Lower Bounds on the Error Probability of Block Codes Based on Improvements on de Caen's Inequality*

Asaf Cohen† Neri Merhav†

soofsoof@tx.technion.ac.il merhav@ee.technion.ac.il

Abstract

New lower bounds on the error probability of block codes with maximum likelihood decoding are proposed. The bounds are obtained by applying a new lower bound on the probability of a union of events, derived by improving on de Caen's lower bound. The new bound includes an arbitrary function to be optimized in order to achieve the tightest results. Since the optimal choice of this function is known, but leads to a trivial and useless identity, we find several useful approximations for it, each resulting in a new lower bound.

For the Additive White Gaussian Noise (AWGN) channel and the Binary Symmetric Channel (BSC), the optimal choice of the optimization function is stated and several approximations are proposed. When the bounds are further specialized to linear codes, the only knowledge on the code used is its weight enumeration. The results are shown to be tighter than the latest bounds in the current literature, such as those by Seguin and by Keren and Litsyn. Moreover, for the BSC, the new bounds widen the range of rates for which the union bound analysis applies, thus improving on the bound to the error exponent compared the de Caen-based bounds.

Index terms - Probability of error, maximum likelihood decoding, probability of a union, lower bound, Gaussian channel, binary symmetric channel, union bound analysis, error exponent.

1 Introduction

Consider the classical coded communication model of transmitting one of M equally likely signals over a communication channel. The error probability of the optimal maximum likelihood decoder is often complicated to evaluate. Thus, to estimate the performance of a given signal set, lower and upper bound on the decoding error probability are required.

Numerous bounds on the error probability of maximum likelihood decoding, based on a wide variety of techniques, can be found in the current literature. We briefly review the ones most related to this paper. Although we mainly refer to the AWGN channel and the BSC, most bounds are applicable

*The material in this paper was presented in part at the 22nd Convention of Electrical and Electronics Engineers in Israel, Tel Aviv, Israel, December 2002, and submitted, in part, to the 2003 IEEE International Symposium on Information Theory, Yokohama, Japan, June-July 2003.

†The authors are with the Faculty of the Electrical Engineering, Technion - I.I.T., Haifa 32000, Israel.

to a wider range of channel models. The most common upper bound on the error probability is the well-known *union bound.* This bound is tight, hence widely used, at low levels of noise. At high levels of noise, the union bound is loose and tighter bounds are required. The best known upper bound for finite block length and high levels of noise is due to Poltyrev [1]. Let ε be a decoding error event and $\mathcal{A}$ be an arbitrary subset of the possible channel outputs (or, equivalently, noise vectors), Poltyrev's bound is based on the following inequality

$$P(\varepsilon) \leq P(\varepsilon, \mathcal{A}) + P(\mathcal{A}^c), \tag{1}$$

where $\mathcal{A}^c$ is the complement of $\mathcal{A}$. For example, to bound the error probability over the BSC, Poltyrev chose the set $\mathcal{A}$ to be the set of all binary words of weight higher than some threshold m, which was later optimized to yield the tightest bound. The value of $P(\varepsilon, \mathcal{A})$ was bounded by the union bound. For the AWGN channel, Poltyrev elaborated on previous techniques used by Hughes [2] and Berlekamp [3]. In this case, $\mathcal{A}$ was chosen to be a circular cone of half-angle θ, whose central line passes through the origin and the transmitted codeword.

For asymptotically infinite block length, define the *error exponent* (also know as the reliability function) of a channel, $E(R)$, as

$$E(R) = \limsup_{N\to\infty} \frac{-\log P_{N,R}(\varepsilon)}{N},$$

where $P_{N,R}(\varepsilon)$ is the minimum value of $P(\varepsilon)$ over all codes of a given block length N and rate R. Hereafter, the base of the logarithm is 2. Although Poltyrev's bound is tight for finite block length, the best known lower bound on the error exponent (upper bound on the error probability) is due to Gallager. In [4], Gallager derived a lower bound on the error exponent by methods of random coding. For low rates, since the average low-rate code is bad, the bound was tightened by methods of expurgation.

As for lower bounds on the error probability, the best known bound for high levels of noise is the *sphere packing* bound [5]. Roughly speaking, the sphere packing bound states that the probability of error is greater than that of a perfect code [6]. For the AWGN channel, for example, the sphere packing bound is derived by noting that the error probability of any code whose codewords lie on a given sphere must be greater than the error probability of a code with the same rate and whose codewords are uniformly distributed over that sphere. Asymptotically in N, the sphere packing bound coincides with the random coding bound [4] for rates higher than a certain critical rate, R_c, thus yielding the exact error exponent for these rates. For rates lower than R_c, several bounds were offered. The *two codewords* bound considers the error caused by a pair of closest codewords. Together with an upper bounds on the minimum distance of a code, such as the bound derived by McEliece, Rodemich, Rumsey and Welch for binary codes [7, pp. 559], bounds tighter than the sphere packing bound can be derived for low rates [8]. For intermediate rates, the *straight line* bound by Shannon, Gallager and Berlekamp [9], connecting any low rate bound to the sphere packing bound, can be used. The latest upper bound on the error exponent of the BSC was derived by Litsyn [10]. The essence of his technique is in a new bound on the distance distribution of codes, and not an improvement of the McEliece-Rodemich-Rumsey-Welch bound, as might have been expected. The latest upper bound on the error exponent of the AWGN was derived by Burnashev [11]. Burnashev showed that by extending the range in which the union bound analysis applies, together with a bound on the distance distribution of codes, the bound on the error exponent can be tightened. We note here that for random codes, random linear codes and *typical* codes from these ensembles the error exponent is known exactly [12].

For finite block length and low levels of noise, the currently best known lower bounds on the error probability are due to Seguin [13] (AWGN) and Keren and Litsyn [14][1] (BSC). While the preceding bounds discussed herein mainly use geometrical arguments in order to evaluate the error probability, Seguin's and Keren and Litsyn's bounds use an alternative approach. The idea is analogous to the union bound technique: to view the probability of error as a probability of a union of events, and use known bound on this probability. When this method is used, the basic events, whose probabilities are to be evaluated directly, are usually the error events when only two or three codewords are involved, hence their evaluation is simple. As a bound on the probability of a union, both Seguin and Keren and Litsyn used a recent lower bound by de Caen [15]. Their techniques will be discussed later in this paper. In [16], Kuai, Alajaji and Takahara drive upper and lower bounds using the same method. Their work includes a bound by the same authors on the probability of a union [17], together with simple algorithms for Kounias' bound [18] and Hunter's bound [19]. However, Kuai, Alajaji and Takahara consider uncoded communication and nonuniform signaling.

In this paper, we derive a new bound on the probability of a union and apply it to derive lower bounds on the error probability of block codes. In Section 2, the new bound on the probability of a union is proposed. In Section 3, this bound is applied to lower bound the error probability over the AWGN channel. The resulting bounds are specialized for Binary Phase Shift Keying (BPSK) modulation of linear codes. In this case, the only knowledge on the code required is its weight enumeration. Numerical analysis results show significant enhancement in performance compared to known bounds in the literature. To the authors' knowledge, for medium and high values of the Signal to Noise Ratio (SNR) the bounds are shown to yield the tightest results currently available. Section 3 also includes a derivation of a new bound based on Kounias' lower bound on the probability of a union. The resulting bound is very simple and performs well for every SNR (superior to Seguin's bound). Section 4 includes analogous derivations for the BSC. Numerical analysis shows enhancement in performance compared to Keren and Litsyn's bound, though in this case the improvement is milder. However, in Section 5, the bounds on the error exponent resulting from the bounds in the preceding sections are discussed. It is shown that the new bounds may be exponentially tighter than the de Caen-based bounds. Section 7 includes a short discussion and suggestions for future work.

2 A New Lower Bound on the Probability of a Union of Events

In this section, we derive a new lower bound on the probability of a union of events. We mainly follow the method used by de Caen in [15], however, the new bound includes a function that can be optimized to yield tighter bounds. This bound will stand at the basis of our analysis tools.

[1]A more detailed paper (unpublished) is also available [6].

Let $\{A_i\}_{i\in\mathcal{I}}$ be any finite set of events in a probability space $(\Omega, \mathcal{F}, P)$. For each $x \in \Omega$, define

$$deg(x) \stackrel{\triangle}{=} |\{i \in \mathcal{I} : x \in A_i\}|.$$

The new lower bound is given by the following theorem.

Theorem 2.1 *Let $\{A_i\}_{i\in\mathcal{I}}$ be any finite set of events in a probability space $(\Omega, \mathcal{F}, P)$. The probability of the union $P(\cup_{i\in\mathcal{I}} A_i)$ is lower bounded by*

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) \geq \sum_{i\in\mathcal{I}} \frac{\left(\sum_{x\in A_i} p(x)m_i(x)\right)^2}{\sum_{j\in\mathcal{I}} \sum_{x\in A_i\cap A_j} p(x)m_i^2(x)}, \tag{2}$$

where $m_i(x) \geq 0$ is any real function on Ω such that the sums on the right hand side (r.h.s.) of (2) converge. Equality in (2) is achieved when

$$m_i(x) = m^*(x) = \frac{1}{deg(x)}, \quad \forall_{i\in\mathcal{I}}. \tag{3}$$

Proof . We first consider the case where Ω is finite. Using a simple counting argument, we have

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) = \sum_{i\in\mathcal{I}} \sum_{x\in A_i} \frac{p(x)}{deg(x)}. \tag{4}$$

Let $m_i(x) \geq 0$ be any real function on Ω. From the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\left(\sum_{x\in A_i} \frac{p(x)}{deg(x)}\right)\left(\sum_{x\in A_i} p(x)m_i^2(x)deg(x)\right) &\geq \left(\sum_{x\in A_i} \sqrt{\frac{p(x)}{deg(x)}}\sqrt{p(x)m_i^2(x)deg(x)}\right)^2 \\ &= \left(\sum_{x\in A_i} p(x)m_i(x)\right)^2,\end{aligned} \tag{5}$$

provided that the sums in (5) converge. Therefore, from (4) and (5),

$$\begin{aligned}P\left(\bigcup_{i\in\mathcal{I}} A_i\right) &\geq \sum_{i\in\mathcal{I}} \frac{\left(\sum_{x\in A_i} p(x)m_i(x)\right)^2}{\sum_{x\in A_i} p(x)m_i^2(x)deg(x)} \\ &= \sum_{i\in\mathcal{I}} \frac{\left(\sum_{x\in A_i} p(x)m_i(x)\right)^2}{\sum_{j\in\mathcal{I}} \sum_{x\in A_i\cap A_j} p(x)m_i^2(x)}.\end{aligned}$$

Note that $m_i(x)$ may be different for each i in the sum over all $i \in \mathcal{I}$. However, in order to achieve equality in (5) we should define

$$m_i(x) = \frac{1}{deg(x)}, \quad \forall_{i\in\mathcal{I}}.$$

For a general probability space, as noted in [15] and [17], since there are only finitely many A_i's, the number of Boolean atoms defined by the A_i's unions and intersections is also finite. Thus, the general space can be reduced to a finite probability space. In this case, the sums in (2) are replaced by the corresponding Lebesgue integrals. □

We shall refer to the choice of $m_i(x) \equiv 1$ as the *trivial* choice of $m_i(x)$. By choosing the trivial choice for $m_i(x)$, we have

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) \geq \sum_{i\in\mathcal{I}} \frac{P(A_i)^2}{\sum_{j\in\mathcal{I}} P(A_i\cap A_j)},$$

which is de Caen's bound [15]. Thus, de Caen's bound is a special case of the bound suggested in Theorem 2.1. In this context, note that a recent improvement of de Caen's bound was given in [17], by Kuai, Alajaji and Takahara. However, one can show ([20]) that both de Caen's bound and Kuai, Alajaji and Takahara's bound are derived by solving the same minimization problem. While the latter is obtained by applying a stronger method than the first, it improves on de Caen's bound by at most 9/8.

The essence of the bound in Theorem 2.1 is the ability to choose an appropriate function $m_i(x)$. To define a proper strategy for choosing $m_i(x)$, first note that any constant multiplier of $m_i(x)$ factors out in (2). Hence, $m_i(x)$ should only define an *essence of behavior*, and not necessarily exact values. When seeking such a behavior, we remember that the optimal value of $m_i(x)$ is $1/deg(x)$. While the function $deg(x)$ is complex to evaluate, usually requires more than the available information on the sets $\{A_i\}_{i\in\mathcal{I}}$, and leads to a trivial identity in (2), its behavior possesses the guidelines for choosing a competent family of approximations. By requiring that any such family of approximations includes the trivial choice of $m_i(x)$, and optimizing the bound over this family, one can assure that the resulting bound is always at least as tight as de Caen's bound.

It is clear that the bound given in Theorem 2.1 does not depend only on the $P(A_i)$'s and $P(A_i \cap A_j)$'s. However, a proper choice of the function $m_i(x)$ may still yield the same computational tractability, while improving on de Caen's bound, achieved by choosing $m_i(x) \equiv 1$. When the computational tractability is of less importance, $m_i(x)$ may be chosen to be constant on *subsets* of the A_i's, yielding more accurate results. Thus, the chosen family of approximations should also reflect the tradeoff between tractability and tightness of the bound.

3 The Additive White Gaussian Noise Channel

We consider the case of uniform signaling over an AWGN channel and maximum likelihood decoding. The transmitted signal is one of M equiprobable signals $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}$ of length K. If $\mathbf{s}_0$ is transmitted, the received signal is $\mathbf{r} = \mathbf{s}_0 + \mathbf{n}$, where $\mathbf{n}$ is a vector of K independent Gaussian random variables with zero mean and variance $\frac{N_0}{2}$ (i.e., the double sided spectral density of the noise is $\frac{N_0}{2}$). The maximum likelihood decoder chooses the closest of the M signals to $\mathbf{r}$, in the Euclidean sense. Thus, the probability of error given that $\mathbf{s}_0$ was sent is

$$P(\varepsilon|\mathbf{s}_0) = P\left(\cup_{i\neq 0}\varepsilon_{0i}|\mathbf{s}_0\right),$$

where

$$\varepsilon_{0i} \stackrel{\triangle}{=} \{\mathbf{r} \in \mathbb{R}^K : ||\mathbf{r} - \mathbf{s}_i|| < ||\mathbf{r} - \mathbf{s}_0||\} \tag{6}$$

and $||\cdot||$ is the Euclidean norm.[2]

In order to use the bound in Theorem 2.1, choose $\mathcal{I} = \{1, \ldots, M-1\}$ and $A_i = \varepsilon_{0i}$. Referring to

[2]Note the strict inequality in (6), a consequence of the assumption that ties are solved in favor of the correct signal. Generally speaking, this assumption is essential when lower bounds on the error probability are discussed. When a continuous probability space is at hand it is of less importance.

(2), the computation of the bound requires the evaluation of the following integrals:

$$\int_{\varepsilon_{0i}} p(\mathbf{r}|\mathbf{s}_0)m(\mathbf{r}|\mathbf{s}_0)d\mathbf{r}, \tag{7}$$

$$\int_{\varepsilon_{0i}\cap\varepsilon_{0j}} p(\mathbf{r}|\mathbf{s}_0)m^2(\mathbf{r}|\mathbf{s}_0)d\mathbf{r}, \tag{8}$$

where

$$p(\mathbf{r}|\mathbf{s}_0) = (\pi N_0)^{-\frac{K}{2}} \exp\left\{-\frac{1}{N_0}\|\mathbf{r}-\mathbf{s}_0\|^2\right\}. \tag{9}$$

Note that since $m_i(\cdot)$ was any function to be optimized, we may choose it to be independent of i, as the optimal value given in (3) suggests. For the trivial choice of $m(\mathbf{r}|\mathbf{s}_0)$, i.e., $m(\mathbf{r}|\mathbf{s}_0) \equiv 1$, the integrals in (7) and (8) are simply the pairwise error probability and triplets error probability. In this case, the resulting bound is the bound given by Seguin in [13, eq. (13)] for any signal set.

However, the essence of the new bound is the ability to choose a proper, non trivial, $m(\mathbf{r}|\mathbf{s}_0)$. The optimal value of $m(\mathbf{r}|\mathbf{s}_0)$ was given in Theorem 2.1,

$$m^*(\mathbf{r}|\mathbf{s}_0) = \frac{1}{deg(\mathbf{r}|\mathbf{s}_0)},$$

where $deg(\mathbf{r}|\mathbf{s}_0)$ is the number of signals which are closer to $\mathbf{r}$ than $\mathbf{s}_0$, i.e.,

$$deg(\mathbf{r}|\mathbf{s}_0) = |\{i : \|\mathbf{r}-\mathbf{s}_i\| < \|\mathbf{r}-\mathbf{s}_0\|\}|. \tag{10}$$

The evaluation of $deg(\mathbf{r}|\mathbf{s}_0)$ is usually very complex and, in many practical cases, infeasible when only the distance spectrum of the code is to be used. Moreover, $m(\mathbf{r}|\mathbf{s}_0)$ should be mathematically endurable so the integrals in (7) and (8) can be easily computed. Nevertheless, we will see that suitable approximations can be found.

A first approximation is derived directly from (10). Since $deg(\mathbf{r}|\mathbf{s}_0)$ is the number of signals in the interior of a sphere of radius $\|\mathbf{r}-\mathbf{s}_0\|$ centered at $\mathbf{r}$, one might suggest that the larger the volume of the sphere, the higher $deg(\mathbf{r}|\mathbf{s}_0)$ is. Namely, $deg(\mathbf{r}|\mathbf{s}_0)$ is monotonically increasing in $\|\mathbf{r}-\mathbf{s}_0\|$. Thus, $m(\mathbf{r}|\mathbf{s}_0)$ might be chosen as

$$m(\mathbf{r}|\mathbf{s}_0) = \exp\left\{-a\|\mathbf{s}_0-\mathbf{r}\|^2\right\}, \tag{11}$$

where $a \geq 0$ is a parameter to be optimized in order to achieve the tightest bound. An exponential behavior was chosen to facilitate the computation of (7) and (8). A drawback of this approximation however, is that it is implicitly based on the infeasible assumption that the signals are uniformly distributed in $\mathbb{R}^K$. Nevertheless, this choice does improve on the trivial choice of $m(\mathbf{r}|\mathbf{s}_0)$, corresponding to $a = 0$, as we will see in Section 6.

Fortunately, for equal-energy signals, a more realistic approximation can be derived in a similar fashion. Since for all i, $\|\mathbf{s}_i\| = \|\mathbf{s}_0\| = \sqrt{E}$, we have

$$deg(\mathbf{r}|\mathbf{s}_0) = |\{i : \langle\mathbf{s}_i,\mathbf{r}\rangle > \langle\mathbf{s}_0,\mathbf{r}\rangle\}| = |\{i : \theta_{\mathbf{r}i} < \theta_{\mathbf{r}0}\}|. \tag{12}$$

where $\langle\cdot,\cdot\rangle$ is the standard dot product and

$$\theta_{\mathbf{r}i} \stackrel{\triangle}{=} \cos^{-1}\left\{\frac{\langle\mathbf{s}_i,\mathbf{r}\rangle}{\|\mathbf{s}_i\|\|\mathbf{r}\|}\right\}, \qquad 0 \leq \theta_{\mathbf{r}i} < \pi.$$

Assuming the signals are uniformly distributed on the *surface* of a sphere of radius $\sqrt{E}$ centered at the origin, equation (12) implies that $deg(\mathbf{r}|\mathbf{s}_0)$ is monotonically increasing with respect to the *absolute value of the angle* between $\mathbf{r}$ and $\mathbf{s}_0$. Thus, $m(\mathbf{r}|\mathbf{s}_0)$ might be chosen as

$$m(\mathbf{r}|\mathbf{s}_0) = \exp\{a\langle \mathbf{s}_0, \mathbf{r}\rangle\}, \tag{13}$$

where, again, $a \geq 0$ is a parameter to be optimized. Clearly, when BPSK modulation of a binary code is used, which is the case drawing our main attention, the signals are of equal energy. However, as noted in [5], equal-energy signals are worth considering anyhow. The assumption that the signals are uniformly distributed on the surface of the sphere cannot, of course, be justified in general. However, it is important to note that since this assumption is at the basis of the sphere packing bound [5, Section 3], which is asymptotically tight for rates higher than R_c, we know that good codes of high rate do have approximately uniform distribution of codewords on the surface of the sphere.

Both suggestions for $m(\mathbf{r}|\mathbf{s}_0)$, defined in (11) and (13), are members of a wider family, characterized by three parameters, a, b and c,

$$m(\mathbf{r}|\mathbf{s}_0) = \exp\left\{-\left(a\|\mathbf{r}\|^2 + b\langle \mathbf{r}, \mathbf{s}_0\rangle + c\|\mathbf{s}_0\|^2\right)\right\}. \tag{14}$$

Although more suggestions for $m(\mathbf{r}|\mathbf{s}_0)$ can be given, we choose to focus on (14). The following proposition introduces the new bound on the error probability for any signal set, using this suggestion. The simpler suggestions discussed earlier easily follow.

Proposition 3.1 *Let $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{M-1}$ be a set of M signals of dimension K for an AWGN channel with spectral density $\frac{N_0}{2}$. The conditional probability of error of a maximum likelihood decoder is lower bounded by*

$$P(\varepsilon|\mathbf{s}_0) \geq e^{(\beta'-2\beta)\|\mathbf{s}_0\|^2}\left(\frac{N_0'}{\sqrt{N_0 N_0''}}\right)^K \sum_{i=1}^{M-1} \frac{Q^2\left(\kappa(\alpha, \mathbf{s}_i, N_0')\right)}{\sum_{j=1}^{M-1} \Psi\left(\rho_{ij}, \kappa(\alpha', \mathbf{s}_i, N_0''), \kappa(\alpha', \mathbf{s}_j, N_0'')\right)}, \tag{15}$$

where

$$\begin{aligned}
Q(x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy, &(16)\\
\Psi(\rho, x', y') &= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{x'}^\infty\int_{y'}^\infty \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}dxdy, &(17)\\
\rho_{ij} &= \frac{\langle \mathbf{s}_i - \mathbf{s}_0, \mathbf{s}_j - \mathbf{s}_0\rangle}{\|\mathbf{s}_i - \mathbf{s}_0\|\|\mathbf{s}_j - \mathbf{s}_0\|}, &(18)\\
\kappa(\alpha, \mathbf{s}_i, N_0) &= \frac{\|\alpha\mathbf{s}_0 - \mathbf{s}_i\|^2 - (\alpha-1)^2\|\mathbf{s}_0\|^2}{\sqrt{2N_0}\|\mathbf{s}_0 - \mathbf{s}_i\|}, &(19)
\end{aligned}$$

with the understanding that $\Psi(1, x, x) = Q(x)$. The constants N_0', N_0'', α, α', β and β' are given by

$$N_0' = \frac{N_0}{1 + aN_0}, \quad N_0'' = \frac{N_0}{1 + 2aN_0}, \quad \alpha = \left(\frac{\frac{1}{N_0} - \frac{b}{2}}{a + \frac{1}{N_0}}\right), \quad \alpha' = \left(\frac{\frac{1}{N_0} - b}{2a + \frac{1}{N_0}}\right)$$

$$\beta = \frac{\left(\frac{1}{N_0} + a\right)\left(\frac{1}{N_0} + c\right) - \left(\frac{1}{N_0} - \frac{b}{2}\right)^2}{\frac{1}{N_0} + a}, \quad \beta' = \frac{\left(\frac{1}{N_0} + 2a\right)\left(\frac{1}{N_0} + 2c\right) - \left(\frac{1}{N_0} - b\right)^2}{\frac{1}{N_0} + 2a}, \tag{20}$$

and $a > -\frac{1}{2N_0}$, b, and c are arbitrary constants.

Proof . We apply the lower bound on the probability of a union given in (2), using equations (7), (8), (9) and (14). Equation (15) easily follows after computing the integrals (7) and (8). Since the computation of these integrals is rather cumbersome, it is included in Appendix A. □

Clearly, choosing $a = b = c = 0$ results in Seguin's bound [13, eq. (13)]. Hence, for the optimal choice of a, b and c the bound in (15) is at least as tight as Seguin's. In this context, it is clear that the asymptotic tightness of Seguin's bound as $E_b/N_0 \to \infty$ remains intact. To restrict ourselves to simpler bounds, when only one parameter can be optimized, we may choose $a = c = a'$, $b = -2a'$, which results in the *norm* bound (i.e., using (11)), or $a = c = 0$, $b = -a'$, which results in the *dot product* bound (i.e., using (13)).

3.1 New Lower Bounds for Linear Codes

The bound given in Proposition 3.1 requires two nested summations over the entire signal set. Thus, it is of very little use for large codes. Analogously to Seguin's derivations, we specialize this bound for linear codes and BPSK modulation. In this case, the resulting bound depends on the code only through its weight enumeration, and is, thus, much easier to evaluate.

Assume that a binary (N, K) linear code $\mathcal{C}$ is used. $\mathbf{s}_i$ is obtained by replacing the zeroes and ones in $\mathbf{c}_i$ with $\sqrt{E_N}$ and $-\sqrt{E_N}$ respectively[3] (BPSK modulation). The energy per bit in this case is $E_b = \frac{NE_N}{K}$. Denote by $w(\mathbf{c})$ the Hamming weight of the codeword $\mathbf{c}$ and by $\mathcal{B} = \{B_0, B_1, \ldots, B_N\}$ the weight enumeration of the code, i.e., B_i is the number of codewords of Hamming weight i. Assuming $\mathbf{c}_0$ is the all-zero codeword, we have

$$\begin{aligned} \|\mathbf{s}_0\|^2 &= NE_N, \\ \|\alpha\mathbf{s}_0 - \mathbf{s}_i\|^2 &= (\alpha-1)^2 NE_N + 4\alpha E_N w(\mathbf{c}_i). \end{aligned}$$

Hence,

$$Q\left(\kappa(\alpha, \mathbf{s}_i, N_0')\right) = Q\left(\sqrt{\frac{\alpha^2 E_N w(\mathbf{c}_i)}{N_0'/2}}\right) \tag{21}$$

and

$$\Psi\left(\rho_{ij}, \kappa(\alpha', \mathbf{s}_i, N_0''), \kappa(\alpha', \mathbf{s}_j, N_0'')\right) = \Psi\left(\rho_{ij}, \sqrt{\frac{\alpha'^2 E_N w(\mathbf{c}_i)}{N_0''/2}}, \sqrt{\frac{\alpha'^2 E_N w(\mathbf{c}_j)}{N_0''/2}}\right), \tag{22}$$

where

$$\rho_{ij} = \frac{w(\mathbf{c}_i\mathbf{c}_j)}{\sqrt{w(\mathbf{c}_i)w(\mathbf{c}_j)}}.$$

The expressions in (21) and (22) can be substituted into (15) directly. However, ρ_{ij} does not depend solely on the code's weight enumeration. In [13], Seguin proved that $\Psi(\rho, x, y)$ is monotonically increasing in ρ. Thus, to derive a bound which depends only on the weight enumeration of the code,

[3]Note that the signals in this case are of length N and dimension K. Signals of length K are achieved by the Gram-Schmidt procedure and the projection of the signals on the new base. The computation of the signals' energy and distance spectrum is, however, clearer when the original signals are treated.

ρ_{ij} should be upper bounded in terms of the weight enumeration alone. Denote by d the minimum distance of the code, for $\mathbf{s}_i \neq \mathbf{s}_j$, an upper bound on ρ_{ij}, derived in [13], is given by

$$\rho_{ij} \leq \varrho(i,j) \stackrel{\triangle}{=} \min\left\{ \sqrt{\frac{w(\mathbf{c}_i)}{w(\mathbf{c}_j)}}, \sqrt{\frac{w(\mathbf{c}_j)}{w(\mathbf{c}_i)}}, \frac{w(\mathbf{c}_i)+w(\mathbf{c}_j)-d}{2\sqrt{w(\mathbf{c}_i)w(\mathbf{c}_j)}} \right\}. \tag{23}$$

All this said, substitute (23) into (22), and the result, together with (21), into (15). Since the resulting summands depend on the code only through the weight enumeration, the summation can be carried out on the possible codewords weights. Finally, when linear codes are used on a binary-input output-symmetric channel with maximum likelihood decoding, the probability of error is independent of the codeword sent (see [21, pp. 86]). Hence, we assume the all-zero codeword was sent, $P(\varepsilon|\mathbf{s}_0) = P(\varepsilon)$, and we have

$$P(\varepsilon) \geq \exp\{(\beta' - 2\beta)NE_N\}\left(\frac{N_0'}{\sqrt{N_0 N_0''}}\right)^K$$
$$\cdot \sum_{i\neq 0} \frac{B_i Q^2\left(\sqrt{\frac{\alpha^2 E_N i}{N_0'/2}}\right)}{Q\left(\sqrt{\frac{\alpha'^2 E_N i}{N_0''/2}}\right) + (B_i - 1)\Psi\left(\varrho_{ii}, \sqrt{\frac{\alpha'^2 E_N i}{N_0''/2}}, \sqrt{\frac{\alpha'^2 E_N i}{N_0''/2}}\right) + \sum_{j\neq 0,i} B_j \Psi\left(\varrho_{ij}, \sqrt{\frac{\alpha'^2 E_N i}{N_0''/2}}, \sqrt{\frac{\alpha'^2 E_N j}{N_0''/2}}\right)} \tag{24}$$

where

$$\varrho_{ij} = \min\left\{ \sqrt{\frac{i}{j}}, \sqrt{\frac{j}{i}}, \frac{i+j-d}{2\sqrt{ij}} \right\}, \tag{25}$$

$\alpha, \alpha', \beta, \beta', N_0'$ and N_0'' are as defined in (20), and $a > -\frac{1}{2N_0}$, b, and c are arbitrary constants.

3.2 Lower Bounds Depending Only on a Subset of the Code

The bound given in (24) requires the complete weight enumeration of the code. When the weight enumeration is not known completely, or when (24) is too complex to evaluate, a simpler, yet very efficient, bound can be offered. Clearly, the error probability of a given code $\mathcal{C}$, $P_{\mathcal{C}}(\varepsilon)$, satisfies $P_{\mathcal{C}}(\varepsilon) \geq P_{\mathcal{C}^*}(\varepsilon)$, where $\mathcal{C}^*$ is any subset of the code $\mathcal{C}$. Hence, any lower bound on $P_{\mathcal{C}^*}(\varepsilon)$ is a lower bound on $P_{\mathcal{C}}(\varepsilon)$. This technique is widely used when lower bounds for low rates are discussed (see, for example, [21, pp. 174]). When the code $\mathcal{C}$ is linear, $\mathcal{C}^*$ is not necessarily linear. Nevertheless, we have

$$P_{\mathcal{C}}(\varepsilon) = P_{\mathcal{C}}(\varepsilon|\mathbf{c}_0) \geq P_{\mathcal{C}^*}(\varepsilon|\mathbf{c}_0).$$

As in [14], we choose

$$\mathcal{C}^* = \mathcal{C}_d^* \stackrel{\triangle}{=} \{\mathbf{c} \in \mathcal{C} : w(\mathbf{c}) = d\} \cup \{\mathbf{c}_0\}.$$

Thus, we may substitute $B_0 = 1$ and $B_i = 0$ for every $i \neq d$ in (24). The resulting lower bound is given in the following corollary.

Corollary 3.2 *Let $\mathcal{C}$ be a binary (N, K) linear code used over the AWGN channel with BPSK modulation. The probability of error of a maximum likelihood decoder is lower bounded by*

$$P(\varepsilon) \geq \frac{\exp\{(\beta' - 2\beta)NE_N\}\left(\frac{N_0'}{\sqrt{N_0 N_0''}}\right)^K B_d Q^2\left(\sqrt{\frac{\alpha^2 E_N d}{N_0'/2}}\right)}{Q\left(\sqrt{\frac{\alpha'^2 E_N d}{N_0''/2}}\right) + (B_d - 1)\Psi\left(\frac{1}{2}, \sqrt{\frac{\alpha'^2 E_N d}{N_0''/2}}, \sqrt{\frac{\alpha'^2 E_N d}{N_0''/2}}\right)} \tag{26}$$

where $\alpha, \alpha', \beta, \beta', N_0'$ and N_0'' are as defined in (20)*, and $a > -\frac{1}{2N_0}$, b, and c are arbitrary constants.*

3.3 Kounias' Bound

We apply Kounias' lower bound [18] to derive a new lower bound, analogously to the preceding derivations in this section. Although Kounias' bound was used by Kuai, Alajaji and Takahara in [16] to derive a lower bound for the AWGN channel, no specialization of the bound for linear codes was done. In this section, in addition to the straightforward specialization for linear codes, we further simplify the bound by using only the sub-code $\mathcal{C}_d^*$. In this case, the customarily-tedious optimization required in Kounias' bound is direct and can be done analytically. The resulting bound is very simple to evaluate and performs better than Seguin's bound (yet, inferior to (26)) for every value of E_b/N_0.

Under the notations of Section 2, Kounias' bound is given by

$$P\left(\bigcup_{i\in\mathcal{I}} A_i\right) \geq \max_{\mathcal{J}\subseteq\mathcal{I}}\left\{\sum_{i\in\mathcal{J}} P(A_i) - \sum_{i,j\in\mathcal{J},i<j} P(A_i\cap A_j)\right\}.$$

Refereing to our problem, utilization of this bound yields

$$P\left(\bigcup_{i\neq 0} \varepsilon_{0i}|\mathbf{s}_0\right) \geq \max_{\mathcal{J}\subseteq\mathcal{M}\setminus\{0\}}\left\{\sum_{i\in\mathcal{J}} Q\left(\frac{\|\mathbf{s}_0-\mathbf{s}_i\|}{\sqrt{2N_0}}\right) - \sum_{i,j\in\mathcal{J},i<j} \Psi\left(\rho_{ij}, \frac{\|\mathbf{s}_0-\mathbf{s}_i\|}{\sqrt{2N_0}}, \frac{\|\mathbf{s}_0-\mathbf{s}_j\|}{\sqrt{2N_0}}\right)\right\}, \quad (27)$$

where $\mathcal{M} = \{0, 1, \ldots, M-1\}$.[4] To specialize the bound for linear codes, note that the r.h.s. of (27) is a decreasing function of Ψ, therefore, we can use ϱ_{ij} as in (24), resulting in a bound depending only on the weight enumeration. Yet, this bound is still tedious to evaluate for large codes, even when the stepwise algorithm suggested in [16] is used. Thus, we choose to limit the search to *subsets of the sub-code* $\mathcal{C}_d^*$. In this case, we have

$$p(\varepsilon) \geq \max_{1\leq b\leq B_d}\left\{bQ\left(\sqrt{\frac{2E_Nd}{N_0}}\right) - \frac{b(b-1)}{2}\Psi\left(\frac{1}{2}, \sqrt{\frac{2E_Nd}{N_0}}, \sqrt{\frac{2E_Nd}{N_0}}\right)\right\}. \quad (28)$$

Since the r.h.s. of (28) is a concave ($\cap$) function of b, and its second derivative with respect to b is constant, the maximum is achieved by comparing the first derivative to zero and taking the closest integer value to the result, provided that it is in the range $\{1, 2, \ldots, B_d\}$. Thus, the maximum is achieved with

$$b^* = \min\left\{\left[\frac{1}{2} + \frac{Q\left(\sqrt{\frac{2E_Nd}{N_0}}\right)}{\Psi\left(\frac{1}{2}, \sqrt{\frac{2E_Nd}{N_0}}, \sqrt{\frac{2E_Nd}{N_0}}\right)}\right], B_d\right\},$$

where $[x]$ is closest integer to x. Consequently, we have

$$p(\varepsilon) \geq b^*Q\left(\sqrt{\frac{2E_Nd}{N_0}}\right) - \frac{b^*(b^*-1)}{2}\Psi\left(\frac{1}{2}, \sqrt{\frac{2E_Nd}{N_0}}, \sqrt{\frac{2E_Nd}{N_0}}\right). \quad (29)$$

[4] Note that the fact that Kounias' bound allows us to use any subset $\mathcal{J} \subseteq \mathcal{M}\setminus\{0\}$ is insignificant since when lower bounds on the error probability are considered, this step is straightforward (refer to Section 3.2). Hence, in this case Kounias' bound is equivalent to the well known Bonferroni's inclusion-exclusion lower bound ([22]).

4 The Binary Symmetric Channel

Analogously to the previous section, in this section, we apply the bound in Theorem 2.1 to derive new lower bounds for maximum likelihood decoding over the BSC. For the sake of simplicity, we consider only linear codes. Bounds for any block code can be derived in a similar fashion.

In this case, the transmitted codeword is one of $M = 2^K$ equiprobable binary codewords $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1}$ of length N. Denote by p the channel crossover probability. Assuming $\mathbf{c}_0$ was sent, let $\mathbf{x} = \mathbf{c}_0 + \mathbf{e}$ be the received word, where $\mathbf{e}$ is the binary error vector. The maximum likelihood decoder chooses the closest of the M codewords to $\mathbf{x}$ in the Hamming sense, i.e., $\hat{i} = \arg\min_i d_H(\mathbf{c}_i, \mathbf{x})$. Thus, the probability of error given that $\mathbf{c}_0$ was sent is

$$P(\varepsilon|\mathbf{c}_0) = P\left(\cup_{i\neq 0}\varepsilon_{0i}|\mathbf{c}_0\right), \tag{30}$$

where

$$\varepsilon_{0i} = \{\mathbf{x} \in GF(2)^N : w(\mathbf{x}+\mathbf{c}_i) < w(\mathbf{x})\}, \tag{31}$$

assuming $\mathbf{c}_0$ is the all-zero codeword.

Our goal is to lower bound the error probability in (30). Again, when the code used is a binary (N, K) linear code $\mathcal{C}$, we wish to express the bound in terms of the code's weight enumeration and the channel crossover probability alone. Since the method developed in Section 3 is general, and can be used in any case where the error probability admits a union form, we focus only on channel-specific derivations. We have,

$$p(\mathbf{x}|\mathbf{c}_0) = p^{w(\mathbf{x})}(1-p)^{N-w(\mathbf{x})} \tag{32}$$

and

$$deg(\mathbf{x}|\mathbf{c}_0) = |\{\mathbf{c}_i \in \mathcal{C}, i \neq 0 : w(\mathbf{x}+\mathbf{c}_i) < w(\mathbf{x})\}|. \tag{33}$$

In the computation of (2), we encounter the summations over $\mathbf{x} \in \varepsilon_{0i}$ and $\mathbf{x} \in \varepsilon_{0i} \cap \varepsilon_{0j}$. The summands are $p(\mathbf{x}|\mathbf{c}_0)m_i(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{c}_0)m_i^2(\mathbf{x})$, respectively. While the dependence of $p(\mathbf{x}|\mathbf{c}_0)$ on $\mathbf{x}$ is only through $w(\mathbf{x})$, $m_i(\mathbf{x})$ is a function to be optimized and hence might, in general, be chosen to be different for each $\mathbf{x}$. To avoid a tedious evaluation of the considered sums, we prefer to reduce the degrees of freedom in choosing $m_i(\mathbf{x})$ by the restriction

$$m_i(\mathbf{x}) = \eta_i(w(\mathbf{x})), \qquad \eta_i : \mathbb{Z}^+ \mapsto \mathbb{R}.$$

Clearly, since $deg(\mathbf{x}|\mathbf{c}_0)$ is not likely to depend only on $w(\mathbf{x})$ when non-trivial codes are discussed, we may assume that the optimal value for $m_i(\mathbf{x})$ cannot be achieved by any function η_i. Nevertheless, we will discover that the function η_i may still be chosen to yield tighter bounds than the one achieved with the trivial choice of $m_i(\mathbf{x})$. To conclude, according to (2), we have

$$P(\varepsilon) \geq \sum_{i=1}^{M} \frac{\left(\sum_{\mathbf{x}\in\varepsilon_{0i}} p^{w(\mathbf{x})}(1-p)^{N-w(\mathbf{x})}\eta_i(w(\mathbf{x}))\right)^2}{\sum_{j=1}^{M}\sum_{\mathbf{x}\in\varepsilon_{0i}\cap\varepsilon_{0j}} p^{w(\mathbf{x})}(1-p)^{N-w(\mathbf{x})}\eta_i^2(w(\mathbf{x}))}, \tag{34}$$

where $\eta_i : \mathbb{Z}^+ \mapsto \mathbb{R}$ is any function to be optimized. In the proceeding subsection, we suggest several possibilities for η_i in the spirit of (3), namely, we seek approximations for $deg(\mathbf{x}|\mathbf{c}_0)$. For the time being, we evaluate (34) for any η_i whose dependence on $\mathbf{x}$ is only through $w(\mathbf{x})$.

We start by calculating the sum over $\mathbf{x} \in \varepsilon_{0i}$. Define the set $\mathcal{N} = \{1, 2, \ldots N\}$, and for every $\mathcal{M} \subseteq \mathcal{N}$ and $\mathbf{x} \in GF(2)^N$ define $\mathbf{x}_{\mathcal{M}}$ to be the sub-word $(\mathbf{x}(m_1), \ldots, \mathbf{x}(m_{|\mathcal{M}|}))$ of $\mathbf{x}$. Let the *support* of $\mathbf{c}_i$ be the set $\mathcal{S}_i = \{j : \mathbf{c}_i(j) = 1\}$. Hence, $\mathbf{x}_{\mathcal{S}_i}$ is the sub-word of $\mathbf{x}$ consisting of $\mathbf{x}$ in the places $\mathbf{c}_i$ equals 1. Referring to (31), a word $\mathbf{x} \in GF(2)^N$ satisfies $\mathbf{x} \in \varepsilon_{0i}$ if under $\mathbf{c}_i$'s support it has more 1's than 0's. The number of 1's or 0's out of $\mathbf{c}_i$'s support is irrelevant. Thus

$$\mathbf{x} \in \varepsilon_{0i} \quad \text{iff} \quad w(\mathbf{x}_{\mathcal{S}_i}) \geq \left\lfloor \frac{w(\mathbf{c}_i)}{2} \right\rfloor + 1.$$

Accordingly,

$$\begin{aligned} P_{num}(w(\mathbf{c}_i)) &\stackrel{\triangle}{=} \sum_{\mathbf{x} \in \varepsilon_{0i}} p^{w(\mathbf{x})}(1-p)^{N-w(\mathbf{x})} \eta_i(w(\mathbf{x})) \\ &= \sum_{l=\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor + 1}^{w(\mathbf{c}_i)} \sum_{m=0}^{N-w(\mathbf{c}_i)} \binom{w(\mathbf{c}_i)}{l} \binom{N-w(\mathbf{c}_i)}{m} p^{l+m}(1-p)^{N-l-m} \eta_i(l+m). \end{aligned} \tag{35}$$

To avoid cumbersome notations, the notation for $P_{num}(w(\mathbf{c}_i))$ does not reflect its dependence on p and the parameter N. On the more technical side, note that choosing a non-trivial η_i prevents us from using the binomial formula to evaluate the inner sum and thus increases the computational complexity. For the codes tested in this work, this tradeoff was worthwhile.

The evaluation of the sum in the denominator is carried out in the same fashion. In this case,

$$\mathbf{x} \in \varepsilon_{0i} \cap \varepsilon_{0j} \quad \text{iff} \quad w(\mathbf{x}_{\mathcal{S}_i}) \geq \left\lfloor \frac{w(\mathbf{c}_i)}{2} \right\rfloor + 1 \qquad \text{and} \qquad w(\mathbf{x}_{\mathcal{S}_j}) \geq \left\lfloor \frac{w(\mathbf{c}_j)}{2} \right\rfloor + 1.$$

Thus, when $\mathbf{c}_i \neq \mathbf{c}_j$, we have

$$\begin{aligned} \widetilde{P}_{den}(\mathbf{c}_i, \mathbf{c}_j) &\stackrel{\triangle}{=} \sum_{\mathbf{x} \in \varepsilon_{0i} \cap \varepsilon_{0j}, i \neq j} p^{w(\mathbf{x})}(1-p)^{N-w(\mathbf{x})} \eta_i^2(w(\mathbf{x})) \\ &= \sum_{l=0}^{w(\mathbf{c}_i\mathbf{c}_j)} \sum_{m=\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor + 1 - l}^{w(\mathbf{c}_i) - w(\mathbf{c}_i\mathbf{c}_j)} \sum_{n=\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor + 1 - l}^{w(\mathbf{c}_j) - w(\mathbf{c}_i\mathbf{c}_j)} \sum_{k=0}^{N-w(\mathbf{c}_i)-w(\mathbf{c}_j)+w(\mathbf{c}_i\mathbf{c}_j)} \binom{w(\mathbf{c}_i\mathbf{c}_j)}{l} \binom{w(\mathbf{c}_i) - w(\mathbf{c}_i\mathbf{c}_j)}{m} \\ &\cdot \binom{w(\mathbf{c}_j) - w(\mathbf{c}_i\mathbf{c}_j)}{n} \binom{N - w(\mathbf{c}_i) - w(\mathbf{c}_j) + w(\mathbf{c}_i\mathbf{c}_j)}{k} p^{l+m+n+k}(1-p)^{N-l-m-n-k} \eta_i^2(l+m+n+k) \end{aligned} \tag{36}$$

where the first sum in (36) is over the intersection of $\mathbf{c}_i$'s and $\mathbf{c}_j$'s supports - $\mathcal{S}_{ij}$, the second is over $\mathcal{S}_i \backslash \mathcal{S}_{ij}$, the third is over $\mathcal{S}_j \backslash \mathcal{S}_{ij}$ and the forth is over $\mathcal{N} \backslash \mathcal{S}_i \backslash \mathcal{S}_j$. When $\mathbf{c}_i = \mathbf{c}_j$, $\widetilde{P}_{den}(\mathbf{c}_i, \mathbf{c}_i) = P_{den}(\mathbf{c}_i)$, where $P_{den}(\mathbf{c}_i)$ is defined just as $P_{num}(\mathbf{c}_i)$, only with $\eta_i(\cdot)$ raised to the power of two.

Clearly, $\widetilde{P}_{den}(\mathbf{c}_i, \mathbf{c}_j)$ does not depend on $\mathbf{c}_i$ and $\mathbf{c}_j$ solely through $w(\mathbf{c}_i)$ and $w(\mathbf{c}_j)$ since it includes the expression $w(\mathbf{c}_i\mathbf{c}_j)$. Thus, its evaluation requires more than the weight enumeration of the code. Recall dealing with an equivalent problem in the AWGN channel, it is clear that to remove the obstacle in specializing the bound to linear codes, the following proposition comes in handy.

Proposition 4.1 *$\widetilde{P}_{den}(\mathbf{c}_i, \mathbf{c}_j)$ is monotonically increasing in $w(\mathbf{c}_i\mathbf{c}_j)$ for any $w(\mathbf{c}_i\mathbf{c}_j) \leq \min\{w(\mathbf{c}_i) - 1, w(\mathbf{c}_j) - 1\}$, $w(\mathbf{c}_i)$, $w(\mathbf{c}_j)$ and $\eta_i : \mathbb{Z}^+ \mapsto \mathbb{R}^+$.*

Referring to (33), it is clear that the demand for positive η_i is not restricting. The proof of Proposition 4.1 is given in Appendix B.1. To utilize Proposition 4.1, define

$$\overline{w}(\mathbf{c}_i\mathbf{c}_j) \stackrel{\triangle}{=} \min\left\{ w(\mathbf{c}_i), w(\mathbf{c}_j), \left\lfloor \frac{w(\mathbf{c}_i) + w(\mathbf{c}_j) - d}{2} \right\rfloor \right\}.$$

Since

$$w(\mathbf{c}_i\mathbf{c}_j) = \frac{w(\mathbf{c}_i) + w(\mathbf{c}_j) - d_H(\mathbf{c}_i, \mathbf{c}_j)}{2},$$

it is clear that $w(\mathbf{c}_i\mathbf{c}_j) \leq \overline{w}(\mathbf{c}_i\mathbf{c}_j)$. Thus,

$$\begin{aligned}
&\widetilde{P}_{den}(\mathbf{c}_i, \mathbf{c}_j) \leq P_{den}(w(\mathbf{c}_i), w(\mathbf{c}_j)) \\
&\quad \triangleq \sum_{l=0}^{\overline{w}(\mathbf{c}_i\mathbf{c}_j)} \sum_{m=\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor+1-l}^{w(\mathbf{c}_i)-\overline{w}(\mathbf{c}_i\mathbf{c}_j)} \sum_{n=\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor+1-l}^{w(\mathbf{c}_j)-\overline{w}(\mathbf{c}_i\mathbf{c}_j)} \sum_{k=0}^{N-w(\mathbf{c}_i)-w(\mathbf{c}_j)+\overline{w}(\mathbf{c}_i\mathbf{c}_j)} \binom{\overline{w}(\mathbf{c}_i\mathbf{c}_j)}{l} \binom{w(\mathbf{c}_i) - \overline{w}(\mathbf{c}_i\mathbf{c}_j)}{m} \\
&\cdot \binom{w(\mathbf{c}_j) - \overline{w}(\mathbf{c}_i\mathbf{c}_j)}{n} \binom{N - w(\mathbf{c}_i) - w(\mathbf{c}_j) + \overline{w}(\mathbf{c}_i\mathbf{c}_j)}{k} p^{l+m+n+k}(1-p)^{N-l-m-n-k}\eta_i^2(l+m+n+k),
\end{aligned} \tag{37}$$

with the understanding that $\binom{n}{k} = 0$ when $k > n$. Now it is possible to derive a bound using only the code's weight enumeration. Thus, the new lower bound on the error probability of a linear code $\mathcal{C}$ on the BSC is given by

$$P(\varepsilon) \geq \sum_{n=1}^{N} \frac{B_n P_{num}^2(n)}{P_{den}(n) + (B_n - 1)P_{den}(n,n) + \sum_{m=1, m\neq n}^{N} B_m P_{den}(n,m)}, \tag{38}$$

where $P_{num}(n)$, $P_{den}(n)$ and $P_{den}(n,m)$ include the function η_i, which can be optimized to yield the tightest bound.

4.1 Approximations to $deg(\mathbf{x}|\mathbf{c}_0)$

We seek a function η_i, of the form

$$\eta_i(w(\mathbf{x})) = \frac{1}{\widetilde{deg}(w(\mathbf{x}))}, \tag{39}$$

where $\widetilde{deg}(w(\mathbf{x}))$ is any approximation of $deg(\mathbf{x}|\mathbf{c}_0)$ whose dependence on $\mathbf{x}$ is only through $w(\mathbf{x})$. Referring to (33), $deg(\mathbf{x}|\mathbf{c}_0)$ is the number of words with Hamming weight less than $w(\mathbf{x})$ in the coset $\mathcal{C} + \mathbf{x}$. Thus, we are interested in the weight enumeration of this coset when the only knowledge on $\mathbf{x}$ is $w(\mathbf{x})$. As a simple example, consider a 1-bit parity check code. Clearly, there are only two cosets in this case. The first is the code itself, i.e., the set of all even-weight words. The second is the set of all odd-weight words. Thus, given a received word $\mathbf{x}$, its weight is sufficient to identify the correct coset and $deg(\mathbf{x}|\mathbf{c}_0)$ is known exactly. However, the evaluation of the error probability for this code is trivial in the first place. There are several codes whose cosets weight enumeration can be found in the current literature. Yet, even for simple codes, the weight enumeration of $\mathcal{C} + \mathbf{x}$ cannot, in general, be evaluated using $w(\mathbf{x})$ alone.

We include here two possible approximations to $deg(\mathbf{x}|\mathbf{c}_0)$ using only the existing information on $\mathbf{x}$ and the code's weight enumeration. In the first approximation, we view $deg(\mathbf{x}|\mathbf{c}_0)$ as $|\mathcal{C}| \cdot P(w(\mathbf{x}+\mathbf{c}) < w(\mathbf{x}))$, where P is the probability measurement inferred by a uniform distribution on the codewords of $\mathcal{C}$. Let $\mathbf{x}$ be fixed and let $\mathbf{c}$ be a codeword chosen randomly with uniform distribution. Considering $w(\mathbf{x} + \mathbf{c})$ as a random variable, by the Chernoff bound, we have

$$P(w(\mathbf{x} + \mathbf{c}) < w(\mathbf{x})) \leq \mathbb{E}\left\{e^{-a(w(\mathbf{x}+\mathbf{c})-w(\mathbf{x}))}\right\}, \tag{40}$$

where the expectation is over all possible codewords $\mathbf{c}$, and $a \geq 0$ is an arbitrary parameter. Clearly,

$$\begin{aligned}\mathbb{E}\left\{e^{-a(w(\mathbf{x}+\mathbf{c})-w(\mathbf{x}))}\right\} &= \frac{1}{|\mathcal{C}|}\sum_{\mathbf{c}\in\mathcal{C}} e^{-a(w(\mathbf{x}+\mathbf{c})-w(\mathbf{x}))} \\ &\leq \frac{1}{|\mathcal{C}|}\sum_{\mathbf{c}\in\mathcal{C}} e^{-a(|w(\mathbf{x})-w(\mathbf{c})|-w(\mathbf{x}))} \\ &= \frac{e^{aw(\mathbf{x})}}{|\mathcal{C}|}\sum_{i=0}^{N} B_i e^{-a|w(\mathbf{x})-i|},\end{aligned} \tag{41}$$

where the inequality in (41) is since $w(\mathbf{x}+\mathbf{c}) \geq |w(\mathbf{x}) - w(\mathbf{c})|$. The approximation for $deg(\mathbf{x}|\mathbf{c}_0)$ is therefore

$$\widetilde{deg}(w(\mathbf{x})) = e^{aw(\mathbf{x})}\sum_{i=0}^{N} B_i e^{-a|w(\mathbf{x})-i|}, \tag{42}$$

where $a \geq 0$ is a parameter to be optimized.

The second approximation uses a different method. Clearly,

$$deg(\mathbf{x}) = \begin{cases} 0 & w(\mathbf{x}) \leq \left\lfloor \frac{d-1}{2} \right\rfloor \\ \sum_{i=0}^{w(\mathbf{x})-1} B_i^{\mathbf{x}} & \left\lfloor \frac{d-1}{2} \right\rfloor < w(\mathbf{x}) < N \\ |\mathcal{C}| - 1 & w(\mathbf{x}) = N, \end{cases} \tag{43}$$

where $B_i^{\mathbf{x}}$ is the number of words of weight i in the coset $\mathcal{C} + \mathbf{x}$. Thus, $deg(\mathbf{x}|\mathbf{c}_0)$ is monotonically increasing in $w(\mathbf{x})$, with known values for $w(\mathbf{x}) = \left\lfloor \frac{d-1}{2} \right\rfloor$ and $w(\mathbf{x}) = N$. By choosing a family of monotone functions (concave or convex) passing through these points, we have the following approximation

$$\widetilde{deg}(w(\mathbf{x})) = \begin{cases} (|\mathcal{C}| - 1)\left(\frac{w(\mathbf{x}) - \left\lfloor \frac{d-1}{2} \right\rfloor}{N - \left\lfloor \frac{d-1}{2} \right\rfloor}\right)^a & w(\mathbf{x}) > \left\lfloor \frac{d-1}{2} \right\rfloor \\ 0 & \text{else}, \end{cases} \tag{44}$$

where $a \geq 0$ is a parameter to be optimized. This approximation is easier to evaluate than the previous one since no summation is required. Moreover, only the size of the code, its length and its minimum distance are used.

4.2 Lower Bounds Using the Sub-code $\mathcal{C}_i^*$ and the Code's Covering Radius

In this section, we consider two variations on the bound given in (38). These two variations will both reduce the complexity of the bound as well as tighten it. Denote by t the *covering radius* of the code $\mathcal{C}$,

$$t = \max_{\mathbf{f}\in GF(2)^N} \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{f}, \mathbf{c}).$$

Namely, t is the maximum number of errors that can be corrected. Clearly,

$$P_{\mathcal{C}}(\varepsilon) = P_{\mathcal{C}}(\varepsilon, w(\mathbf{x}) \leq t) + P(w(\mathbf{x}) > t).$$

As noted in [14], when lower bounds on the error probability are discussed, an upper bound $M \geq t$ can be used if t is not known. Furthermore, as in Section 3.2, since

$$P_{\mathcal{C}}(\varepsilon, w(\mathbf{x}) \leq M) \geq P_{\mathcal{C}_i^*}(\varepsilon, w(\mathbf{x}) \leq M),$$

where $\mathcal{C}_i^* = \{\mathbf{c} \in \mathcal{C} : w(\mathbf{c}) = i\} \cup \{\mathbf{c}_0\}$, we may compute the bound disregarding all codewords of weight other than i. Although the numerical analysis shows that best results are achieved with $i = d$, we prefer this general form for later reference. Consequently, we have the following proposition.

Proposition 4.2 *Let $\mathcal{C}$ be any linear code over $GF(2)$ of length N and minimum distance d. Let B_i be the number of codewords of hamming weight i, $d \leq i \leq N - \lceil \frac{d}{2} \rceil$. The decoding error probability on a BSC with crossover probability p is lower bounded by*

$$P(\varepsilon) \geq LB_i(\eta_i, p) \triangleq \frac{B_i \bar{P}_{num}^2(i)}{\bar{P}_{den}(i) + (B_i - 1)\bar{P}_{den}(i,i)} + P_M, \tag{45}$$

where

$$\bar{P}_{num}(i) \triangleq \sum_{l=\lfloor \frac{i}{2} \rfloor + 1}^{i} \sum_{m=0}^{M-l} \binom{i}{l} \binom{N-i}{m} p^{l+m} (1-p)^{N-l-m} \eta_i(l+m), \tag{46}$$

$$\bar{P}_{den}(i) \triangleq \sum_{l=\lfloor \frac{i}{2} \rfloor + 1}^{i} \sum_{m=0}^{M-l} \binom{i}{l} \binom{N-i}{m} p^{l+m} (1-p)^{N-l-m} \eta_i^2(l+m), \tag{47}$$

$$\bar{P}_{den}(i,i) \triangleq \sum_{l=0}^{i-\lceil \frac{d}{2} \rceil} \sum_{m=\lfloor \frac{i}{2} \rfloor + 1 - l}^{\lceil \frac{d}{2} \rceil} \sum_{n=\lfloor \frac{i}{2} \rfloor + 1 - l}^{\lceil \frac{d}{2} \rceil} \sum_{k=0}^{M-l-m-n} \binom{i - \lceil \frac{d}{2} \rceil}{l} \binom{\lceil \frac{d}{2} \rceil}{m} \binom{\lceil \frac{d}{2} \rceil}{n} \binom{N - i - \lceil \frac{d}{2} \rceil}{k} \cdot p^{l+m+n+k} (1-p)^{N-l-m-n-k} \eta_i^2(l+m+n+k), \tag{48}$$

$$P_M \triangleq \sum_{l=M+1}^{N} \binom{N}{l} p^l (1-p)^{N-l} \tag{49}$$

and $\eta_i : \mathbb{Z}^+ \mapsto \mathbb{R}^+$ is any function to be optimized.

Note that the demand $i \leq N - \lceil \frac{d}{2} \rceil$ is not restricting, since when $i > N - \lceil \frac{d}{2} \rceil$ there is only one codeword with weight i, and the sub-code $\mathcal{C}_i^*$ is degenerated.

Proof . Based on the preceding discussion, substitute $B_n = 0$ for every $n \neq i$ in (38). To use the covering radius of the code, evading high values of $w(\mathbf{x})$, we may alter the expressions in (35) and (37) to include only words $\mathbf{x}$ with weight smaller than M by changing the upper bound of the last summation in each expression. P_M is the probability of more than M bit errors. $\square$

Note that when $\lfloor \frac{i}{2} \rfloor + 1 - l > \lceil \frac{d}{2} \rceil$ the sums over m and n in (48) are empty. Thus, the value of $\bar{P}_{den}(i,i)$ is unchanged if we sum over $\lfloor \frac{i}{2} \rfloor + 1 - \lceil \frac{d}{2} \rceil \leq l \leq i - \lceil \frac{d}{2} \rceil$ instead of $0 \leq l \leq i - \lceil \frac{d}{2} \rceil$.

To choose a proper η_i, we return to Section 4.1. Although the approximations therein refer to the bound given in (38), i.e., when the whole code is used, we find them useful in (45) for two main reasons. First, since η_i defines only an essence of behavior, the approximations in Section 4.1 may be sufficient as are. Second, even if several variations are required, the *methods* suggested in Section 4.1 are still applicable. For example, in Section 6, where the bound given in Proposition 4.2 is utilized, the following variation of (44) was used

$$\widetilde{deg}(w(\mathbf{x})) = \begin{cases} (B_d - 1) \left(\frac{w(\mathbf{x}) - \lfloor \frac{d-1}{2} \rfloor}{N - \lfloor \frac{d-1}{2} \rfloor} \right)^a & w(\mathbf{x}) > \lfloor \frac{d-1}{2} \rfloor \\ 0 & \text{else.} \end{cases} \tag{50}$$

Finally, we refer to the asymptotic tightness of the bounds presented in this section. Let $UB(p)$ be the union bound for the BSC, i.e.,

$$P(\varepsilon) \leq UB(p) \stackrel{\triangle}{=} \sum_{i=1}^{N} B_i P(i), \tag{51}$$

where $P(i)$ is the pairwise error probability given by

$$P(i) = \sum_{l=\lfloor \frac{i}{2} \rfloor+1}^{i} \binom{i}{l} p^l (1-p)^{i-l}.$$

The following proposition states that the new bounds are tight for $p \to 0$.

Proposition 4.3 *Consider the bound in Proposition 4.2 when $i = d$. Then, for any function $\eta_d > 0$ which is independent of p,*

$$\lim_{p \to 0} \frac{LB_d(\eta_d, p)}{UB(p)} = 1.$$

Note that the condition on η_d is not restricting since η_d depends on the code rather than the channel. Proposition 4.3 is proved in Appendix B.2.

5 Upper Bounds on the Error Exponent

In this section, we calculate the upper bound on the error exponent resulting from the bound given in Proposition 4.2. We prove that a non trivial choice of η_i may result in a tighter bound on the error exponent than the one resulting from a de Caen-based bound, and identify the optimal choice of η_i. It is important to mention, however, that only bounds for specific codes are discussed, and not bounds on the error exponent of the BSC in general.

We first introduce the required notations. Let $\{\mathcal{C}_N\}$ be any sequence of codes, each of which is of length N and minimum distance d_N. For every $d_N < i \leq N$, denote by δ_i the ratio $\frac{i}{N}$. Let B_i^N be the number of codewords of weight i in each code. We consider only sequences of codes for which the limits $\lim_{N\to\infty} \frac{1}{N} \log B_i^N$ and $\lim_{N\to\infty} \frac{d_N}{N}$ exists, and denote their values by $E_B^{\delta_i}$ and δ_d, respectively. Let $F(N)$ and $G(N)$ be any two functions. If

$$\lim_{N\to\infty} \frac{1}{N} \log F(N) \leq \lim_{N\to\infty} \frac{1}{N} \log G(N)$$

we write $\frac{1}{N} \log F(N) \preceq \frac{1}{N} \log G(N)$, namely, $F(N)$ is exponentially smaller than $G(N)$.

To calculate the upper bound on the error exponent resulting from the bound in Proposition 4.2 and analyze the results, we substitute $M = N$ in equations (45) to (49), i.e., assume no knowledge on the covering radius is available. This assumption only weakens the bound, mainly at high values of p. Thus, for any $d_N \leq i \leq N - \lceil \frac{d_N}{2} \rceil$, the following bound is considered

$$P(\varepsilon) \geq \frac{B_i^N P_{num}^2(i)}{P_{den}(i) + (B_i^N - 1) P_{den}(i,i)}, \tag{52}$$

where $P_{num}(i)$, $P_{den}(i)$ and $P_{den}(i,i)$ were defined in Section 4.

For easy reference and facile understanding of this section, we briefly introduce the outline of the analysis and summarize the main results. Consider the bound in (52). We wish to calculate the resulting bound on the error exponent, and to identify the optimal choice of the function η_i. Clearly, since the denominator of the r.h.s. of (52) is a sum of two expressions, the exponential behavior of the bound in (52) depends on which expression dominates. In the first part of the analysis, whose results are given by Proposition 5.1, we show that this observation translates to a *condition on the code*, which determines the value of the new bound on the error exponent in each case.

In the second part of the analysis, whose results are given by Corollary 5.2 and the discussion following it, we analyze the condition on the code and the resulting bound on the error exponent when this condition is satisfied. It is shown therein, that if the difference between the triplets error exponent and the pairwise error exponent is not too small (i.e., the rate of the code is not too large), the condition on the code is satisfied, and the resulting bound on the error exponent is given by

$$-\frac{1}{N}\log P(\varepsilon) \preceq -\delta_i \log\sqrt{4p(1-p)} - E_B^{\delta_i},$$

for any $\delta_d \leq \delta_i \leq 1-\frac{1}{2}\delta_d$. Namely, one can use the union bound to derive a valid lower bound on the error probability. In this case, we say that the *union bound analysis applies*, namely, the union bound is exponentially tight.

This far, we have not considered the choice of the function η_i. Our main result, given by Proposition 5.3, is that while it can be easily proved that when the condition on the code is satisfied, the trivial η_i is optimal, this is not the case when it is not satisfied. In this case, a non trivial η_i can extend the range of rates for which the union bound analysis applies, thus achieving a tighter bound on the error exponent. The optimal value of η_i, the range of rates for which the union bound analysis applies and a quantification of the improvement over the bound with trivial η_i are given in Proposition 5.3 and the discussion which follows.

5.1 Analysis

We start with several definitions. For $t=\delta_t N$, define

$$E_\eta^{\delta_i}(\delta_t) = \lim_{N\to\infty} -\frac{1}{N}\log \eta_{\delta_i N}(t). \tag{53}$$

Since $\eta_i(t)$ is any function to be optimized, we may reduce the set of possible functions to assure that the limit in (53) exists. For $\eta_i(t)\equiv 1$, we have $E_\eta^{\delta_i}(\delta_t)\equiv 0$. Analogously to the previous sections, we denote this case as the *trivial* choice of $E_\eta^{\delta_i}(\delta_t)$. For any δ_d and $\delta_d \leq \delta_i \leq 1-\frac{1}{2}\delta_d$, define the following regions in $[0,1]^2$ and $[0,1]^4$, respectively

$$\mathcal{D}_1 = \left\{(\delta_l,\delta_m)\in[0,1]^2 : \frac{\delta_i}{2}\leq\delta_l\leq\delta_i, 0\leq\delta_m\leq 1-\delta_i\right\} \tag{54}$$

and

$$\begin{aligned}\mathcal{D}_2 = \Big\{(\delta_l,\delta_m,\delta_n,\delta_k)\in[0,1]^4 : \quad & \tfrac{1}{2}(\delta_i-\delta_d)\leq\delta_l\leq\delta_i-\tfrac{\delta_d}{2}, \tfrac{\delta_i}{2}-\delta_l\leq\delta_m\leq\tfrac{\delta_d}{2},\\ & \tfrac{\delta_i}{2}-\delta_l\leq\delta_n\leq\tfrac{\delta_d}{2}, 0\leq\delta_k\leq 1-\delta_i-\tfrac{\delta_d}{2}\Big\}.\end{aligned} \tag{55}$$

Let $H(x)$ be the binary entropy function

$$H(x) = -x\log(x) - (1-x)\log(1-x).$$

Define

$$E_1^{\delta_i}(\delta_l, \delta_m, p) \stackrel{\triangle}{=} -\delta_i H\left(\frac{\delta_l}{\delta_i}\right) - (1-\delta_i) H\left(\frac{\delta_m}{1-\delta_i}\right) + (\delta_l + \delta_m)\log\left(\frac{1-p}{p}\right) - \log(1-p) \tag{56}$$

and

$$\begin{aligned} E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p) \stackrel{\triangle}{=} & -(\delta_i - \delta_d/2) H\left(\frac{\delta_l}{\delta_i - \delta_d/2}\right) - \frac{\delta_d}{2} H\left(\frac{\delta_m}{\delta_d/2}\right) - \frac{\delta_d}{2} H\left(\frac{\delta_n}{\delta_d/2}\right) \\ & -(1-\delta_i - \delta_d/2) H\left(\frac{\delta_k}{1-\delta_i-\delta_d/2}\right) + (\delta_l + \delta_m + \delta_n + \delta_k)\log\left(\frac{1-p}{p}\right) - \log(1-p). \end{aligned} \tag{57}$$

Under these definitions, we have the following proposition.

Proposition 5.1 *Let $\{\mathcal{C}_N\}$ be a sequence of codes for the BSC. Let p be the crossover probability of the channel. Then, for any $\delta_d \leq \delta_i \leq 1 - \frac{1}{2}\delta_d$, and for any piecewise continuous function $E_\eta^{\delta_i} : [0,1] \mapsto \mathbb{R}^+$, we have*

$$\begin{aligned} -\frac{1}{N}\log P(\varepsilon) \preceq 2\min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l, \delta_m, p) + E_\eta^{\delta_i}(\delta_l + \delta_m)\right\} - E_B^{\delta_i} \\ - \min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l, \delta_m, p) + 2E_\eta^{\delta_i}(\delta_l + \delta_m)\right\} \end{aligned} \tag{58}$$

if

$$\begin{aligned} E_B^{\delta_i} \leq \min_{\mathcal{D}_2}\left\{E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p) + 2E_\eta^{\delta_i}(\delta_l + \delta_m + \delta_n + \delta_k)\right\} \\ - \min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l, \delta_m, p) + 2E_\eta^{\delta_i}(\delta_l + \delta_m)\right\} \end{aligned} \tag{59}$$

and

$$\begin{aligned} -\frac{1}{N}\log P(\varepsilon) \preceq 2\min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l, \delta_m, p) + E_\eta^{\delta_i}(\delta_l + \delta_m)\right\} \\ - \min_{\mathcal{D}_2}\left\{E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p) + 2E_\eta^{\delta_i}(\delta_l + \delta_m + \delta_n + \delta_k)\right\} \end{aligned} \tag{60}$$

otherwise.

The condition in (59) is a condition on the code's parameter[5] $E_B^{\delta_i}$ (hereafter referred to as the *condition on the code*). The essence of Proposition 5.1, is the fact that the new bound on the error exponent is given by one of two different expressions, corresponding to the cases where condition (59) is satisfied or not. The complete proof of Proposition 5.1 is given in Appendix C.1.

We first analyze the case where the sequence of codes satisfies condition (59). From Appendix C.1, it is clear that $\min_{\mathcal{D}_1} E_1^{\delta_i}(\delta_l, \delta_m, p)$ is the error exponent for the pairwise error probability, while $\min_{\mathcal{D}_2} E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p)$ is the error exponent for triplets. Hence, we expect to have

$$\min_{\mathcal{D}_2} E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p) \geq \min_{\mathcal{D}_1} E_1^{\delta_i}(\delta_l, \delta_m, p),$$

which means that the set of possible functions $E_\eta^{\delta_i}$ for which the r.h.s. of (59) is nonnegative is not empty (it includes at least the trivial choice). In such a case, condition (59) is not trivial, and the following corollary is constructive.

[5] We will see later that this is the condition for applying the union bound analysis. As noted in [11], this condition can be referred to as a condition on the code's parameter $E_B^{\delta_i}$, or a condition on the code's rate, R.

Corollary 5.2 *Let δ_i, $\delta_d \le \delta_i \le 1 - \frac{1}{2}\delta_d$, be fixed. Let $\mathcal{S}_\eta$ be a set of functions $E_\eta^{\delta_i} : [0,1] \mapsto \mathbb{R}$, indexed by η, which includes the trivial choice. Suppose that $\{\mathcal{C}_N\}$ is a sequence of codes for which condition* (59) *is satisfied for every value of $p \in \mathcal{P}$, for some $\mathcal{P} \subseteq (0, \frac{1}{2})$, and for every choice of $E_\eta^{\delta_i} \in \mathcal{S}_\eta$. Then, for every $p \in \mathcal{P}$, the trivial choice of $E_\eta^{\delta_i}$ minimizes the upper bound in* (58) *over all choices of $E_\eta^{\delta_i} \in \mathcal{S}_\eta$, and we have*

$$-\frac{1}{N}\log P(\varepsilon) \preceq -\delta_i \log\sqrt{4p(1-p)} - E_B^{\delta_i}. \tag{61}$$

Proof . When condition (59) is satisfied and $E_\eta^{\delta_i}(\delta_t) \equiv 0$, we have

$$-\frac{1}{N}\log P(\varepsilon) \preceq \min_{\mathcal{D}_1} E_1^{\delta_i}(\delta_l, \delta_m, p) - E_B^{\delta_i}. \tag{62}$$

Subtracting the r.h.s. of (62) from the r.h.s. of (58), we have

$$\begin{aligned}
2\min_{\mathcal{D}_1}&\left\{E_1^{\delta_i}(l,m,p) + E_\eta^{\delta_i}(l+m)\right\} - E_B^{\delta_i} - \min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(l,m,p) + 2E_\eta^{\delta_i}(l+m)\right\} \\
&\qquad - \left(\min_{\mathcal{D}_1} E_1^{\delta_i}(l,m,p) - E_B^{\delta_i}\right) \\
= \min_{\mathcal{D}_1}&\left\{2E_1^{\delta_i}(l,m,p) + 2E_\eta^{\delta_i}(l+m)\right\} \\
&\qquad - \left(\min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(l,m,p) + 2E_\eta^{\delta_i}(l+m)\right\} + \min_{\mathcal{D}_1} E_1^{\delta_i}(l,m,p)\right) \ge 0
\end{aligned}$$

for any $E_\eta^{\delta_i} \in \mathcal{S}_\eta$. Thus, when condition (59) is satisfied, $E_\eta^{\delta_i}(\delta_t) \equiv 0$ is the optimal choice and (62) is the resulting bound. Finally, it is easy to show that

$$\begin{aligned}
\min_{\mathcal{D}_1} E_1^{\delta_i}(\delta_l, \delta_m, p) &= E_1^{\delta_i}(\delta_i/2, p(1-\delta_i), p) \\
&= -\delta_i \log\sqrt{4p(1-p)},
\end{aligned}$$

and (61) immediately follows. □

At this point, several remarks are in order. The bound

$$-\frac{1}{N}\log P(\varepsilon) \preceq \min_{\mathcal{D}_1} E_1^{\delta_i}(\delta_l, \delta_m, p) = -\delta_i \log\sqrt{4p(1-p)}$$

is the well known *two codewords* bound, where $-\log\sqrt{4p(1-p)}$ is the *Bhattacharyya distance* for the BSC (see, for example, [21, pp. 88]). However, the bound in (61) implies that, under certain conditions, when there are exponentially many codewords of weight i, the exponent $E_B^{\delta_i}$ can be subtracted, yielding a tighter upper bound. This is to say that a union bound analysis results in a valid upper bound on the error exponent (a lower bound on the error probability). Thus, by optimizing the bound on δ_i, i.e., choosing the correct sub-code, the union bound analysis gives the true error exponent for the code[6]. The fact that union bound analysis yields the true error exponent

[6]The union bound, given by $P(\varepsilon|\mathbf{c}_0) \le \sum_{w=1}^{N} B_w P(\varepsilon_{0w}|\mathbf{c}_0)$, has only polynomially many summands. Merely one of them determines the exponential behavior. Consequently, if we calculate a lower bound on the error probability using this sub-code, and find out that the union bound analysis applies, this is the true exponential behavior. In this context, it is clear that if our choice of $E_\eta^{\delta_i}$ yields the true error exponent, no other $E_\eta^{\delta_i}$ is required.

for random codes is well known [23]. In [12], Barg and Forney used this argument to derive the exact error exponents for typical codes from Shannon's random code ensemble as well as typical codes from a random linear code ensemble. Yet, the bound in (61) is valid for any given code, as long as condition (59) is satisfied.

Thus far, the results given by Proposition 5.1 were analyzed only as long as the condition on the code is satisfied. The main statement in Corollary 5.2 is that the union bound analysis applies in this case. However, note that the r.h.s. of (59) includes the function $E_\eta^{\delta_i}$, which can be optimized. The most important result of this section, as we will see below, is that by choosing a non trivial $E_\eta^{\delta_i}$, the range of rates for which the union bound analysis applies can be widened. To see this, the minimization $\min_{\mathcal{D}_2} E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p)$ should be discussed.

Using the Karush-Kuhn-Tucker conditions ([24]), one can show[7] that

$$\min_{\mathcal{D}_2} E_2^{\delta_i}(\delta_l, \delta_m, \delta_n, \delta_k, p) = E_2^{\delta_i}(\delta_{l_2}, \delta_{m_2}, \delta_{n_2}, \delta_{k_2}, p),$$

where

$$\begin{aligned} \delta_{m_2} &= \delta_i/2 - \delta_{l_2}, \\ \delta_{n_2} &= \delta_i/2 - \delta_{l_2}, \\ \delta_{k_2} &= p\left(1 - \delta_i - \frac{\delta_d}{2}\right), \end{aligned}$$

and δ_{l_2} is the only root (with respect to δ_l) of the following cubic equation

$$\frac{\delta_l\left(\frac{\delta_d}{2} - \frac{\delta_i}{2} + \delta_l\right)^2}{\left(\frac{\delta_i}{2} - \delta_l\right)^2\left(\delta_i - \frac{\delta_d}{2} - \delta_l\right)} = \frac{1-p}{p}, \tag{63}$$

such that $(\delta_{l_2}, \delta_{m_2}, \delta_{n_2}, \delta_{k_2}) \in \mathcal{D}_2$. Since this solution is rather cumbersome to analyze, we handle here only the special case where $\delta_i = \delta_d$, namely, the sub-code $\mathcal{C}_d^*$ is used. In this case, equation (63) has a simple solution and our course of action and choice of $E_\eta^{\delta_i}$ becomes clearer. The general case is analogous, and yields similar results. We return to it at the end of this section.

When $\delta_i = \delta_d$, (63) simplifies to

$$\left(\frac{\delta_l}{\frac{\delta_d}{2} - \delta_l}\right)^3 = \frac{1-p}{p},$$

yielding the following solution to the minimization of $E_2^{\delta_d}(\delta_l, \delta_m, \delta_n, \delta_k, p)$ over $\mathcal{D}_2$

$$\begin{aligned} \delta_{l_2} &= \frac{\frac{\delta_d}{2}}{1 + \sqrt[3]{\frac{p}{1-p}}}, & \delta_{m_2} &= \frac{\delta_d}{2} - \delta_{l_2} = \frac{\frac{\delta_d}{2}\sqrt[3]{\frac{p}{1-p}}}{1+\sqrt[3]{\frac{p}{1-p}}}, \\ \delta_{k_2} &= p\left(N - \frac{3\delta_d}{2}\right), & \delta_{n_2} &= \frac{\delta_d}{2} - \delta_{l_2} = \frac{\frac{\delta_d}{2}\sqrt[3]{\frac{p}{1-p}}}{1+\sqrt[3]{\frac{p}{1-p}}}, \end{aligned}$$

The solutions of the minimization of $E_1^{\delta_d}(\delta_l, \delta_m, p)$ over $\mathcal{D}_1$ are

$$\delta_{l_1} = \frac{\delta_d}{2}, \qquad \delta_{m_1} = p(1 - \delta_d).$$

[7] The complete derivations can be found in [25]

Define $C(E_B^{\delta_d},p)$ as

$$C(E_B^{\delta_d},p)\stackrel{\triangle}{=}E_B^{\delta_d} - \Big(E_2^{\delta_d}(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2},p) - E_1^{\delta_d}(\delta_{l_1},\delta_{m_1},p)\Big)$$

and $M(p)$ as

$$M(p)\stackrel{\triangle}{=}\min_{\mathcal{D}_2\cap\mathcal{D}_2'} E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) - E_2^{\delta_d}(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2},p),$$

where $\mathcal{D}_2' = \{(\delta_l,\delta_m,\delta_n,\delta_k)\in[0,1]^4 : \delta_l+\delta_m+\delta_n+\delta_k \le \delta_{l_1}+\delta_{m_1}\}$. All this said, the following is the main proposition in this section.

Proposition 5.3 *For any $0 < C(E_B^{\delta_d},p) \le M(p)$, the optimal choice of $E_\eta^{\delta_d}$ is given by*

$$E_\eta^{\delta_d}(\delta_t) = \begin{cases} C(E_B^{\delta_d},p) & \delta_t > \delta_{l_1}+\delta_{m_1} \\ 0 & else, \end{cases} \tag{64}$$

and we have

$$-\frac{1}{N}\log P(\varepsilon) \preceq -\delta_d \log\sqrt{4p(1-p)} - E_B^{\delta_d}. \tag{65}$$

Observe that the requirement $C(E_B^{\delta_d},p)\le 0$ is simply the condition on the code (i.e., equation (59)), with the trivial $E_\eta^{\delta_d}$. Thus, by using a de Caen-based bound, one can only show that the union bound analysis applies when $C(E_B^{\delta_d},p)\le 0$. However, since it can be easily proved that $M(p)>0$ for any $0<p<\frac{1}{2}$, Proposition 5.3 states that by choosing a non trivial $E_\eta^{\delta_d}$, the union bound analysis can be shown to apply in a wider range, $C(E_B^{\delta_d},p)\le M(p)$. Furthermore, in Appendix C.2, where Proposition 5.3 is proved, we show that when $0 < C(E_B^{\delta_d},p) \le M(p)$, the union bound analysis tightens the bound on the error exponent, with respect to the bound with the trivial $E_\eta^{\delta_d}$, by exactly $C(E_B^{\delta_d},p)$. When $C(E_B^{\delta_d},p)>M(p)$, and the new bound does not result in union bound analysis, $E_\eta^{\delta_d}$ as defined in (64) can still tighten the bound with respect to the trivial $E_\eta^{\delta_d}$, this time by as much as $M(p)$, regardless of $C(E_B^{\delta_d},p)$.

We give here only an intuitive explanation for Proposition 5.3. The complete proof can be found in Appendix C.2. We wish to prove that the union bound analysis, namely, the bound in (58) with the trivial $E_\eta^{\delta_d}$, may be applicable even when $C(E_B^{\delta_d},p)>0$. Observe that the r.h.s. of (59) is the difference between two minimization problems. Suppose that there exists a function $E_\eta^{\delta_d}$, such that the result of the minimization over $\mathcal{D}_2$ is increased with respect to the trivial $E_\eta^{\delta_d}$, while the result of the minimization over $\mathcal{D}_1$ is unchanged. If this is possible, the value of the r.h.s. of (59) is increased, thus the range in which the union bound analysis apply is widened. The bound in (58) is the same as it was with the trivial $E_\eta^{\delta_d}$, since the proposed $E_\eta^{\delta_d}$ does not change the result of the minimization over $\mathcal{D}_1$. To see that such an $E_\eta^{\delta_d}$ does exist, observe that both $E_2^{\delta_d}$ and $E_1^{\delta_d}$ are convex functions, and their minimization points satisfy $\delta_{l_2}+\delta_{m_2}+\delta_{n_2}+\delta_{k_2} > \delta_{l_1}+\delta_{m_1}$ for every $0<p<\frac{1}{2}$. Thus, the step function suggested in (64) can change the result of the minimization over $\mathcal{D}_2$ without changing the result of the minimization over $\mathcal{D}_1$. The threshold value $M(p)$ is due to the fact that the proposed step function cannot unlimitedly increase the result of the minimization over $\mathcal{D}_2$. For more intuition on the choice of $E_\eta^{\delta_d}$, remember that for any received word $\mathbf{x}$, the optimal value of η_i is $1/deg(\mathbf{x})$. Since $deg(\mathbf{x})$ is a non decreasing function of $w(\mathbf{x})$, and the size of any coset is 2^{RN}, when $R\neq 0$ we expect $deg(\mathbf{x})$ to grow exponentially with N, at least when $w(\mathbf{x})=N$ (in this case the exponent is exactly R). Thus, for a reasonable choice of η_i, there exist $\delta_{t_0} \leq 1$ such that for any $\delta_t \geq \delta_{t_0}$ we have $E_\eta^{\delta_i}(\delta_t) > 0$ and $E_\eta^{\delta_i}(\delta_t) = 0$ otherwise. It is clear that the function $E_\eta^{\delta_d}$ suggested in (64) answers to this restraint.

To conclude this discussion, we return to the general case of the sub-code $\mathcal{C}_i^*$. As explained earlier, the equations required here are cubic, with cumbersome coefficients. Yet, a closed form solution for these equations exists, and is easily handled using Matlab's symbolic toolbox. We can follow the derivations above (and the proof in Appendix C.2) step by step and find out that the inequality $\delta_{l_2} + \delta_{m_2} + \delta_{n_2} + \delta_{k_2} > \delta_i/2 + p(1-\delta_i)$ is still valid, hence Proposition 5.3 stands solid for any sub-code $\mathcal{C}_i^*$, not necessarily $\mathcal{C}_d^*$. Thus, to derive the tightest bound on the error exponent, one can optimize the bound over all possible *sub-codes*, as long as the union bound analysis applies. For example, in [12] Barg and Forney compute the error exponent for random and typical codes. Their derivations are based on the fact that union bound analysis applies for these codes, i.e., optimization on the sub-code is implicitly used. However, even when only the sub-code $\mathcal{C}_d^*$ is used, the new bounds given in this work may be interesting. For example, consider the recently discovered family of binary linear codes with exponentially many minimum distance codewords [26]. For these codes, the union bound analysis allows us to subtract the rate of the minimum distance codewords from the two codewords bound, resulting in a tighter bound on the error exponent. Widening the range of the union bound analysis is, in this case, beneficial.

As for the AWGN channel and the bound derived in Section 3, we may follow the steps in this section directly, though, it is important to note, we do not seek the optimal optimization function, only the optimal value of the parameters. The results, thus, are not as sharp as these for the BSC. Yet, it is possible to show that the new bound on the error probability given in Corollary 3.2 results in a tighter bound on the error exponent than the de Caen-based analogue [25].

6 Results

In this section, several examples with well known codes are given and the results of the numerical analysis are shown.

6.1 AWGN channel

Before the numerical results for the lower bounds are introduced, we address several computational issues. First, the definition of $Q(\cdot)$ as given in (16) requires an integration over an infinite set. Instead, an alternative form by Craig[8] [28] was used

$$Q(x) = \frac{1}{\pi}\int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta \qquad x \geq 0.$$

[8] also appearing in [27], with a simpler proof.

As for $\Psi(\cdot,\cdot,\cdot)$, an expression given by Simon and Divsalar in [29] was used

$$\Psi(\rho,x,y) = \frac{1}{2\pi}\int_0^{\pi/2-\tan^{-1}(y/x)} \frac{\sqrt{1-\rho^2}}{1-\rho\sin 2\theta}\exp\left\{-\frac{x^2}{2}\frac{1-\rho\sin 2\theta}{(1-\rho^2)\sin^2\theta}\right\}d\theta + \frac{1}{2\pi}\int_0^{\tan^{-1}(y/x)} \frac{\sqrt{1-\rho^2}}{1-\rho\sin 2\theta}\exp\left\{-\frac{y^2}{2}\frac{1-\rho\sin 2\theta}{(1-\rho^2)\sin^2\theta}\right\}d\theta.$$

We compare the new lower bounds for linear codes, presented in Section 3, with several known bounds in the current literature. For the sake of simplicity, only three new bounds are discussed. The first is the *norm bound - whole code*, i.e., the bound given in (24) with $a = c = a'$ and $b = -2a'$. The second is the *dot product bound - sub-code $\mathcal{C}_d^*$*, i.e., the bound given in (26) with $a = c = 0$ and $b = -a'$. The third bound is Kounias' lower bound as given in (29). The new bounds are compared to Seguin's lower bound [13], Shannon's lower bound [5] and Poltyrev's upper bound [1][9]. The results for the codes BCH(63,24) and Golay(23,12) are given in Figures 1 and 2, respectively. For the sake of clarity, Figure 2 does not include Kounias' bound. It is only slightly superior to Seguin's.

It is clear that the new bounds perform better than Seguin's for any value of E_b/N_0. This can be seen both for the bound using the whole code (as in Seguin's bound) and for the bound using only the sub-code $\mathcal{C}_d^*$. To the authors' knowledge, for high values of E_b/N_0, where the new bounds are superior to Shannon's lower bound, they establish the best know results in the current literature.

Consider the limiting cases of $\frac{E_b}{N_0} \to \infty$ and $\frac{E_b}{N_0} \to 0$. While non-trivial values of the parameters a, b and c yield strictly tighter bounds for intermediate values of $\frac{E_b}{N_0}$, it is not so in these cases. When $\frac{E_b}{N_0} \to \infty$, Seguin's bound is optimal ([13, Section 5]), in the sense that the ratio with the union bound tends to unity. Therefore, no non-trivial values of the parameters a, b and c yield tighter results. Note that, however, the rate of convergence may be faster with non-trivial parameters. This is also the case when $\frac{E_b}{N_0} \to 0$, though to see this, unwieldy limit computations are required. The behavior in these limiting cases is evident in the graphs, although the case where $\frac{E_b}{N_0} \to 0$ is most apparent when $\frac{E_b}{N_0} < -3dB$, a range not included in the presented graphs.

6.2 BSC

We compare the new lower bound given in Proposition 4.2 with several known bounds in the current literature, when the code used is $BCH(63,24)$. Figure 3 includes the new bound with $i = d$ and the approximation (50). The new bound with the trivial choice of η_i is not plotted since the results are very similar to Keren and Litsyn's. For reference, three bounds are plotted: Poltyrev's upper bound [1], Keren and Litsyn's lower bound [14] and the sphere packing lower bound. Referring to Keren and Litsyn's bound [14], the bound presented in [14] is based on the same techniques, namely, de Caen's bound [15], the subset $\mathcal{C}_d^*$ is used and words with weight higher than the covering radius are considered erroneous. However, even when $\eta_i(w) \equiv 1$, the bound in Proposition 4.2 is not identical to [14]. The major difference is the fact that in [14], the set $\{\mathbf{x} \in GF(2)^N : \exists_i \quad w(\mathbf{x}+\mathbf{c}_i) < w(\mathbf{x}), w(\mathbf{x}) \leq t\}$ is

[9] As noted in [30], the upper bound given in [1] does not take into account a subset of the space (the lower half of the cone) when upper bounding the error probability. Indeed, this part has negligible probability, yet, it is necessary to obtain a rigorous bound. However, following the derivations in [30], it is easy to verify the validity of the bound for the codes tested in this work.

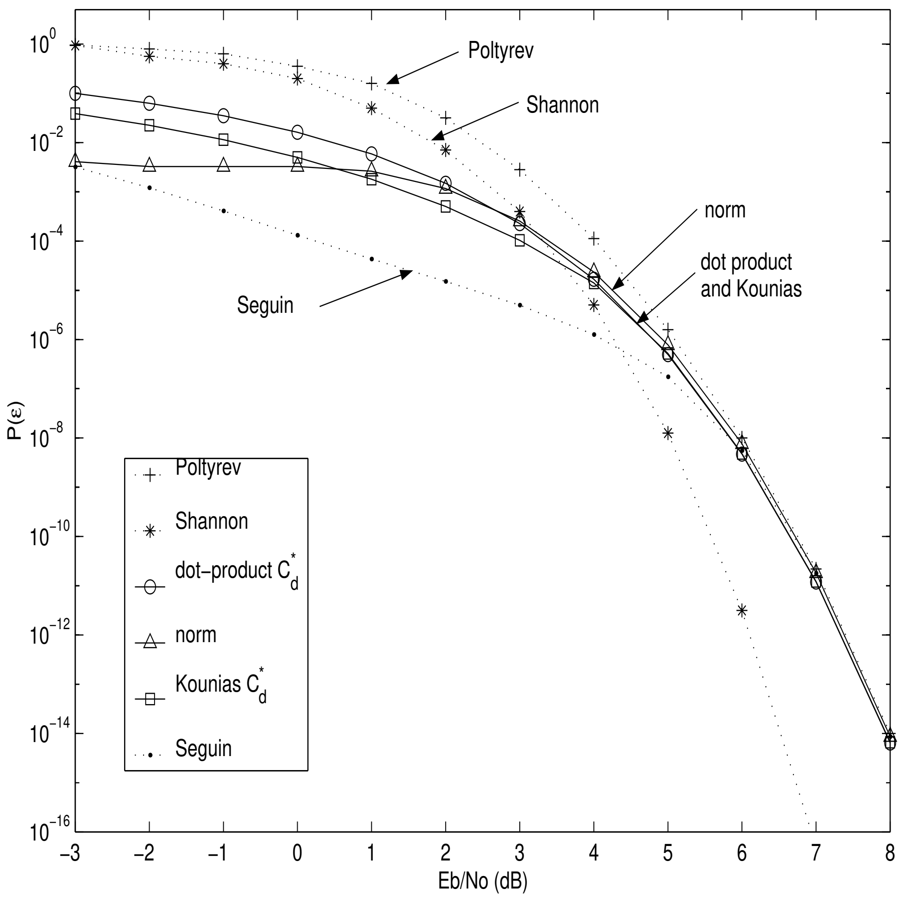

Figure 1: Bounds on the decoding error probability of BCH(63,24), AWGN channel. The new lower bounds *norm - whole code*, *dot product - sub-code* $\mathcal{C}_d^*$ and Kounias' are shown. For reference, Poltyrev's upper bound and Shannon's and Seguin's lower bounds are given.

partitioned to constant weight subsets and de Caen's bound is employed to each subset separately. This partition simplifies several computations and instead of Proposition 4.1 a more ad hoc approach can be used. However, in this way, the usage of Theorem 2.1 instead of de Caen's bound is burdensome. It is also important to note that the bound in [14] is easier to evaluate since the summations required are simpler.

The new bound is at least as good as Keren and Litsyn's bound for every value of p. The improvement is obvious for high values of p, however, for lower, and more realistic values of p, where Keren and Litsyn's bound is superior to the sphere packing bound, the improvement is scarce. Nevertheless, it is clear that a non trivial choice of the optimizing function η_i yields better results.

Finally, we compare the new bound on the error exponent derived in Section 5 with the de Caen-based analogue. As a simple example, we use the codes derived by Ashikhmin Barg and Vlăduţ [26]. Since this family of codes has exponentially many minimum distance codewords, it can be used as an example for the bound in Proposition 5.3, namely, when the sub-code $\mathcal{C}_d^*$ is used. Figure 4 includes the results. The two upmost curves are the discussed bound, with trivial $E_\eta^{\delta_d}$ above and non trivial $E_\eta^{\delta_d}$ below. The horizontal line is the value of $E_B^{\delta_d}$. The lowermost curve is the condition on the code

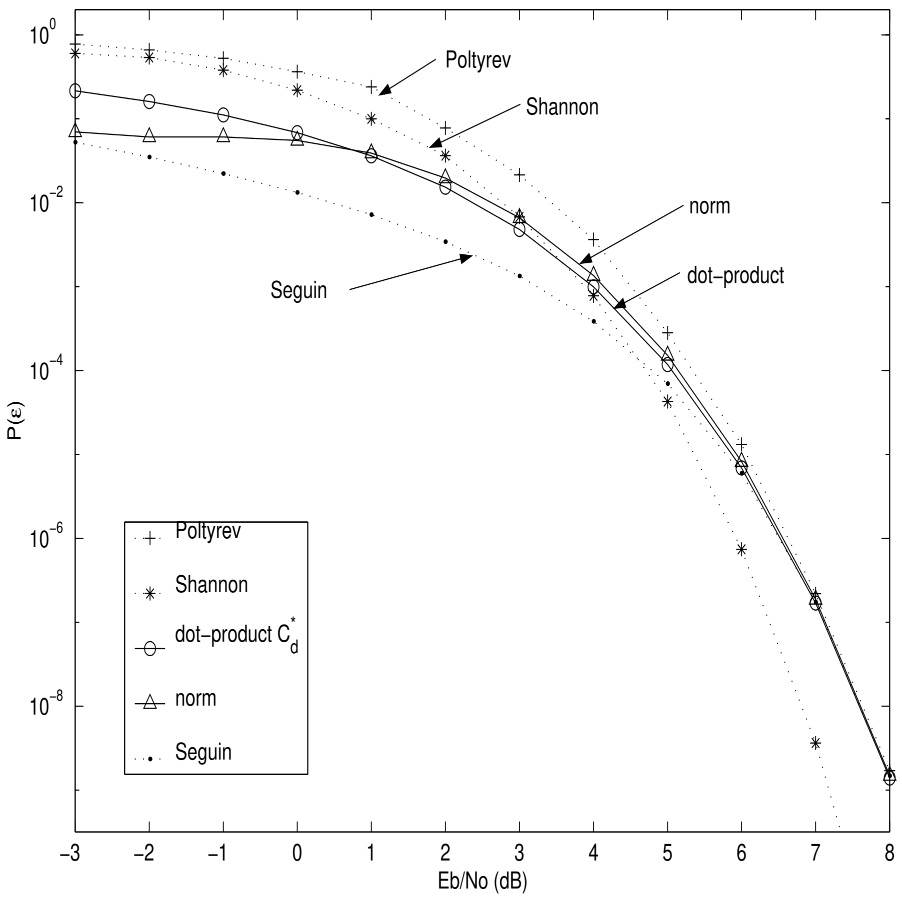

Figure 2: Bounds on the decoding error probability of Golay(23,12), AWGN channel. The new lower bounds *norm - whole code* and *dot product - sub-code* $\mathcal{C}_d^*$ are shown. For reference, Poltyrev's upper bound and Shannon's and Seguin's lower bounds are given.

for this case. It is clear that for values of p for which the condition is not satisfied, non trivial $E_B^{\delta_d}$ tightens the bound. It is also clear that the improvement is achieved by continuing the usage of union bound analysis, until a certain threshold is exceeded. From this point on, the union bound analysis does not apply, yet the bound with non trivial $E_\eta^{\delta_d}$ is still tighter.

7 Discussion

In this paper, new lower bounds on the error probability of a given block code were derived. In the first part of the paper, a new lower bound on the probability of a union of events was introduced. As explained therein, the bound improves on de Caen's inequality by having the ability to optimize the result over a wide family of functions. Moreover, the optimal function is known, though not always mathematically endurable, thus may act as a guiding light in the optimization process. This lower bound was used as a framework for deriving the new bounds on the error probability. It was shown that these bounds are tighter than the best known bounds in the current literature for finite block length and low values of noise. Hence, the new bound on the probability of a union gives a powerful

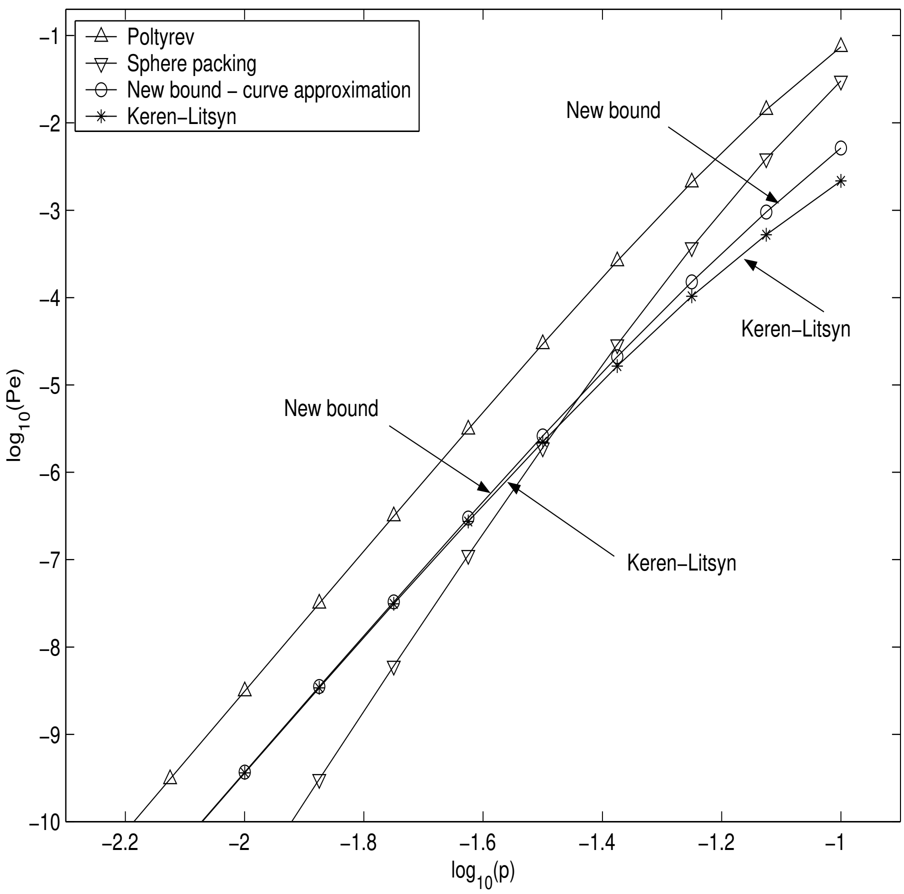

Figure 3: Bounds on the decoding error probability of BCH(63,24), $M = 15$, BSC. The new bound, based on the approximation given in (44) is given. For reference, Poltyrev's upper bound, Keren and Litsyn's lower bound and the sphere packing lower bound are given.

framework for deriving lower bounds on the error probability.

As for future work, note that the bounds on the error exponent, derived in Section 5, are applicable only for specific codes, with known distance distribution. To derive upper bound on the reliability function of the BSC, the conditions for union bound analysis given in this paper can be used, together with known or new bounds on the distance distribution of binary (or binary linear) codes. In this case, future work may refer to the bounds and techniques appearing in the works of Litsyn [10] and Burnashev [11]. Finally, we note that since the new bound on the probability of a union suggests a framework for deriving bounds on the error probability, new bounds can be derived for different channel models. Moreover, the proposed bounds may be improved by seeking new families of functions for optimization.

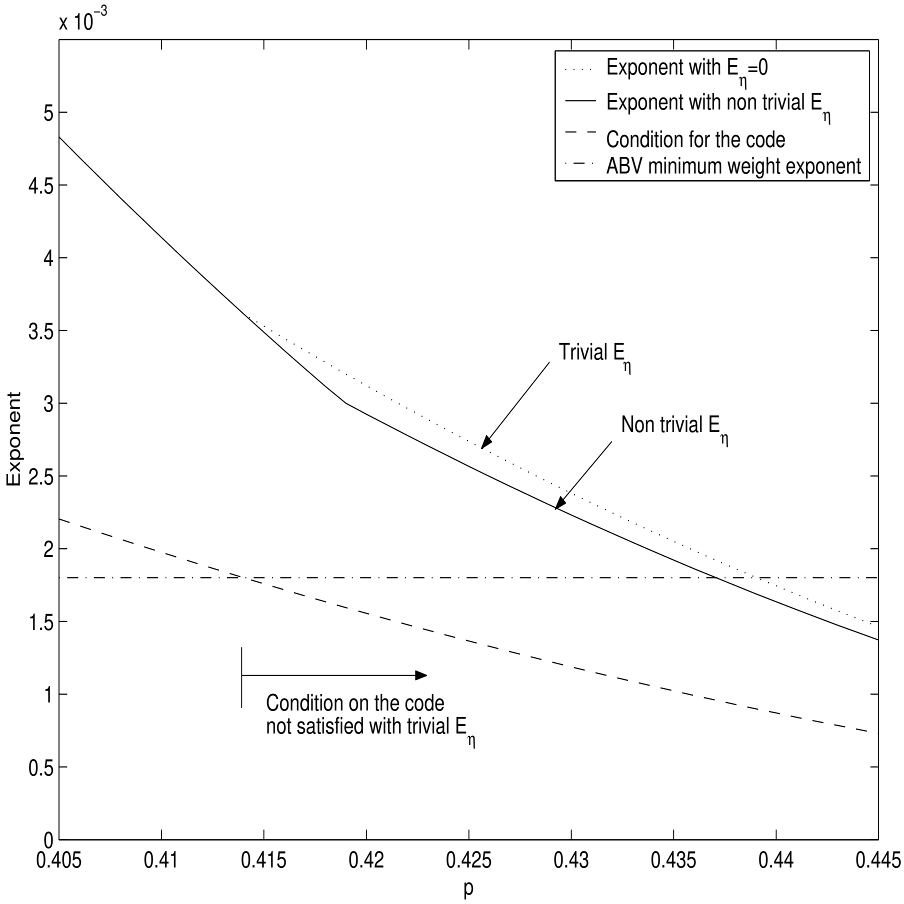

Figure 4: Bounds on the error exponent, BSC, ABV codes and the sub-code $\mathcal{C}_d^*$. The two upmost curves are the discussed bound, with trivial $E_\eta^{\delta_d}$ above and non trivial $E_\eta^{\delta_d}$ below. The horizontal line is the value of $E_B^{\delta_d}$. The lowermost curve is the condition on the code.

A Computation of the Integrals Required for Proposition 3.1

We first compute the integral

$$\int_{\varepsilon_{0i}} p(\mathbf{r}|\mathbf{s}_0)m(\mathbf{r}|\mathbf{s}_0)d\mathbf{r},$$

where ε_{0i}, $p(\mathbf{r}|\mathbf{s}_0)$ and $m(\mathbf{r}|\mathbf{s}_0)$ were defined in (6), (9) and (14) respectively. We have

$$\begin{aligned}
&\int_{\varepsilon_{0i}} p(\mathbf{r}|\mathbf{s}_0)m(\mathbf{r}|\mathbf{s}_0)d\mathbf{r} \\
&= \int_{\varepsilon_{0i}} (\pi N_0)^{-\frac{K}{2}} \exp\left\{-\frac{1}{N_0}||\mathbf{r}-\mathbf{s}_0||^2\right\} \exp\left\{-(a||\mathbf{r}||^2 + b\langle\mathbf{r},\mathbf{s}_0\rangle + c||\mathbf{s}_0||^2)\right\}d\mathbf{r} \\
&= \exp\{-\beta||\mathbf{s}_0||^2\}\left(\frac{N_0'}{N_0}\right)^{\frac{K}{2}} \int_{\varepsilon_{0i}} (\pi N_0')^{-\frac{K}{2}} \exp\left\{-\frac{1}{N_0'}\left\|\mathbf{r}-\alpha\mathbf{s}_0\right\|^2\right\}d\mathbf{r} \\
&= \exp\{-\beta||\mathbf{s}_0||^2\}\left(\frac{N_0'}{N_0}\right)^{\frac{K}{2}} P\left(\{\mathbf{r}\in\mathbb{R}^K : ||\mathbf{r}-\mathbf{s}_i|| < ||\mathbf{r}-\mathbf{s}_0||\} | \mathbf{r} = \tilde{\mathbf{n}} + \alpha\mathbf{s}_0\right) \\
&= \exp\{-\beta||\mathbf{s}_0||^2\}\left(\frac{N_0'}{N_0}\right)^{\frac{K}{2}} P\left(\frac{\langle\tilde{\mathbf{n}},\mathbf{s}_0-\mathbf{s}_i\rangle}{\sqrt{\frac{N_0'}{2}}||\mathbf{s}_0-\mathbf{s}_i||} < \frac{(\alpha-1)^2||\mathbf{s}_0||^2 - ||\alpha\mathbf{s}_0-\mathbf{s}_i||^2}{\sqrt{2N_0'}||\mathbf{s}_0-\mathbf{s}_i||}\right),
\end{aligned}$$

where N_0', α and β are defined by

$$\begin{aligned} N_0' &= \frac{N_0}{1+aN_0}, \qquad a \neq -\frac{1}{N_0}, \\ \alpha &= \left(\frac{\frac{1}{N_0}-\frac{b}{2}}{a+\frac{1}{N_0}}\right), \\ \beta &= \frac{\left(\frac{1}{N_0}+a\right)\left(\frac{1}{N_0}+c\right)-\left(\frac{1}{N_0}-\frac{b}{2}\right)^2}{\frac{1}{N_0}+a}, \end{aligned}$$

$\tilde{\mathbf{n}}$ is a K-dimensional vector of i.i.d. $\mathcal{N}\left(0, \frac{N_0'}{2}\right)$ random variables and we assume $N_0' > 0$, i.e., $a > -\frac{1}{N_0}$. Finally, since

$$X_i' \triangleq \frac{\langle \tilde{\mathbf{n}}, \mathbf{s}_0 - \mathbf{s}_i \rangle}{\sqrt{\frac{N_0'}{2}}\|\mathbf{s}_0-\mathbf{s}_i\|}$$

is an $\mathcal{N}(0,1)$ random variable, we have

$$\begin{aligned} &\int_{\varepsilon_{0i}} p(\mathbf{r}|\mathbf{s}_0) m(\mathbf{r}|\mathbf{s}_0) d\mathbf{r} \\ &\quad = \exp\left\{-\beta\|\mathbf{s}_0\|^2\right\}\left(\frac{N_0'}{N_0}\right)^{\frac{K}{2}} Q\left(\frac{\|\alpha\mathbf{s}_0-\mathbf{s}_i\|^2-(\alpha-1)^2\|\mathbf{s}_0\|^2}{\sqrt{2N_0'}\|\mathbf{s}_0-\mathbf{s}_i\|}\right), \end{aligned}$$

where $Q(\cdot)$ is the error function defined in (16).

As for the integral in (8), analogously to the preceding derivations, we have

$$\begin{aligned} &\int_{\varepsilon_{0i}\cap\varepsilon_{0j}} p(\mathbf{r}|\mathbf{s}_0) m^2(\mathbf{r}|\mathbf{s}_0) d\mathbf{r} \\ &= \exp\left\{-\beta'\|\mathbf{s}_0\|^2\right\}\left(\frac{N_0''}{N_0}\right)^{\frac{K}{2}} \int_{\varepsilon_{0i}\cap\varepsilon_{0j}} (\pi N_0'')^{-\frac{K}{2}} \exp\left\{-\frac{1}{N_0''}\left\|\mathbf{r}-\alpha'\mathbf{s}_0\right\|^2\right\} d\mathbf{r} \\ &= \exp\left\{-\beta'\|\mathbf{s}_0\|^2\right\}\left(\frac{N_0''}{N_0}\right)^{\frac{K}{2}} P\left(\{\mathbf{r}\in\mathbb{R}^K : \|\mathbf{r}-\mathbf{s}_i\| < \|\mathbf{r}-\mathbf{s}_0\|, \|\mathbf{r}-\mathbf{s}_j\| < \|\mathbf{r}-\mathbf{s}_0\|\} | \mathbf{r} = \hat{\mathbf{n}} + \alpha'\mathbf{s}_0\right) \\ &= \exp\left\{-\beta'\|\mathbf{s}_0\|^2\right\}\left(\frac{N_0''}{N_0}\right)^{\frac{K}{2}} P\left(X_i'' < \frac{(\alpha'-1)^2\|\mathbf{s}_0\|^2-\|\alpha'\mathbf{s}_0-\mathbf{s}_i\|^2}{\sqrt{2N_0''}\|\mathbf{s}_0-\mathbf{s}_i\|}, X_j'' < \frac{(\alpha'-1)^2\|\mathbf{s}_0\|^2-\|\alpha'\mathbf{s}_0-\mathbf{s}_j\|^2}{\sqrt{2N_0''}\|\mathbf{s}_0-\mathbf{s}_j\|}\right), \end{aligned}$$

where now

$$N_0'' = \frac{N_0}{1+2aN_0}, \qquad a \neq -\frac{1}{2N_0}, \tag{66}$$

$$\alpha' = \left(\frac{\frac{1}{N_0}-b}{2a+\frac{1}{N_0}}\right), \tag{67}$$

$$\beta' = \frac{\left(\frac{1}{N_0}+2a\right)\left(\frac{1}{N_0}+2c\right)-\left(\frac{1}{N_0}-b\right)^2}{\frac{1}{N_0}+2a},$$

we assuming $N_0'' > 0$, i.e., $a > -\frac{1}{2N_0}$, $\hat{\mathbf{n}}$ is a K-dimensional vector of i.i.d. $\mathcal{N}\left(0, \frac{N_0''}{2}\right)$ random variables and

$$X_i'' \triangleq \frac{\langle \hat{\mathbf{n}}, \mathbf{s}_0 - \mathbf{s}_i \rangle}{\sqrt{\frac{N_0''}{2}}\|\mathbf{s}_0-\mathbf{s}_i\|}$$

is an $\mathcal{N}(0,1)$ random variable. It is easy to verify that

$$\mathbb{E}\{X_i'' X_j''\} = \frac{\langle \mathbf{s}_i-\mathbf{s}_0, \mathbf{s}_j-\mathbf{s}_0\rangle}{\|\mathbf{s}_i-\mathbf{s}_0\|\|\mathbf{s}_j-\mathbf{s}_0\|} = \rho_{ij},$$

where ρ_{ij} was defined in (18). Consequently,

$$\int_{\varepsilon_{0i}\cap\varepsilon_{0j}} p(\mathbf{r}|\mathbf{s}_0)m^2(\mathbf{r}|\mathbf{s}_0)d\mathbf{r} = \exp\left\{-\beta'\|\mathbf{s}_0\|^2\right\}\left(\frac{N_0''}{N_0}\right)^{\frac{K}{2}}$$
$$\cdot\,\Psi\left(\rho_{ij}, \frac{\|\alpha'\mathbf{s}_0-\mathbf{s}_i\|^2-(\alpha'-1)^2\|\mathbf{s}_0\|^2}{\sqrt{2N_0''}\|\mathbf{s}_0-\mathbf{s}_i\|}, \frac{\|\alpha'\mathbf{s}_0-\mathbf{s}_j\|^2-(\alpha'-1)^2\|\mathbf{s}_0\|^2}{\sqrt{2N_0''}\|\mathbf{s}_0-\mathbf{s}_j\|}\right),$$

where $\Psi(\cdot,\cdot,\cdot)$ is the bivariate normal distribution defined in (17).

B Proofs of Propositions 4.1 and 4.3

B.1 Proof of Proposition 4.1

The proof is as follows. We examine a simpler expression than $\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j)$, in which the sum is only over words $\mathbf{x}\in\varepsilon_{0i}\cap\varepsilon_{0j}, i\neq j$ with constant weight $w(\mathbf{x})=u$, i.e.,

$$\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j,u)\triangleq\sum_{l=0}^{w(\mathbf{c}_i\mathbf{c}_j)}\ \sum_{m=\lfloor\frac{w(\mathbf{c}_i)}{2}\rfloor+1-l}^{w(\mathbf{c}_i)-w(\mathbf{c}_i\mathbf{c}_j)}\ \sum_{n=\lfloor\frac{w(\mathbf{c}_j)}{2}\rfloor+1-l}^{w(\mathbf{c}_j)-w(\mathbf{c}_i\mathbf{c}_j)}\binom{w(\mathbf{c}_i\mathbf{c}_j)}{l}\binom{w(\mathbf{c}_i)-w(\mathbf{c}_i\mathbf{c}_j)}{m}$$
$$\cdot\binom{w(\mathbf{c}_j)-w(\mathbf{c}_i\mathbf{c}_j)}{n}\binom{N-w(\mathbf{c}_i)-w(\mathbf{c}_j)+w(\mathbf{c}_i\mathbf{c}_j)}{u-l-m-n}p^u(1-p)^{N-u}\eta_i(u). \quad (68)$$

Since $\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j)=\sum_u\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j,u)$, if $\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j,u)$ is monotonically increasing in $w(\mathbf{c}_i\mathbf{c}_j)$ the proposition is proved. Observe that $p^u(1-p)^{N-u}\eta_i(u)\geq 0$ does not affect the behavior of $\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j,u)$ as a function of $w(\mathbf{c}_i\mathbf{c}_j)$. Hence, it is enough to prove that $\widetilde{P}_{den}(\mathbf{c}_i,\mathbf{c}_j,u)/(p^u(1-p)^{N-u}\eta_i(u))$ is monotonic in $w(\mathbf{c}_i\mathbf{c}_j)$. This expression is, however, simply the number of words in a sub set of $GF(2)^N$, which we denote by $V_{ij}(u,w)$, where $w=w(\mathbf{c}_i\mathbf{c}_j)$. To examine the behavior of $|V_{ij}(u,w)|$ as a function of w, we assume that one codeword is fixed, without loss of generality $\mathbf{c}_i$, and instead of $\mathbf{c}_j$ introduce a dummy codeword $\mathbf{c}_{j'}$, satisfying $w(\mathbf{c}_{j'})=w(\mathbf{c}_j)$ and $w(\mathbf{c}_i\mathbf{c}_{j'})=w+1$. Thus, the only difference in (68) is $w(\mathbf{c}_i\mathbf{c}_{j'})$ instead of $w(\mathbf{c}_i\mathbf{c}_j)$. Let $V_{ij}^c(u,w)=GF(2)^N\setminus V_{ij}(u,w)$. We wish to prove that

$$|V_{ij'}(u,w+1)|-|V_{ij}(u,w)|=|V_{ij'}(u,w+1)\cap V_{ij}^c(u,w)|-|V_{ij}(u,w)\cap V_{ij'}^c(u,w+1)|\geq 0$$

for any w, u, $w(\mathbf{c}_i)$ and $w(\mathbf{c}_j)$.

Consider the set $V_{ij'}(u,w+1)\cap V_{ij}^c(u,w)$. To count the number of words in this set, we examine the example in Figure 5. For the sake of simplicity, we group the 1's of each codeword together. Clearly, the *size* of the considered set is invariant under this permutation. Let $\mathbf{x}$ be a word in this set. First, $\mathbf{x}$ must satisfy $w(\mathbf{x})=u$. Second, $w(\mathbf{x}_{\mathcal{S}_i})\geq\lfloor\frac{w(\mathbf{c}_i)}{2}\rfloor+1$, since $\mathbf{x}\in\varepsilon_{0i}$. However, in order for $\mathbf{x}$ to satisfy $\mathbf{x}\in V_{ij'}(u,w+1)$ but $\mathbf{x}\notin V_{ij}(u,w)$, only due to a shift of one bit (from $w(\mathbf{c}_i\mathbf{c}_j)=w$ to $w(\mathbf{c}_i\mathbf{c}_{j'})=w+1$), we must have

$$w(\mathbf{x}_{\mathcal{S}_j})=\left\lfloor\frac{w(\mathbf{c}_j)}{2}\right\rfloor \qquad \text{and} \qquad w(\mathbf{x}_{\mathcal{S}_{j'}})=\left\lfloor\frac{w(\mathbf{c}_j)}{2}\right\rfloor+1.$$

$\mathbf{c}_i$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1											
$\mathbf{c}_j$												1	1	1	1	1	1	1	1	1						
$\mathbf{c}_{j'}$											1	1	1	1	1	1	1	1	1							

$\mathbf{x}$	$\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor - l \leq m \leq w(\mathbf{c}_i) - w - 1$	1	$0 \leq l \leq w$	$\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor - l$	0	$u - \lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor - m - 1$

Figure 5: Example of $\mathbf{c}_i$, $\mathbf{c}_j$, $\mathbf{c}_{j'}$ and $\mathbf{x}$. The word $\mathbf{x}$ is divided to six parts, each one includes the possible number of 1's. The inequalities for m and l define the summation bounds in (69).

Therefore, $\mathbf{x}_{\mathcal{S}_j \setminus \mathcal{S}_{j'}} = 0$, $\mathbf{x}_{\mathcal{S}_{j'} \setminus \mathcal{S}_j} = 1$ and $w(\mathbf{x}_{\mathcal{S}_j \cap \mathcal{S}_{j'}}) = \lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor$. Accordingly, we have

$$\begin{aligned}
&|V_{ij'}(u, w+1) \cap V_{ij}^c(u, w)| \\
&= \sum_{l=0}^{w} \binom{w}{l} \sum_{m=\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor - l}^{w(\mathbf{c}_i)-w-1} \binom{w(\mathbf{c}_i) - w - 1}{m} \binom{w(\mathbf{c}_j) - w - 1}{\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor - l} \binom{N - w(\mathbf{c}_i) - w(\mathbf{c}_j) + w}{u - m - 1 - \lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor}. \qquad (69)
\end{aligned}$$

To evaluate $V_{ij}(u, w) \cap V_{ij'}^c(u, w+1)$, note that now $\mathbf{x}_{\mathcal{S}_j \setminus \mathcal{S}_{j'}} = 1$ and $\mathbf{x}_{\mathcal{S}_{j'} \setminus \mathcal{S}_j} = 0$. Thus,

$$\begin{aligned}
&|V_{ij}(u, w) \cap V_{ij'}^c(u, w+1)| \\
&= \sum_{l=0}^{w} \binom{w}{l} \sum_{m=\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor - l + 1}^{w(\mathbf{c}_i)-w-1} \cdot \binom{w(\mathbf{c}_i) - w - 1}{m} \binom{w(\mathbf{c}_j) - w - 1}{\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor - l} \binom{N - w(\mathbf{c}_i) - w(\mathbf{c}_j) + w}{u - m - 1 - \lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&|V_{ij'}(u, w+1) \cap V_{ij}^c(u, w)| - |V_{ij}(u, w) \cap V_{ij'}^c(u, w+1)| \\
&\qquad = \sum_{l=0}^{w} \binom{w}{l} \cdot \left[\binom{w(\mathbf{c}_i) - w - 1}{\lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor - l} \binom{w(\mathbf{c}_j) - w - 1}{\lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor - l} \binom{N - w(\mathbf{c}_i) - w(\mathbf{c}_j) + w}{u - \lfloor \frac{w(\mathbf{c}_i)}{2} \rfloor - \lfloor \frac{w(\mathbf{c}_j)}{2} \rfloor + l - 1} \right],
\end{aligned}$$

an expression which is clearly non negative for any $0 \leq w \leq \min\{w(\mathbf{c}_i) - 1, w(\mathbf{c}_j) - 1\}$, $w(\mathbf{c}_i)$, $w(\mathbf{c}_j)$, and u.

B.2 Proof of Proposition 4.3

The proof is as follows. First, taking $M = N$ only weakens the bound, since (45) is a monotonic decreasing function of M. We have

$$P(\varepsilon) \geq LB(p) = \frac{B_d P_{num}^2(d)}{P_{den}(d) + (B_d - 1)P_{den}(d, d)},$$

where p is the crossover probability of the channel appearing in the definitions of $P_{num}(d)$, $P_{den}(d)$ and $P_{den}(d, d)$. The proof is analogous to Seguin's proof, as it appears in [13]. As an upper bound we use the union bound given in (51). We have

$$\begin{aligned}
\frac{LB(p)}{UB(p)} &= \frac{B_d P_{num}^2(d)}{\left(\sum_{i=1}^{N} B_i P(i)\right) (P_{den}(d) + (B_d - 1)P_{den}(d, d))} \\
&= \frac{B_d}{\left(\sum_{i=1}^{N} B_i \frac{P(i)}{P_{num}(d)}\right) \left(\frac{P_{den}(d)}{P_{num}(d)} + (B_d - 1)\frac{P_{den}(d,d)}{P_{num}(d)}\right)}.
\end{aligned}$$

Observe that both $P(i)$ and $P_{num}(d)$ go to 0 as p goes to 0, hence we may apply l'Hopital's rule until one of them is a non zero constant. Since the expression with the lowest power of p is the first to yield a non zero constant after successive differentiations, we have

$$\lim_{p\to 0}\frac{P(i)}{P_{num}(d)} = \begin{cases} 0 & i > d \\ \frac{1}{\eta_i(\lfloor\frac{d}{2}\rfloor+1)} & i = d \\ \infty & i < d, \end{cases}$$

thus,

$$\lim_{p\to 0}\sum_{i=1}^{N} B_i\frac{P(i)}{P_{num}(d)} = \frac{B_d}{\eta_i(\lfloor\frac{d}{2}\rfloor+1)}.$$

Using the same method we have

$$\lim_{p\to 0}\frac{P_{den}(d)}{P_{num}(d)} = \eta_i\left(\left\lfloor\frac{d}{2}\right\rfloor+1\right)$$

and

$$\lim_{p\to 0}\frac{P_{den}(d,d)}{P_{num}(d)} = 0,$$

therefore,

$$\lim_{p\to 0}\frac{LB(p)}{UB(p)} = 1.$$

C Proofs of Propositions 5.1 and 5.3

C.1 Proof of Proposition 5.1

The proof is as follows. We wish to determine the exponential rate (as $N\to\infty$) of the r.h.s. of (52). First, consider $P_{num}(i)$. Remembering that

$$\binom{N}{k} \doteq 2^{NH(\frac{k}{N})},$$

we have

$$\lim_{N\to\infty} -\frac{1}{N}\log\left(\binom{i}{l}\binom{N-i}{m}p^{l+m}(1-p)^{N-l-m}\eta_i(l+m)\right) =$$
$$-\delta_i H\left(\frac{\delta_l}{\delta_i}\right) - (1-\delta_i)H\left(\frac{\delta_m}{1-\delta_i}\right) - (\delta_l+\delta_m)\log(p) - (1-\delta_l-\delta_m)\log(1-p) + E_\eta^{\delta_i}(\delta_l+\delta_m)$$
$$= E_1^{\delta_i}(\delta_l,\delta_m,p) + E_\eta^{\delta_i}(\delta_l+\delta_m).$$

Thus, since the exponential rate is determined by the summand with the maximal exponent, we have

$$\lim_{N\to\infty} -\frac{1}{N}\log P_{num}(i) = \min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l,\delta_m,p) + E_\eta^{\delta_i}(\delta_l+\delta_m)\right\}.$$

Note that since $E_1^{\delta_i}(\delta_l,\delta_m,p)$ is continuous and $E_\eta^{\delta_i}(\delta_l+\delta_m)$ is piecewise continuous, for large enough N the minimum can be taken over $\mathcal{D}_1$, a continuous interval, ignoring the requirements for rational values of δ_l and δ_m. The requirements for integer values in the summation bounds of (35) were also relaxed for the same reason. The same applies for $P_{den}(i)$ as well, thus, we have

$$\lim_{N\to\infty} -\frac{1}{N}\log P_{den}(i) = \min_{\mathcal{D}_1}\left\{E_1^{\delta_i}(\delta_l,\delta_m,p) + 2E_\eta^{\delta_i}(\delta_l+\delta_m)\right\}.$$

As for $P_{den}(i,i)$, using the same arguments, we have

$$\lim_{N\to\infty} -\frac{1}{N}\log P_{den}(i,i) = \min_{\mathcal{D}_2}\left\{E_2^{\delta_i}(\delta_l,\delta_m,\delta_n,\delta_k,p) + 2E_\eta^{\delta_i}(\delta_l+\delta_m+\delta_n+\delta_k)\right\}.$$

To conclude, observe that when considering the exponent of the sum $P_{den}(i) + (B_i - 1)P_{den}(i,i)$, we distinguish between two cases. The first is when

$$\lim_{N\to\infty} -\frac{1}{N}\log P_{den}(i) \leq \lim_{N\to\infty} -\frac{1}{N}\log\left((B_i-1)P_{den}(i,i)\right),$$

namely, when condition (59) is satisfied, $P_{den}(i)$ dominates $(B_i - 1)P_{den}(i,i)$ and we have (58). The second is when condition (59) is not satisfied, $(B_i - 1)P_{den}(i,i)$ dominates $P_{den}(i)$ and we have (60).

C.2 Proof of Proposition 5.3

The proof is as follows. First, we show that indeed $M(p) > 0$ for every $0 < p < \frac{1}{2}$. Observe that, for every $0 < p < \frac{1}{2}$, we have

$$\delta_{l_2} + \delta_{m_2} + \delta_{n_2} + \delta_{k_2} > \delta_{l_1} + \delta_{m_1}, \tag{70}$$

with equality in (70) only for $p = 0, \frac{1}{2}$. Thus, by the convexity of $\mathcal{D}_2$ and the strict convexity of $E_2^{\delta_d}$, we have

$$\min_{\mathcal{D}_2\cap\mathcal{D}_2'} E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) > \min_{\mathcal{D}_2} E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) = E_2^{\delta_d}(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2},p).$$

Note that the minimum over $\mathcal{D}_2 \cap \mathcal{D}_2'$ can be calculated[10]using the Karush-Kuhn-Tucker conditions ([24]).

We may now consider the minimization problems in the r.h.s. of (59), where $\delta_i = \delta_d$, $E_\eta^{\delta_d}$ is as defined in (64) and $C(E_B^{\delta_d},p) > 0$. Since $E_\eta^{\delta_d} \geq 0$, and $E_\eta^{\delta_d}(\delta_{l_1} + \delta_{m_1}) = 0$, where $(\delta_{l_1},\delta_{m_1})$ is the minimizing point of $E_1^{\delta_d}$, it is clear that

$$\min_{\mathcal{D}_1}\left\{E_1^{\delta_d}(\delta_l,\delta_m,p) + 2E_\eta^{\delta_d}(\delta_l+\delta_m)\right\} = \min_{\mathcal{D}_1} E_1^{\delta_d}(\delta_l,\delta_m,p),$$

for every $C(E_B^{\delta_d},p) > 0$. Namely, the result of the minimization over $\mathcal{D}_1$ is unchanged. However, when considering the minimization of $E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) + 2E_\eta^{\delta_d}(\delta_l+\delta_m+\delta_n+\delta_k)$ over $\mathcal{D}_2$, the value of $C(E_B^{\delta_d},p)$ is important. Since $\delta_{l_2}+\delta_{m_2}+\delta_{n_2}+\delta_{k_2} > \delta_{l_1}+\delta_{m_1}$, the step function $E_\eta^{\delta_d}$, defined in (64), "lifts" $E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p)$ in a range which includes its minimizing point. For large enough $C(E_B^{\delta_d},p)$, i.e., $C(E_B^{\delta_d},p) > M(p)$, the new minimum must be at a new point, $(\delta_l{}',\delta_m{}',\delta_n{}',\delta_k{}') \in \mathcal{D}_2\cap\mathcal{D}_2'$, yielding

$$\begin{aligned}\min_{\mathcal{D}_2}\left\{E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) + 2E_\eta^{\delta_d}(\delta_l+\delta_m+\delta_n+\delta_k)\right\} &= \min_{\mathcal{D}_2\cap\mathcal{D}_2'} E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p)\\ &= E_2^{\delta_d}(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2},p) + M(p).\end{aligned}$$

However, for smaller values of $C(E_B^{\delta_d},p)$, $(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2})$ remains the minimizing point, yielding

$$\min_{\mathcal{D}_2}\left\{E_2^{\delta_d}(\delta_l,\delta_m,\delta_n,\delta_k,p) + 2E_\eta^{\delta_d}(\delta_l+\delta_m+\delta_n+\delta_k)\right\} = E_2^{\delta_d}(\delta_{l_2},\delta_{m_2},\delta_{n_2},\delta_{k_2},p) + C(E_B^{\delta_d},p).$$

This far, we have proved that when $0 < C(E_B^{\delta_d},p) \leq M(p)$, $E_\eta^{\delta_d}$ as defined in (64) alters the condition on the code in such a way that it is not repealed, and the union bound analysis apply. It remains to prove that this is the optimal choice when $0 < C(E_B^{\delta_d}, p) \leq M(p)$, and to quantify the improvement over the bound with the trivial $E_\eta^{\delta_d}$ for every $C(E_B^{\delta_d}, p) > 0$. To see this, Subtract the r.h.s. of (58) from the r.h.s. of (60). Requiring the result to be negative is no other than the condition on the code (59). Namely, when the condition on the code is not satisfied, and (60) is valid, the bound in (58) is tighter. Thus, in this case, the best choice of $E_\eta^{\delta_d}$ can improve the error exponent by no more than equalizing it to (58). Since this can be done by the $E_\eta^{\delta_d}$ proposed in (64), we draw the conclusion that it is the optimal choice. Another, more intuitive, explanation for this result is obtained by noticing that no tighter lower bound on the error probability, calculated on a sub-code $\mathcal{C}_d^*$, can be achieved, than the one which coincides with the union bound. The improvement over the bound with the trivial $E_\eta^{\delta_d}$ is simply the change in the value of the r.h.s. of (60) caused by our choice of $E_\eta^{\delta_d}$, which is $C(E_B^{\delta_d}, p)$ when $0 < C(E_B^{\delta_d}, p) \leq M(p)$, and $M(p)$ when $C(E_B^{\delta_d}, p) > M(p)$.

Acknowledgment

The authors wish to thank S. Litsyn and I. Sason for several interesting discussions and fruitful comments.

References

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

[2] B. Hughes, "On the error probability of signals in additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 151–155, January 1991.

[3] E. R. Berlekamp, "The technology of error correcting codes," *Proc. IEEE*, vol. 68, no. 8, pp. 564–593, May 1980.

[4] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, pp. 3–18, January 1965.

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

[6] O. Keren and S. Litsyn, "A simple lower bound on the probability of decoding error over a BSC," Unpublished notes, 2001.

[7] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Company, New York, 1977.

[8] R. J. McEliece and J. K. Omura, "An improved upper bound on the block coding error exponent for bynary-input discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 23, pp. 611–613, September 1977.

[9] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inform. Contr.*, vol. 10, pp. 65–103 (Part 1); 522–552 (Part 2), 1967.

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

[11] M. V. Burnashev, "On the relation between the code spectrum and the decoding error probability," *Probl. Inform. Transm.*, vol. 36, no. 4, pp. 285–304, 2000.

[12] A. Barg and G. D. Forney, "Random codes: minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2568–2573, September 2002.

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

[14] O. Keren and S. Litsyn, "A lower bound on the probability of decoding error over a BSC channel," *The 21st IEEE Electrical and Electronic Engineers in Israel*, pp. 271–273, 2000.

[15] D. de Caen, "A lower bound on the probability of a union," *Discr. Math.*, vol. 169, pp. 217–220, 1997.

[16] H. Kuai and G. Takahara F. Alajaji, "Tight error bounds for nonuniform signaling over AWGN channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2712–2718, November 2000.

[17] H. Kuai, F. Alajaji, and G. Takahara, "A lower bound on the probability of a finite union of events," *Discr. Math.*, vol. 215, pp. 147–158, March 2000.

[18] E. G. Kounias, "Bounds on the probability of a union, with applications," *Ann. Math. Statist.*, vol. 39, no. 6, pp. 2154–2158, 1968.

[19] D. Hunter, "An upper bound for the probability of a union," *J. Appl. Probab.*, vol. 13, pp. 597–603, 1976.

[20] A. Dembo, "Unpublished notes," Communicated by I. Sason, 2000.

[21] A.J. Viterbi and J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, Singapore, 1979.

[22] J. Galambos and I. Simonelli, *Bonferroni-type Inequalities with Applications*, Springer, 1996.

[23] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inform. Theory*, vol. 19, pp. 244–246, March 1973.

[24] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, second edition, 1999.

[25] A. Cohen, "Lower bounds on the error probability of a given block code," M.S. thesis, Technion, I.I.T., 2002.

[26] A. Ashikhmin, A. Barg, and S. Vlăduţ, "Linear codes with exponentialy many light vectors," *Journal of Combinatorial Theory*, vol. A 96, no. 2, pp. 396–399, November 2001.

[27] M.-S. Alouini and A. J. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Trans. Commun.*, vol. 47, no. 9, pp. 1324–1334, September 1999.

[28] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two dimentional signal constalletions," *IEEE MILCOM91 Conf. Rec., Boston, MA*, pp. 25.5.1–25.5.5, 1991.

[29] M. K. Simon and D. Divsalar, "Some new twists to problems involving the Gaussian probability integral," *IEEE Trans. commun.*, vol. 46, no. 2, pp. 200–210, February 1998.

[30] I. Sason and S. Shamai (Shitz), "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 24–47, January 2000.