively.*

Performing an analysis similar to that of the previous section we obtain

Theorem 6. *For some R_0^*, a function of p, we have*

$$E(R,p) \leq \begin{cases} -A(\delta_{\mathrm{LP}}(R)) - R + 1 - h(\delta_{\mathrm{LP}}(R)) & 0 \leq R \leq R_0^* \\ \max\limits_{0\leq\lambda\leq\delta_{\mathrm{LP}}(R)} \max\limits_{\lambda\leq\omega\leq\delta_{\mathrm{LP}}(R)} B(\omega,\lambda) - A(\lambda) & R_0^* \leq R \end{cases} \quad (8)$$

where A and B are defined as in Equations (1) *and* (4) *respectively.*

Example. To show that (7) improves over (3), let $p = 0.01$. Then from (6) we obtain $R_0 \approx 0.271$. From (7) we find that the bound (5) is valid for $R \leq R_0^* \approx 0.388$. See Figure 1 for a graph of the known error bounds including our new bounds.

Remark. Experience leads us to believe that the maximums in the equation are achieved for $\omega = \lambda = \delta_{\mathrm{LP}}(R)$ which would give us the bound

$$E(R,p) \leq \begin{cases} -A(\delta_{\mathrm{LP}}(R)) - R + 1 - h(\delta_{\mathrm{LP}}(R)) & 0 \leq R \leq R_0^* \\ B(\delta_{\mathrm{LP}}(R), \delta_{\mathrm{LP}}(R)) - A(\delta_{\mathrm{LP}}(R)) & R_0^* \leq R \end{cases}$$

However this has proved too difficult to verify analytically due to the awkward nature of the η maximization term in the definition of $B(\omega,\lambda)$.

The basic idea of the estimation method is from [4] 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

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

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

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

One of the main questions in applying this inequality and further ideas of [4] is the choice of the sets Y_{ij}. We construct the Y_{ij}'s via sets $X_{ij} \subset \mathbb{F}_2^n$, where

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

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 $T_i = \{x_j :$

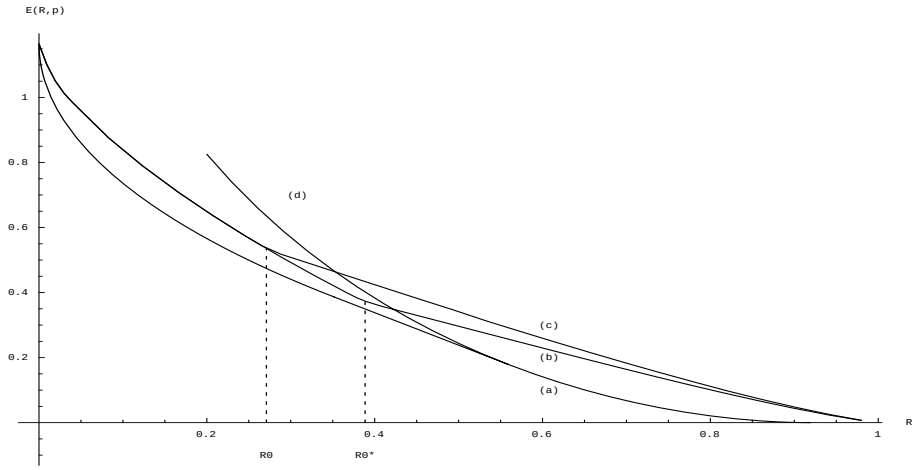

Figure 1: Bounds on the Error Exponent for the Binary Symmetric Channel for $p = 0.01$. Curve (a) is a combination of the best lower bounds on the error exponent. Curve (b) is the new upper bound given by Theorem 5. Curve (c) is the upper bound given by Theorem 3. Curve (d) is the sphere-packing bound $E_{\rm sp}(R,p)$. Note that $E_{\rm sp}(R,p)$ is better than (b) from $R \approx 0.422$; the straight-line bound (not shown) further improves the results.

$d_{ij} = w\}$ for each codeword x_i. We then arbitrarily index all pairs (x_i, x_j) with $d_{ij} = w$ by integers s_{ij}. 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 $x \in Y_{im} \cap Y_{in}$. Then $x \in X_{im}$ and $x \notin \bigcup_{k \in T(i,m)} X_{ik}$ gives that $s_{in} > s_{im}$. However we also have $x \in X_{in}$ and $x \notin \bigcup_{k \in T(i,n)} X_{ik}$ and this gives that $s_{im} > s_{in}$ which is a contradiction.

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

$$\mathbb{P}_i(Y_{ij}) \geq \mathbb{P}_i(X_{ij})\,(1 - K_{ij}),$$

where $K_{ij} = \sum_{k \in T(i,j)} \mathbb{P}_i(X_{ik}|X_{ij})$.

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

Lemma 7. *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 8. *For all codewords x_i, x_j and x_k such that $d_{ij} = d_{ik} = w$ and $d_{jk} = l$ we have $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_i(X_{ik}|X_{ij}) = B(\omega,\lambda)$ where $B(\omega,\lambda)$ is defined in Theorem 4.*

Proof. First consider

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

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 9. *If $\omega \leq \lambda \leq 2\omega$ then $B(\omega, \lambda) \leq B(\omega, \omega)$. If $\lambda \leq \omega$ then $B(\lambda, \lambda) \leq B(\omega, \lambda)$*

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

$$P_e \geq \frac{1}{M} \sum_{i=1}^{M} \sum_{j: d_{ij}=w} \mathbb{E}\left(\mathbb{P}_i\left(Y_{ij}\right)\right) \tag{9}$$

This equation will be the basis for our new bound on the error exponent but before deriving this bound we have two final preliminaries. Firstly we define B_w^i to be the "local" distance distribution, i.e., the number of neighbors of the ith codeword at distance w. Secondly we shall say that a subset of codewords is of *substantial size* if its size has the same exponential order as M. Note that for a family of codes $(C_i)_{i \geq 1}$ where C_i has length n and rate R, we can consider $(C_i')_{i \geq 1}$, a family of codes were 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

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

Theorem 10. *Consider any code C of sufficiently large length n, rate R and with $\frac{1}{n} \log B_{\omega n}^i \geq f(\omega)$ for some ω and bounding function f. Construct Y_{ij}, X_{ij} and K_{ij} as described above for all (i, j) pairs with $d_{ij} = \omega n$. If $\{x_j | K_{ij} > 1/2$ for some $i\}$ is not a substantially sized subcode then*

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

Proof. Wlog. we can assume that $K_{ij} \leq 1/2$ for all code words x_i and x_j in our code since removing codewords in $\{x_j | K_{ij} > 1/2 \text{ for some } i\}$ from our code gives a substantially sized subcode in which $K_{ij} \leq 1/2$ for all code words x_i and x_j. Hence

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

□

Theorem 11. *Consider any code C of sufficiently large length n and rate R and an $\omega \in [0,1]$. Construct Y_{ij}, X_{ij} and K_{ij} as described above for all (i,j) pairs with $d_{ij} = \omega n$. If $\{x_j | K_{ij} > 1/2$ for some $i\}$ is a substantially sized subcode then there exists a $0 \leq \lambda \leq 2\omega$ such that there is a substantial number of codewords with at least $2^{-nB(\omega,\lambda)}$ neighbors. We call λ a "nuisance level" for ω. Furthermore*

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

To prove this we need the following lemmas.

Lemma 12. *[5] 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 $n_k, 1 \leq k \leq K$. 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 0 and τ. Then*

$$\mathbb{P}(t_k \leq n_k, k = 1, \ldots K) \geq \frac{1}{4}\min_{1\leq k\leq K}\frac{n_k}{r_k}.$$

We then can prove the following lemma:

Lemma 13. *Recall that, for a given (i,j) pair, K_{ij} is a random variable. 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*

$$\mathbb{P}(K_{ij} \leq 1/2) \geq \frac{1}{8(n+1)}\min_{l\in\Lambda}\frac{2^{-nB(\omega,\lambda)}}{B_w^i}$$

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.

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

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 $n_l = N_{w,l}$ Then let $\mu_l = |x_m \in R_{w,l} : s_{im} < s_{ij}|$ Therefore

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

By the previous lemma we have

$$\mathbb{P}\left(\mu_l \leq n_l \forall l \in \Lambda\right) \geq \frac{1}{4} \min_{l \in \Lambda} \frac{n_l}{r_l}.$$

The theorem then follows from the fact that $|R_{w,l}| \leq B_w^i$. □

Proof of Theorem 11. For each codeword $x \in \{x_j | K_{ij} \geq 1/2 \text{ for some } i\}$ there exists λ from Lemma 13 such that x has at least $2^{-nB(\omega,\lambda)}$ neighbors at relative distance λ. Now since there exists a substantial number of such x and there are only n distinct values of λ there exists a λ_1 such that a substantial number of the codewords have at least $2^{-nB(\omega,\lambda_1)}$ neighbors at a relative distance λ_1 and that

$$\mathbb{P}\left(K_{ij} \leq 1/2\right) \geq \frac{1}{8(n+1)} \frac{2^{-nB(\omega,\lambda_1)}}{B_w^i}.$$

Now

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

and so from Lemma 13 and Eq. (9) we get

$$\begin{aligned}
P_e &\geq \frac{1}{M} \sum_{i=1}^{M} \sum_{j : d_{ij} = w} \mathbb{E}(\mathbb{P}_i\left(Y_{ij}\right)) \geq \frac{1}{M} \sum_{i=1}^{M} \sum_{j : d_{ij} = w} \frac{2^{nA(\omega)}}{16(n+1)} \min_{l \in \Lambda} \frac{2^{-nB(\omega,\lambda)}}{B_w^i} \\
&\geq \frac{1}{M} \sum_{i=1}^{M} \frac{2^{nA(\omega)}}{16(n+1)} 2^{-nB(\omega,\lambda)} = 2^{n(A(\omega) - B(\omega,\lambda)) + o(n)}
\end{aligned}$$

□

Proof of Theorem 5. Pick any code C of rate R and sufficiently large length n. By Theorem 2 there exists an $\omega \in [0, G(\alpha,\tau)]$ such that $\frac{1}{n} \log B_{\omega n} \geq \mu(\omega)$. As discussed in [2], [4], the code C contains a subcode C' of size $M' \geq M/n^2$ such that for all codewords x_i in this subcode

$$\frac{1}{n} \log B_{\omega n}^i > \mu(\omega),$$

where $\mu(\omega) = \mu(R, \alpha, \omega)$ is the same as in Theorem 2 . Since the subcode is substantially sized we may now consider this subcode as our new code.

For this choice of ω construct Y_{ij}, X_{ij} and K_{ij} for all (i,j) pairs with $d_{ij} = \omega n$. Hence by Theorems 10 and 11 we get

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \begin{cases} -\mu(\omega) - A(\omega) & \text{if no nuisance level exists for } \omega \\ B(\omega,\lambda_1) - A(\omega) & \text{if a nuisance level } \lambda_1 \text{ exists for } \omega \end{cases}$$

Hence we get

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \max\{-\mu(\omega), B(\omega,\lambda_1)\} - A(\omega).$$

Now if $\lambda_1 \geq \omega$ then $B(\omega,\lambda_1) \leq B(\omega,\omega)$ and so we get

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \max\{-\mu(\omega), B(\omega,\omega)\} - A(\omega). \tag{10}$$

If $\lambda_1 < \omega$ then we use the fact from Theorem 11 that for a substantial number of codewords x_i, $B^i_{\lambda_1 n} \geq 2^{-nB(\omega,\lambda_1)}$. 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 10 and 11 we get

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \begin{cases} B(\omega,\lambda_1) - A(\lambda_1) & \text{if no nuisance level exists for } \lambda_1 \\ B(\lambda_1,\lambda_2) - A(\lambda_1) & \text{if a nuisance level } \lambda_2 \text{ exists for } \lambda_1 \end{cases}$$

Hence we get

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \max\{B(\omega,\lambda_1), B(\lambda_1,\lambda_2)\} - A(\lambda_1).$$

If $\lambda_2 \geq \lambda_1$ then $B(\lambda_1,\lambda_2) \leq B(\lambda_1,\lambda_1) \leq B(\omega,\lambda_1)$ then

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq B(\omega,\lambda_1) - A(\lambda_1).$$

If $\lambda_2 < \lambda_1$ then we use the fact that for a substantial number of codewords x_i, $B^i_{\lambda_2 n} \geq 2^{-nB(\lambda_1,\lambda_2)}$ 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

$$\frac{1}{n}\log\frac{1}{P_e(C,p)} \leq \max\{B(\lambda_{i-1},\lambda_i), B(\lambda_i,\lambda_{i+1})\} - A(\lambda_i).$$

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

$$\begin{aligned} \frac{1}{n}\log\frac{1}{P_e(C,p)} &\leq \max\{B(\lambda_{f-1},\lambda_f), B(\lambda_f,\lambda_{f+1})\} - A(\lambda_f) \\ &\leq \max\{B(\lambda_{f-1},\lambda_f), B(\lambda_f,\lambda_f)\} - A(\lambda_f) \\ &\leq B(\omega,\lambda_f) - A(\lambda_f). \end{aligned}$$

Now for our code either this equation or Eq. (10) is valid and so, since our choice of C was arbitrary among codes of rate R, we have shown that there exists ω and $\lambda \leq \omega$ such that

$$\min_{C \subseteq \{0,1\}^n, R(C)=R} \frac{1}{n} \log \frac{1}{P_e(C,p)} \leq \max(-\mu(\omega) - A(\omega), B(\omega,\lambda) - A(\lambda))$$

Therefore there exists $\omega \in G(\alpha,\tau)$ and $\lambda \leq \omega$ such that

$$E(R,p) \leq \max(-\mu(\omega) - A(\omega), B(\omega,\lambda) - A(\lambda))$$

□

References

[1] A. Ashikhmin and A. Barg, *Binomial moments of the distance distribution: Bounds and applications*, IEEE Trans. Inform. Theory **45** (1999), no. 2, 438–452.

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

[3] M. V. Burnashev, *A new lower bound for the α-mean error of parameter transmission over the white Gaussian channel*, IEEE Trans. Inform. Theory **30** (1984), no. 1, 23–34.

[4] ———, *On the relation between the code spectrum and the decoding error probability*, Problemy Peredachi Informatsii **36** (2000), no. 4, 3–24.

[5] M. V. Burnashev and Y. A. Kutoyants, *On Minimal α-Mean Error Parameter Transmission Over a Poisson Channel*, IEEE Trans. Inform. Theory **47** (2001), no. 6, 2505–2515.

[6] R. G. Gallager, *Low-density parity-check codes*, MIT Press, Cambridge, MA, 1963.

[7] S. Litsyn, *New upper bounds on error exponents*, IEEE Trans. Inform. Theory **45** (1999), no. 2, 385–398.

[8] 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. Inform. Theory **23** (1977), no. 5, 611–613.

[9] 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. Inform. Theory **23** (1977), no. 2, 157–166.

[10] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, *Lower bounds to error probability for codes on discrete memoryless channels, II*, Information and Control **10** (1967), 522–552.