CCIT Report #597
August 2006

Improved Sphere-Packing Bound Targeting Codes of Short to Moderate Block Lengths and Applications

Gil Wiechman Igal Sason

Technion – Israel Institute of Technology
Haifa 32000, Israel
{igillw@tx, sason@ee}.technion.ac.il

September 2, 2006

Abstract

This paper derives an improved sphere-packing (ISP) bound targeting codes of short to moderate block lengths. We first review classical results, i.e., the 1959 sphere-packing (SP59) bound of Shannon for the Gaussian channel, and the 1967 sphere-packing (SP67) bound for discrete memoryless channels. A recent improvement on the SP67 bound, as suggested by Valembois and Fossorier, is also discussed. These concepts are used for the derivation of a new bound (referred to as the ISP bound) which is uniformly tighter than the SP67 bound and its recent improved version. Under a mild condition, the ISP bound is applicable to general memoryless channels, and some of its applications are exemplified. Its tightness is studied by comparing it with bounds on the ML decoding error probability, and computer simulations of iteratively decoded turbo-like codes. The paper also presents a technique which performs the entire calculation of the SP59 bound in the logarithmic domain, thus facilitating the exact calculation of the SP59 bound for moderate to large block lengths without the need for asymptotic approximations. It is shown that the ISP bound suggests an interesting alternative to the SP59 bound, especially for digital modulations of high spectral efficiency.

Index terms – Block codes, list decoding, maximum-likelihood decoding, phase shift keying modulation, sphere-packing bounds.

1 Introduction

The introduction of turbo-like codes which closely approach the Shannon capacity limit with moderate block lengths stirred up new interest in studying the limits of code performance as a function of the block length (see, e.g., [4, 7, 8, 9, 12, 16, 18, 19]).

The 1959 sphere-packing (SP59) bound of Shannon [13] serves for the evaluation of the performance limits of block codes whose transmission takes place over an AWGN channel. The bound is expressed in terms of the block length and rate of the code; it does not take into account the modulation used, but only assumes that the signals are of equal energy. This lower bound on the decoding error probability is used as a reference for quantifying the sub-optimality of codes with their practical decoding algorithms; by comparing computer simulations for the performance obtained by turbo-like codes over a wide range of rates and block sizes, it was exemplified in the literature that the gap between the sphere-packing bounds and the performance of these codes under efficient iterative decoding algorithms can be reduced below 1 dB.

The 1967 sphere-packing (SP67) bound, derived by Shannon, Gallager and Berlekamp [14], provides a lower bound on the decoding error probability of block codes as a function of their block length and code rate, and it applies to arbitrary discrete memoryless channels. Like the random coding bound of Gallager [5], the SP67 bound decays to zero exponentially with the block length. Further, the error exponent of the SP67 bound is a monotonic decreasing and convex function of the rate which is positive at rates below the channel capacity. This error exponent is tight at the portion of the rate region between the critical rate (R_{c}) and the channel capacity; for this important rate region, the error exponents of the SP67 and the random coding bounds coincide [14, Part 1].

The SP67 bound fails to provide informative results for codes of small to moderate block lengths. This is due to the original focus in [14] on asymptotic analysis. In their paper [18], Valembois and Fossorier revisited the SP67 bound in order to improve its tightness for codes of short to moderate block lengths, and also to extend its validity to memoryless continuous-output channels (e.g., the binary-input AWGN channel). The motivation for the study in [18] was strengthened by the outstanding performance of codes defined on graphs even with moderate block lengths. The remarkable improvement in the tightness of the SP67 bound was exemplified in [18] for the case of BPSK signaling over the AWGN channel, and it was shown that a tightened version of the SP67 bound provides an interesting alternative to the SP59 bound [13].

In this work, we derive an improved sphere-packing bound (referred to as the ISP bound) which further enhances the tightness of this bounding technique for codes of short to moderate block lengths. Under a mild condition, the validity of this new bound is extended to general memoryless channels, and it is applied to M-ary PSK block coded modulation schemes whose transmission take place over an AWGN channel. The tightness of the ISP bound is studied by comparing it with the random coding upper bound of Gallager [5], the tangential-sphere bound of Poltyrev [6, 10], classical and recent sphere-packing bounds (see [13, 14, 18]), as well as its comparison with computer simulations of iteratively decoded turbo-like codes. The tightness of the ISP bound for the Gaussian channel is also examined by calculating the regions of code lengths and rates for which this bound outperforms the SP59 bound and the capacity-limit bound (CLB). To this end, we present a technique to perform the entire calculation of the SP59 bound in the logarithmic domain; this technique circumvents numerical difficulties, and facilitates an exact calculation of the SP59 bound for moderate to large block lengths without the need for the asymptotic approximations in [13].

The paper is structured as follows: Section 2 reviews the concepts used in the derivation of the SP67 bound [14, Part 1], and its recent improvements in [18] targeting codes of short

to moderate block lengths. Section 3 introduces the ISP bound which further enhances the tightness of the bound in [18] and extends its validity for memoryless channels; the derivation of this bound relies on concepts and notation presented in Section 2. Section 4 starts by reviewing the SP59 bound of Shannon [13], and presenting the numerical algorithm used in [18] for calculating this bound. The numerical instability of this algorithm for moderate to large block lengths motivates the derivation of a new algorithm in Section 4 for the exact calculation of the SP59 bound, irrespectively of the block length. Section 5 provides numerical results which serve to compare the tightness of the ISP bound in Section 3 with the SP59 bound of Shannon [13] and the recent sphere-packing bound in [18]. The tightness of the ISP bound is exemplified in Section 5 for M-ary phase-shift-keying (PSK) block coded modulation schemes whose transmission takes place over the AWGN channel, and also for the binary erasure channel (BEC). We conclude our discussion in Section 6. Technical calculations are relegated to the appendices.

2 The 1967 Sphere-Packing Bound and Improvements

In this section, we outline the main steps in the derivation of the SP67 bound. We then survey the improvements to the bound, as suggested in [18], which also extend the validity of the bound to memoryless discrete-input continuous-output channels. This serves as a preparatory stage for presenting an improved sphere-packing bound in the next section which further enhances the tightness of the sphere-packing bounding technique for codes of short to moderate block lengths, and extends its use to general memoryless channels. For a comprehensive tutorial review of classical sphere-packing bounds (i.e., the SP59 and SP67 bounds) and recent improvements in [18], the reader is referred to [12, Chapter 5].

2.1 The 1967 Sphere-Packing Bound

Let us consider a block code $\mathcal{C}$ which consists of M codewords each of length N, and denote its codewords by $\mathbf{x}_1, \ldots, \mathbf{x}_M$. Assume that $\mathcal{C}$ is transmitted over a discrete memoryless channel (DMC) and the decoding is performed by a list decoder; for each received sequence $\mathbf{y}$, the decoder outputs a list of at most L integers belonging to the set $\{1, 2, \ldots, M\}$ which correspond to the indices of the codewords. A list decoding error is declared if the index of the transmitted codeword does not appear in the list. In [14], the authors derive a lower bound on the decoding error probability of an arbitrary block code with M codewords of length N, and an arbitrary list decoding scheme whose size is limited to L. The particular case where $L = 1$ clearly provides a lower bound on the performance under optimal ML decoding.

Let $\mathcal{Y}_m$ denote the set of output sequences $\mathbf{y}$ for which message m is on the decoding list, and define $P_m(\mathbf{y}) \triangleq \Pr(\mathbf{y}|\mathbf{x}_m)$. The probability of list decoding error when message m is sent is given by

$$P_{\mathrm{e},m} = \sum_{y \in \mathcal{Y}_m^c} P_m(\mathbf{y}) \ . \tag{1}$$

For the block code and list decoder under consideration, let $P_{\mathrm{e,max}}$ designate the maximal value of $P_{\mathrm{e},m}$ where $m \in \{1, 2, \ldots, M\}$. Assuming that the codewords are equally likely to be transmitted, the average decoding error probability is given by

$$P_{\mathrm{e}} = \frac{1}{M} \sum_{m=1}^{M} P_{\mathrm{e},m}. \tag{2}$$

Referring to a list decoder of size at most L, the code rate (in nats per symbol use) is defined as $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$.

The derivation of the SP67 bound is divided into three main steps. The first step is the derivation of upper and lower bounds on the error probability of a code consisting of two codewords only. The authors prove in [14] the following theorem:

Theorem 2.1 (Upper and Lower Bounds on the Pairwise Error Probability). [14, Theorem 5]: Let $P_1(\mathbf{y})$ and $P_2(\mathbf{y})$ be two probability assignments on a discrete set of sequences, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be disjoint decision regions for these sequences, $P_{\mathrm{e},1}$ and $P_{\mathrm{e},2}$ be given by (1), and assume that $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$ for at least one sequence $\mathbf{y}$. Then, for all $s \in (0,1)$

$$P_{\mathrm{e},1} > \frac{1}{4}\exp\Big(\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\Big) \tag{3}$$

or

$$P_{\mathrm{e},2} > \frac{1}{4}\exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\Big) \tag{4}$$

where

$$\mu(s) \triangleq \ln\Big(\sum_{\mathbf{y}} P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s\Big) \qquad 0 < s < 1. \tag{5}$$

Furthermore, for an appropriate choice of $\mathcal{Y}_1$ and $\mathcal{Y}_2$

$$P_{\mathrm{e},1} \leq \exp\Big(\mu(s) - s\mu'(s)\Big)$$

and

$$P_{\mathrm{e},2} \leq \exp\Big(\mu(s) + (1-s)\mu'(s)\Big).$$

The function μ is non-positive and convex over the interval $(0,1)$. The convexity of μ is strict unless $\frac{P_1(\mathbf{y})}{P_2(\mathbf{y})}$ is constant over all the sequences $\mathbf{y}$ for which $P_1(\mathbf{y})P_2(\mathbf{y}) \neq 0$. Moreover, the function μ is strictly negative over the interval $(0,1)$ unless $P_1(\mathbf{y}) = P_2(\mathbf{y})$ for all $\mathbf{y}$.

The initial motivation given for Theorem 2.1 is the calculation of bounds on the error probability of a two-word code. However, it is valid for any pair of probability assignments P_1 and P_2 and decision regions $\mathcal{Y}_1$ and $\mathcal{Y}_2$ which form a partitioning of the output vector space.

In the continuation of the derivation of the SP67 bound in [14], this theorem is used in order to keep control of the size of a decision region of a particular codeword without directly referring to other codewords. To this end, an arbitrary probability tilting measure f_N is introduced in [14] over all N-length sequences of channel outputs, requiring that it is factorized in the form

$$f_N(\mathbf{y}) = \prod_{n=1}^{N} f(y_n) \tag{6}$$

for an arbitrary output sequence $\mathbf{y} = (y_1, \ldots, y_N)$; the size of the set $\mathcal{Y}_m$ is defined as

$$F(\mathcal{Y}_m) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}_m} f_N(\mathbf{y}). \tag{7}$$

Next, [14] relies on Theorem 2.1 in order to relate the conditional error probability $P_{\mathrm{e},m}$ and $F(\mathcal{Y}_m)$ for fixed composition codes; this is done by associating $\Pr(\mathbf{y}|\mathbf{x}_m)$ and $f_N(\mathbf{y})$ with $P_1(\mathbf{y})$ and $P_2(\mathbf{y})$, respectively. Theorem 2.1 is applied as described above to derive a parametric lower bound on the size of the decision region $\mathcal{Y}_m$ or the conditional error probability $P_{\mathrm{e},m}$. Using a simple upper bound on the smallest size of the set $\mathcal{Y}_m$ where $m \in \{1, \ldots, M\}$, and by upper

bounding the conditional error probability of the corresponding codeword by $P_{\text{e,max}}$, a lower bound on the maximal error probability is obtained. Next, the probability assignment $f \triangleq f_s$ is optimized in [14], so as to get the tightest (i.e., maximal) lower bound within this form while considering a code whose composition minimizes the bound (so that the bound holds for all fixed composition codes). A solution for this min-max problem, as provided in [14], leads to the following theorem which gives a lower bound on the maximal block error probability of an arbitrary fixed composition block code (for a more detailed review of these concepts, see [12, Section 5.3]).

Theorem 2.2 (Sphere-Packing Lower Bound on the Maximal Decoding Error Probability for Fixed Composition Codes). [14, Theorem 6]: Let $\mathcal{C}$ be a *fixed composition code* of M codewords and block length N. Assume that the transmission of $\mathcal{C}$ takes place over a DMC, and let $P(j|k)$ be the set of transition probabilities characterizing this channel (where $j \in \{1,\ldots,J\}$ and $k \in \{1,\ldots,K\}$ designate the channel input and output alphabets, respectively). For an arbitrary list decoder where the size of the list is limited to L, the *maximal error probability* ($P_{\text{e,max}}$) satisfies

$$P_{\text{e,max}} \geq \exp\left[-N\left(E_{\text{sp}}\Big(R - \frac{\ln 4}{N} - \varepsilon\Big) + \sqrt{\frac{8}{N}}\ln\Big(\frac{e}{\sqrt{P_{\min}}}\Big) + \frac{\ln 4}{N}\right)\right]$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$ is the rate of the code, $P_{\min}$ designates the smallest non-zero transition probability of the DMC, the parameter ε is an arbitrarily small positive number, and the function E_{sp} is given by

$$E_{\text{sp}}(R) \triangleq \sup_{\rho \geq 0}\big(E_0(\rho) - \rho R\big) \tag{8}$$

$$E_0(\rho) \triangleq \max_{\mathbf{q}} E_0(\rho, \mathbf{q}) \tag{9}$$

$$E_0(\rho, \mathbf{q}) \triangleq -\ln\left(\sum_{j=1}^{J}\Big[\sum_{k=1}^{K} q_k P(j|k)^{\frac{1}{1+\rho}}\Big]^{1+\rho}\right). \tag{10}$$

The maximum in the RHS of (9) is taken over all probability vectors $\mathbf{q} = (q_1,\ldots,q_K)$, i.e., over all $\mathbf{q}$ with K non-negative components summing to 1.

The reason for considering fixed composition codes is that the optimal probability distribution f_s depends on the composition of the codewords. The derivation of the improved sphere-packing bound in Section 3 is based on the observation that for a wide class of channels, the optimal probability assignment f_s and the lower bound on the error probability are independent of the codeword composition. Therefore, it is possible to adopt the technique used for deriving Theorem 2.2 towards the derivation of a lower bound on the maximal error probability of an arbitrary block code.

The next step in the derivation of the SP67 bound is the application of Theorem 2.2 towards the derivation of a lower bound on the maximal error probability of an arbitrary block code. This is performed by lower bounding the maximal error probability of the code by the maximal error probability of its largest fixed composition subcode. Since the number of possible compositions is polynomial in the block length, one can lower bound the rate of the largest fixed composition subcode by $R - O\left(\frac{\ln N}{N}\right)$ where R is the rate of the original code. Clearly, the rate loss caused by considering this subcode vanishes when the block length tends to infinity; however, it loosens of the bound for short to moderate length codes. Finally, the bound on the maximal error probability is transformed into a bound on the average error probability by considering an

expurgated code which contains half of the codewords of the original code with the lowest decoding error probability. This finally leads to the SP67 bound [14].

Theorem 2.3 (The 1967 Sphere-Packing Bound for Discrete Memoryless Channels). [14, Theorem 2]: Let $\mathcal{C}$ be an arbitrary block code whose transmission takes place over a DMC. Assume that the DMC is specified by the set of transition probabilities $P(j|k)$ where $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, J\}$ designate the channel input and output alphabets, respectively. Assume that the code $\mathcal{C}$ forms a set of M codewords of length N (i.e., each codeword is a sequence of N letters from the input alphabet), and consider an arbitrary list decoder where the size of the list is limited to L. Then, the *average decoding error probability* of the code $\mathcal{C}$ satisfies

$$P_{\mathrm{e}}(N, M, L) \geq \exp\left\{-N\left[E_{\mathrm{sp}}\left(R - O_1\left(\frac{\ln N}{N}\right)\right) + O_2\left(\frac{1}{\sqrt{N}}\right)\right]\right\}$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$, the error exponent $E_{\mathrm{sp}}(R)$ is introduced in (8), and the terms

$$\begin{aligned} O_1\left(\frac{\ln N}{N}\right) &= \frac{\ln 8}{N} + \frac{K \ln N}{N} \\ O_2\left(\frac{1}{\sqrt{N}}\right) &= \sqrt{\frac{8}{N}}\ \ln\left(\frac{e}{\sqrt{P_{\min}}}\right) + \frac{\ln 8}{N} \end{aligned} \tag{11}$$

scale like $\frac{\ln N}{N}$ and the inverse of the square root of N, respectively (hence, they vanish as we let N tend to infinity), and $P_{\min}$ denotes the smallest non-zero transition probability of the DMC.

2.2 Improvements on the 1967 Sphere-Packing Bound Introduced in [18]

In [18], Valembois and Fossorier revisit the derivation of the SP67 bound, this time with the intention of making the bound useful for codes with short to moderate block lengths. They present four modifications to the classical derivation in [14] which improve the pre-exponent of the SP67 bound. The new bound derived in [18] is also valid for memoryless channels with continuous output (as opposed to the SP67 bound which is only valid for DMCs). It is applied to the binary-input AWGN channel, and it is also compared with the SP59 bound which is valid for any set of equal energy signals transmitted over the AWGN channel; this comparison shows that the recent bound in [18] provides an interesting alternative to the SP59 bound, especially for high code rates. In this section, we review the improvement suggested in [18] and present the resulting bound.

The first modification suggested in [18] is the addition of a free parameter in the derivation of the lower bound on the decoding error probability of two-word codes; this free parameter is used in conjunction with Chebychev's inequality, and it is later optimized in order to get the tightest bound within this form.

A second improvement presented in [18] is related to a simplification in [14] where the second derivative of the function μ, as is defined in (5), is upper bounded by $\frac{e}{\sqrt{P_{\min}}}$. This bound results in no asymptotic loss, but it can loosen the bound for short to moderate code lengths. By using the exact value of μ'' instead, the tightness of the resulting bound is further improved in [18]. This modification also makes the bound suitable to memoryless channels with continuous output, as it is no longer required that $P_{\min}$ is positive. It should be noted that this causes a small discrepancy in the derivation of the bound; the derivation of a lower bound on the error probability which is *uniform* over all fixed composition codes relies on finding the composition which minimizes the lower bound. This optimization problem is solved in [14] for the case where the upper bound on μ'' is applied and the same composition is used [18], without

checking whether it is still that minimizing composition. However, as we see in the next section, for a wide class of channels the value of the bound is independent of the code composition and therefore the VF bound stays valid. This class of channels includes all memoryless binary-input output-symmetric (MBIOS) channels; in particular, it includes the binary symmetric channel (BSC), and the binary-input AWGN channel considered in [18].

A third improvement in [18] concerns the particular selection of the value of $\rho \geq 0$ which leads to the derivation of Theorem 2.3. In [14], ρ is set to be the value $\tilde{\rho}$ which minimizes the error exponent of the SP67 bound (i.e., the upper bound on the error exponent). This choice emphasizes the similarity between the error exponents of the SP67 lower bound and the Gallager random coding upper bound, hence proving that the error exponent of the SP67 bound is tight for all rates above the critical rate of the channel. In order to tighten the bound for the finite length case, [18] chooses the value of ρ to be ρ^* which provides the tightest possible lower bound on the decoding error probability. The asymptotic accuracy of the original SP67 bound implies that as the block length tends to infinity, $\tilde{\rho} \to \rho^*$; however, for codes of finite block length, this simple observation tightens the bound with almost no penalty in the computational cost of the resulting bound.

The fourth observation made in [18] concerns the final stage in the derivation of the SP67 bound. In order to get a lower bound on the maximal error probability of an arbitrary block code, the derivation in [14] considers the maximal error probability a fixed composition subcode of the original code. In [14], a simple lower bound on the size of the largest fixed composition subcode is given; namely, the size of the largest fixed composition subcode is not less than the size of the entire code divided by the number of possible compositions. Since this number of compositions is equal at most to the number of possible ways to divide N symbols into K types, this value is given by $\binom{N+K-1}{K-1}$. To simplify the final expression of the SP67, [14] applies the upper bound $\binom{N+K-1}{K-1} \leq N^K$. Since this expression is polynomial is the block length N, there is no asymptotic loss to the error exponent. However, by using the exact expression for the number of possible compositions, the bound in [18] is tightened for codes of short to moderate block lengths. Applying these four modifications in [18] yields an improved lower bound on the decoding error probability of block codes transmitted over memoryless channels with finite input alphabets. As mentioned above, these modifications also extend the validity of the new bound to discrete-time memoryless channels with continuous outputs. However, the requirement of a finite input alphabet still remains, as it is required in order to apply the bound to arbitrary block codes, and not only to fixed composition codes. The VF bound [18] is given in the following theorem:

Theorem 2.4 (Improvement on the 1967 Sphere-Packing Bound for Discrete Memoryless Channels). [18, Theorem 7]: Under the assumptions and notation used in Theorem 2.3, the *average decoding error probability* satisfies

$$P_{\mathrm{e}}(N, M, L) \geq \exp\Big\{-N \widetilde{E}_{\mathrm{sp}}(R, N)\Big\}$$

where

$$\widetilde{E}_{\mathrm{sp}}(R, N) \triangleq \sup_{x > \frac{\sqrt{2}}{2}} \Big\{ E_0(\rho_x) - \rho_x \Big(R - O_1\Big(\frac{\ln N}{N}, x\Big)\Big) + O_2\Big(\frac{1}{\sqrt{N}}, x, \rho_x\Big)\Big\}$$

and

$$
\begin{aligned}
R &\triangleq \frac{\ln\left(\frac{M}{L}\right)}{N} \\
O_1\Big(\frac{\ln N}{N}, x\Big) &\triangleq \frac{\ln 8}{N} + \frac{\ln\binom{N+K-1}{K-1}}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N} \\
O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) &\triangleq x\sqrt{\frac{8}{N}\sum_{k=1}^{K} q_{k,\rho}\nu_k^{(2)}(\rho)} + \frac{\ln 8}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}
\end{aligned}
\tag{12}
$$

$$
\begin{aligned}
\nu_k^{(1)}(\rho) &\triangleq \frac{\sum_{j=1}^{J}\beta_{j,k,\rho}\ln\frac{\beta_{j,k,\rho}}{P(j|k)}}{\sum_{j=1}^{J}\beta_{j,k,\rho}} \\
\nu_k^{(2)}(\rho) &\triangleq \frac{\sum_{j=1}^{J}\beta_{j,k,\rho}\ln^2\frac{\beta_{j,k,\rho}}{P(j|k)}}{\sum_{j=1}^{J}\beta_{j,k,\rho}} - \left[\nu_k^{(1)}(\rho)\right]^2 \\
\beta_{j,k,\rho} &\triangleq P(j|k)^{\frac{1}{1+\rho}}\cdot\left(\sum_{k'} q_{k',\rho}P(j|k')^{\frac{1}{1+\rho}}\right)^{\rho}
\end{aligned}
$$

where $\mathbf{q}_\rho \triangleq (q_{1,\rho},\ldots,q_{K,\rho})$ designates the input distribution which maximizes $E_0(\rho,\mathbf{q})$ in (9), and the parameter ρ_x is determined by solving the equation

$$
R - O_1\Big(\frac{\ln N}{N}, x\Big) = -\frac{1}{\rho_x}\sum_{k} q_{k,\rho_x}\nu_k^{(1)}(\rho_x) + \frac{x}{\rho_x}\sqrt{\frac{2}{N}\sum_{k=1}^{K} q_{k,\rho}\nu_k^{(2)}(\rho)}.
$$

3 An Improved Sphere-Packing Bound

In this section, we derive an improved lower bound on the decoding error probability which utilizes the sphere-packing bounding technique. This bound is referred to as the improved sphere-packing (ISP) bound, and its validity is extended to a wide class of discrete-time memoryless channels.

To keep the notation simple, we derive the ISP bound under the assumption that the communication takes place over a DMC. This assumption allows us to follow the first steps of the proof of the SP67 bound in [14]. However, the derivation of the bound is justified later for a wider class of memoryless channels with discrete or continuous input and output alphabets. Some remarks are given at the end of the derivation. In the continuation of this section, the bound is particularized for M-ary PSK block coded modulation schemes with coherent detection over the AWGN channel.

3.1 Derivation of the New Sphere-Packing Bound

We start our analysis by following the derivation of the SP67 bound, as given in [14], where we take advantage of the improvements suggested in [18]. We show that under a mild condition on memoryless communication channels, the derivation of the sphere-packing bound can be modified so that the intermediate step of bounding the maximal error probability for fixed composition codes can be skipped. This allows the tightening of the sphere-packing bound, and also the extension of its validity to the case where the channel input as well as the channel output are continuous. We begin the derivation by introducing the modified lower bound on

the decoding error probability for a code book of two codewords, as presented in [18]; although this part simply follows the idea in [18] and the analysis in [14], this preparatory stage for the derivation of the bound for a general code is introduced here since it does not appear explicitly in [18] (though it is straightforward in light of the analysis in [14] and the suggested modification in [18]). The novelty here is by moving directly to the derivation of the sphere-packing bound for a general block code, assuming a list decoder of size L, without the need to derive the bound first for fixed composition codes (thus differing from the derivation of the sphere-packing bounds in [14, 18]).

Decoding Error Probability for Two Codewords We start the analysis by considering the decoding error probability of a codebook of two codewords, $\mathbf{x}_1$ and $\mathbf{x}_2$, whose transmission takes place over a DMC. Define $P_m(\mathbf{y}) \triangleq P(\mathbf{y}|\mathbf{x}_m)$ (where $m = 1, 2$). Following the notation in [14], we define the log-likelihood ratio (LLR) as

$$D(\mathbf{y}) = \ln\left(\frac{P_2(\mathbf{y})}{P_1(\mathbf{y})}\right)$$

and the probability distribution

$$Q_s(\mathbf{y}) = \frac{P_1(\mathbf{y})^{1-s} P_2(\mathbf{y})^s}{\sum_{\mathbf{y}'} P_1(\mathbf{y}')^{1-s} P_2(\mathbf{y}')^s}$$

for any $0 < s < 1$. Based on the introduction of the function μ in (5), it can be easily verified that (see [14])

$$\begin{aligned}
\mu'(s) &= \mathbb{E}_{Q_s}\big(D(\mathbf{y})\big) && (13)\\
\mu''(s) &= \mathrm{Var}_{Q_s}\big(D(\mathbf{y})\big) && (14)\\
P_1(\mathbf{y}) &= \exp\big(\mu(s) - sD(\mathbf{y})\big)\, Q_s(\mathbf{y}) && (15)\\
P_2(\mathbf{y}) &= \exp\big(\mu(s) + (1-s)D(\mathbf{y})\big)\, Q_s(\mathbf{y}) && (16)
\end{aligned}$$

where $\mathbb{E}_Q$ and Var_Q stand, respectively, for the statistical expectation and variance w.r.t. a probability distribution Q. Let us define the set

$$\mathcal{Y}_s^x \triangleq \Big\{\mathbf{y} : |D(\mathbf{y}) - \mu'(s)| \le x\sqrt{2\mu''(s)}\Big\}, \quad x > 0. \tag{17}$$

In the original derivation of the SP67 bound in [14], x was set to one; this free parameter is introduced in [18] in order to tighten the bound for finite-length block codes. By applying Chebyshev's inequality to (17), and relying on the equalities in (13) and (14), we get

$$\sum_{\mathbf{y}\in\mathcal{Y}_s^x} Q_s(\mathbf{y}) > 1 - \frac{1}{2x^2} \tag{18}$$

where this result is meaningful only for $x > \frac{\sqrt{2}}{2}$.

Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be the decoding regions of the two codewords $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Let also $\mathcal{A}^c$ designate the complementary of a set $\mathcal{A}$. We now bound the conditional error probability given that the first codeword $\mathbf{x}_1$ is transmitted, and get

$$\begin{aligned}
P_{\mathrm{e},1} &= \sum_{\mathbf{y}\in\mathcal{Y}_1^c} P_1(\mathbf{y})\\
&\ge \sum_{\mathbf{y}\in\mathcal{Y}_1^c \bigcap \mathcal{Y}_s^x} P_1(\mathbf{y})\\
&= \sum_{\mathbf{y}\in\mathcal{Y}_1^c \bigcap \mathcal{Y}_s^x} \exp\Big(\mu(s) - sD(\mathbf{y})\Big)\, Q_s(\mathbf{y})
\end{aligned}$$

where the last transition follows from (15). For every $\mathbf{y} \in \mathcal{Y}_s^x$, we get from (17)

$$\mu'(s) - x\sqrt{2\mu''(s)} \leq D(\mathbf{y}) \leq \mu'(s) + x\sqrt{2\mu''(s)}$$

and therefore

$$P_{\mathrm{e},1} \geq \exp\Big(\mu(s) - s\mu'(s) - s\ x\ \sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y} \in \mathcal{Y}_1^c \bigcap \mathcal{Y}_s^x} Q_s(\mathbf{y})\ , \qquad x > \frac{\sqrt{2}}{2}. \tag{19}$$

Following the same steps w.r.t. the conditional error probability of the second codeword and using (16), gives

$$P_{\mathrm{e},2} \geq \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\ x\ \sqrt{2\mu''(s)}\Big) \sum_{\mathbf{y} \in \mathcal{Y}_2^c \bigcap \mathcal{Y}_s^x} Q_s(\mathbf{y})\ , \qquad x > \frac{\sqrt{2}}{2}. \tag{20}$$

Since the sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ form a disjoint partitioning of the set of output vectors $\mathcal{Y}^N$, then

$$\sum_{\mathbf{y} \in \mathcal{Y}_1^c \bigcap \mathcal{Y}_s^x} Q_s(\mathbf{y}) + \sum_{\mathbf{y} \in \mathcal{Y}_2^c \bigcap \mathcal{Y}_s^x} Q_s(\mathbf{y}) = \sum_{\mathbf{y} \in \mathcal{Y}_s^x} Q_s(\mathbf{y})$$

and therefore, at least one of the sums in the LHS of this equality is necessarily not below half of the value of the RHS of this equality. From (18), at least one of these two sums should be not less than $\frac{1}{2}\left(1 - \frac{1}{2x^2}\right)$. By combining this result with (19) and (20), then for every $s \in (0,1)$

$$P_{\mathrm{e},1} > \left(\frac{1}{2} - \frac{1}{4x^2}\right)\ \exp\Big(\mu(s) - s\mu'(s) - s\ x\ \sqrt{2\mu''(s)}\Big) \tag{21}$$

or

$$P_{\mathrm{e},2} > \left(\frac{1}{2} - \frac{1}{4x^2}\right)\ \exp\Big(\mu(s) + (1-s)\mu'(s) - (1-s)\ x\ \sqrt{2\mu''(s)}\Big). \tag{22}$$

Lower Bound on the Decoding Error Probability of General Block Codes Let us now consider a block code $\mathcal{C}$ of length N with M codewords, denoted by $\{\mathbf{x}_m\}_{m=1}^M$; assume that the transmission takes place over a DMC with transition probabilities $P(j|k)$, where $k \in \{1, \ldots, K\}$ and $j \in \{1, \ldots, J\}$ designate the channel input and output alphabets, respectively. To this end, we rely on the result of the previous section which is valid for any pair of probability measures (P_1 and P_2). Let f_N be an arbitrary probability measure defined over the set of length-N sequences of the channel output, and which can be factorized as in (6). We refer to the pair of probability measures given by

$$P_1(\mathbf{y}) \triangleq \Pr(\mathbf{y}|\mathbf{x}_m), \quad P_2(\mathbf{y}) \triangleq f_N(\mathbf{y}) \tag{23}$$

where $\mathbf{x}_m$ is an arbitrary codeword of the code $\mathcal{C}$. Let $\mathcal{Y}_m$ be the decision region of the codeword $\mathbf{x}_m$, and let its size be defined as in (7). By combining (5) and (23), we define

$$\mu(s, m, f_N) \triangleq \ln\left(\sum_{\mathbf{y}} \Pr(\mathbf{y}|\mathbf{x}_m)^{1-s} f_N(\mathbf{y})^s\right), \quad 0 < s < 1. \tag{24}$$

By associating $\mathcal{Y}_m$ and $\mathcal{Y}_m^c$ with the two decision regions for the probability measures P_1 and P_2, respectively, we obtain from (19), (20) and the above setting that

$$P_{\mathrm{e},m} > \left(\frac{1}{2} - \frac{1}{4x^2}\right)\ \exp\Big(\mu(s, m, f_N) - s\mu'(s, m, f_N) - s\ x\ \sqrt{2\mu''(s, m, f_N)}\Big) \tag{25}$$

or

$$F(\mathcal{Y}_m) > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Big(\mu(s,m,f_N) + (1-s)\mu'(s,m,f_N) - (1-s)\; x\; \sqrt{2\mu''(s,m,f_N)}\Big) \quad (26)$$

where $x > \frac{\sqrt{2}}{2}$.

Let us denote by q_k^m the fraction of appearances of the letter k in the codeword $\mathbf{x}_m$. By assumption, the communication channel is memoryless and the function f_N is a probability measure which is factorized according to (6). Hence, for every $m \in \{1, 2, \ldots, M\}$, the function $\mu(s, m, f_N)$ in (24) is expressible in the form

$$\mu(s,m,f_N) = N \sum_{k=1}^{K} q_k^m \mu_k(s,f) \quad (27)$$

where

$$\mu_k(s,f) \triangleq \ln\left(\sum_{j=1}^{J} P(j|k)^{1-s} f(j)^s\right), \quad 0 < s < 1. \quad (28)$$

Substituting (27) in (25) and (26), then for every $s \in (0,1)$

$$P_{\mathrm{e},m} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\sum_k q_k^m \big(\mu_k(s,f) - s\mu_k'(s,f)\big) - s\; x\; \sqrt{\frac{2\sum_k q_k^m \mu_k''(s,f)}{N}}\right)\right\} \quad (29)$$

or

$$F(\mathcal{Y}_m) > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\sum_k q_k^m \big(\mu_k(s,f) + (1-s)\mu_k'(s,f)\big) - (1-s)\; x\; \sqrt{\frac{2\sum_k q_k^m \mu_k''(s,f)}{N}}\right)\right\}. \quad (30)$$

For $s \in (0,1)$, we choose the function f to be f_s, as is given in [14, Eqs. (4.18)-(4.20)]. Namely, for $0 < s < 1$, let $\mathbf{q}_s = \{q_{1,s}, \ldots, q_{K,s}\}$ satisfy the inequalities

$$\sum_j P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} \geq \sum_j \alpha_{j,s}^{\frac{1}{1-s}}\;; \quad \forall k \quad (31)$$

where

$$\alpha_{j,s} \triangleq \sum_{k=1}^{K} q_{k,s} P(j|k)^{1-s}. \quad (32)$$

The function f_s is given by

$$f_s(j) = \frac{\alpha_{j,s}^{\frac{1}{1-s}}}{\sum_{j'=1}^{J} \alpha_{j',s}^{\frac{1}{1-s}}}, \quad j \in \{1, \ldots, J\}. \quad (33)$$

Note that the input distribution $\mathbf{q}_s$ is *independent of the code* $\mathcal{C}$, as it only depends on the channel statistics. By multiplying both sides of (31) by $q_{k,s}$ and summing over k (where $\sum_k q_{k,s} = 1$), we get

$$\sum_k \left\{ q_{k,s} \sum_j P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} \right\} \geq \sum_j \alpha_{j,s}^{\frac{1}{1-s}}.$$

Examining the LHS of the above inequality gives

$$\begin{aligned}\sum_k \left\{ q_{k,s} \sum_j P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} \right\} &= \sum_j \left\{ \alpha_{j,s}^{\frac{s}{1-s}} \sum_k q_{k,s} P(j|k)^{1-s} \right\} \\ &= \sum_j \alpha_{j,s}^{\frac{1}{1-s}} \end{aligned} \tag{34}$$

where the last equality follows from (32). Let us now assume that for every $0 < s < 1$, the support of $\mathbf{q}_s$ consists of the entire input alphabet. By our assumption, $q_{k,s} \neq 0$ for any k, thus by combining (31) and (34), we obtain that for all values of k

$$\sum_j P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} = \sum_j \alpha_{j,s}^{\frac{1}{1-s}}. \tag{35}$$

Note that this equality holds for all values of k due to our assumption that $q_{k,s} \neq 0$ for all k; otherwise, this equality may not hold for those values of k for which $q_{k,s}$ is zero. From (28), since f in general is allowed to depend on the parameter s (as we examine the validity of the bound for any individual value of $s \in (0,1)$), we get

$$\begin{aligned}\mu_k(s, f_s) &= \ln \sum_j P(j|k)^{1-s} f_s(j)^s \\ &\overset{(a)}{=} \ln \sum_j P(j|k)^{1-s} \alpha_{j,s}^{\frac{s}{1-s}} - s \ln \sum_j \alpha_{j,s}^{\frac{1}{1-s}} \\ &\overset{(b)}{=} (1-s) \ln \sum_j \alpha_{j,s}^{\frac{1}{1-s}} \\ &\overset{(c)}{=} (1-s) \ln \left(\sum_j \left[\sum_k q_{k,s} P(j|k)^{1-s} \right]^{\frac{1}{1-s}} \right) \end{aligned} \tag{36}$$

where (a) follows from the definition of f_s in (32) and (33), (b) follows from (35), and (c) follows from (32). Under the setting $s = \frac{\rho}{1+\rho}$, since the conditions on $\mathbf{q}_s$ in (31) are identical to the conditions on the input distribution $\mathbf{q} = \mathbf{q}_s$ which maximizes $E_0(\frac{s}{1-s}, \mathbf{q})$ as stated in [5, Theorem 4], then

$$\begin{aligned}\mu_k(s, f_s) &= (1-s) \ln \left(\sum_j \left[\sum_k q_{k,s} P(j|k)^{\frac{1}{1+\frac{s}{1-s}}} \right]^{1+\frac{s}{1-s}} \right) \\ &= -(1-s)\, E_0 \left(\frac{s}{1-s}, \mathbf{q}_s \right) \\ &= -(1-s)\, E_0 \left(\frac{s}{1-s} \right), \quad 0 < s < 1 \end{aligned} \tag{37}$$

where E_0 is given by (9). From (37), since μ_k is independent of k (let its common value for all k be denoted by $\mu_0(s, f_s)$), then from (29) and (30), it follows that for $0 < s < 1$

$$P_{\mathrm{e},m} > \left(\frac{1}{2} - \frac{1}{4x^2} \right) \exp \left\{ N \left(\mu_0(s, f_s) - s \mu_0'(s, f_s) - s\, x \sqrt{\frac{2 \mu_0''(s, f_s)}{N}} \right) \right\} \tag{38}$$

or

$$F_s(\mathcal{Y}_m) > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\mu_0(s,f_s) + (1-s)\mu_0'(s,f_s) - (1-s)\,x\,\sqrt{\frac{2\mu_0''(s,f_s)}{N}}\right)\right\} \quad (39)$$

where $F_s(\mathcal{Y}_m) \triangleq \sum_{\mathbf{y}\in\mathcal{Y}_m} f_{s,N}(\mathbf{y})$. Similarly yo [14], we relate $F_s(\mathcal{Y}_m)$ to the number of codewords M and the size of the decoding list L by observing that

$$\sum_{m=1}^{M} F_s(\mathcal{Y}_m) = \sum_{m=1}^{M}\sum_{\mathbf{y}\in\mathcal{Y}_m} f_{s,N}(\mathbf{y}) \le L.$$

The last inequality holds since each $\mathbf{y}\in\mathcal{Y}^N$ appears in at most L subsets $\{\mathcal{Y}_m\}_{m=1}^M$ and also $\sum_{\mathbf{y}} f_{s,N}(\mathbf{y}) = 1$. It follows that for each $s\in(0,1)$, there exists an index $m_s \in \{1,2,\ldots,M\}$ such that $F_s(\mathcal{Y}_{m_s}) \le \frac{L}{M}$. Substituting this in (38) and (39), and upper bounding P_{e,m_s} by the maximum over m of $P_{\mathrm{e},m}$ (this maximal error probability is denoted by $P_{\mathrm{e,max}}$) implies that for $0<s<1$

$$P_{\mathrm{e,max}} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\mu_0(s,f_s) - s\mu_0'(s,f_s) - s\,x\,\sqrt{\frac{2\mu_0''(s,f_s)}{N}}\right)\right\} \quad (40)$$

or

$$\frac{L}{M} > \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\left\{N\left(\mu_0(s,f_s) + (1-s)\mu_0'(s,f_s) - (1-s)\,x\,\sqrt{\frac{2\mu_0''(s,f_s)}{N}}\right)\right\}. \quad (41)$$

A lower bound on the maximum error probability can be obtained from (40) by substituting any value of $s\in(0,1)$ for which the inequality in (41) does not hold. In particular we choose a value $s=s_x$ such that the inequality in (41) is replaced by equality, i.e.,

$$\begin{aligned}\frac{L}{M} &= \exp(-NR)\\ &= \left(\frac{1}{2} - \frac{1}{4x^2}\right) \exp\Bigg\{N\Big(\mu_0(s_x,f_{s_x}) + (1-s_x)\,\mu_0'(s_x,f_{s_x})\\ &\qquad -(1-s_x)\,x\,\sqrt{\frac{2\mu_0''(s_x,f_{s_x})}{N}}\Big)\Bigg\} \end{aligned} \quad (42)$$

where $R \triangleq \frac{\ln\left(\frac{M}{L}\right)}{N}$ designates the code rate in nats per channel use. Note that the existence of a solution $s=s_x$ to equation (42) can be demonstrated in a similar way to the arguments in [14, Eqs. (4.28)–(4.35)] for the non-trivial case where the sphere-packing bound does not reduce to the trivial inequality $P_{\mathrm{e,max}} \ge 0$. This particular value of s is chosen since for large enough N, the RHS of (40) is monotonically decreasing while the RHS of (41) is monotonically increasing for $s\in(0,1)$; thus, this choice is optimal for large enough N. This particular choice of s_x also allows to get a simpler representation of the bound on $P_{\mathrm{e,max}}$. Rearranging equation (42) gives

$$\mu_0'(s_x,f_{s_x}) = -\frac{1}{1-s_x}\left[R + \mu_0(s_x,f_{s_x}) + \frac{1}{N}\,\ln\left(\frac{1}{2} - \frac{1}{4x^2}\right)\right] + x\sqrt{\frac{2\mu_0''(s_x,f_{s_x})}{N}}.$$

Substituting $s=s_x$ and the last equality into (40) yields that

$$\begin{aligned}P_{\mathrm{e,max}} &> \exp\Bigg\{N\Bigg(\frac{\mu_0(s_x,f_{s_x})}{1-s_x} + \frac{s_x}{1-s_x}\left(R + \frac{1}{N}\ln\Big(\frac{1}{2} - \frac{1}{4x^2}\Big)\right)\\ &\qquad -s_x\,x\,\sqrt{\frac{8\mu_0''(s_x,f_{s_x})}{N}} + \frac{1}{N}\ln\Big(\frac{1}{2} - \frac{1}{4x^2}\Big)\Bigg)\Bigg\}. \end{aligned}$$

By applying (37) and defining $\rho_x \triangleq \frac{s_x}{1-s_x}$ we get

$$P_{\text{e,max}} > \exp\left\{-N\left(E_0(\rho_x) - \rho_x\left[R - \frac{\ln 4}{N} + \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}\right] + s_x\, x\, \sqrt{\frac{8\mu_0''(s_x, f_{s_x})}{N}} + \frac{\ln 4}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}\right)\right\}.$$

Note that the above lower bound on the maximal error probability holds for an arbitrary block code of length N and rate R. The selection of ρ_x is similar to the selection in [18] and gives the tightest lower bound within this form.

In order to transform the lower bound on the maximum error probability into a lower bound on the average error probability, we expurgate the original block code. In this standard approach, we look at the expurgated code which is comprised of the $\frac{M}{2}$ codewords with the lowest error probability. The average error probability of the original code is not below one-half of the maximal word error probability of the expurgated code. Since the rate of the expurgated code is $R' = R - \frac{\ln 2}{N}$ nats per channel use (the reduction in the rate by $\frac{\ln 2}{N}$ follows from reducing the size of the code by one-half), we get a lower bound on the average error probability of the original code which reads

$$P_{\text{e}} > \exp\left\{-N\left(E_0(\rho_x) - \rho_x\left[R - \frac{\ln 8}{N} + \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}\right] + s_x\, x\, \sqrt{\frac{8\mu_0''(s_x, f_{s_x})}{N}} + \frac{\ln 8}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}\right)\right\}$$

where $R' - \frac{\ln 4}{N}$ is replaced in the RHS of the bound above by $R - \frac{\ln 8}{N}$, $\rho_x = \frac{s_x}{1-s_x}$ and $s_x \in (0,1)$ is implicitly given as a solution of the equation

$$R - \frac{\ln 8}{N} + \frac{\ln\left(2-\frac{1}{x^2}\right)}{N} = -\mu_0(s_x, f_{s_x}) - (1-s_x)\mu_0'(s_x, f_{s_x}) + (1-s_x)\, x\, \sqrt{\frac{2\mu_0''(s_x, f_{s_x})}{N}}.$$

Finally, we optimize over the parameter $x \in (\frac{\sqrt{2}}{2}, \infty)$ in order to get the tightest lowest bound of this form.

The derivation above only relies on the fact that the channel is memoryless, but does not rely on the fact that the input or output alphabets are discrete. As mentioned in Section 2.2, the original derivation of the SP67 bound in [14] relies on the fact that the input and output alphabets are finite in order to bound the second derivative of μ by $\frac{e}{\sqrt{P_{\min}}}$, where $P_{\min}$ designates the smallest non-zero transition probability of the channel. This requirement was relaxed in [18] to the requirement that only the input alphabet is finite; to this end, the second derivative of the function μ is calculated, thus the above upper bound on this second derivative is replaced by its exact value. However, the requirement for a finite input alphabet remains in [18] due to the fact that the derivation still relies on considering a fixed composition subcode of the original code, and therefore requires that the number of possible compositions for a given length N is finite. The derivation in this section circumvents the use of fixed composition codes, and as a by product, it also relaxes the requirement of a finite input alphabet. The validity of the derivation for memoryless continuous-input and continuous-output channels is provided in the continuation (see Remark 3.4). This leads to the following theorem, which provides an improved sphere-packing lower bound on the error probability of block codes transmitted over memoryless channels.

Theorem 3.1 (An Improved Sphere-Packing (ISP) Bound for Memoryless Channels). Let $\mathcal{C}$ be an arbitrary block code consisting of M codewords, each of length N. Assume that $\mathcal{C}$ is transmitted over a memoryless channel which is specified by the transition probabilities (or densities) $P(j|k)$ where $k \in \mathcal{K}$ and $j \in \mathcal{J}$ designate the channel input and output alphabets, respectively. Assume an arbitrary list decoder where the size of the list is limited to L. If the support of $\mathbf{q}_s$ which satisfies the inequalities in (31) consists of the entire input alphabet for all $0 < s < 1$, then the *average decoding error probability* satisfies

$$P_{\mathrm{e}}(N, M, L) \geq \exp\Big\{-N\widetilde{E}_{\mathrm{sp}}(R, N)\Big\}$$

where

$$\widetilde{E}_{\mathrm{sp}}(R, N) \triangleq \sup_{x > \frac{\sqrt{2}}{2}} \Big\{E_0(\rho_x) - \rho_x\Big(R - O_1\Big(\frac{1}{N}, x\Big)\Big) + O_2\Big(\frac{1}{\sqrt{N}}, x, \rho_x\Big)\Big\} \tag{43}$$

and the function E_0 is introduced in (9),

$$\begin{aligned} R &\triangleq \frac{1}{N}\ \ln\Big(\frac{M}{L}\Big) \\ O_1\Big(\frac{1}{N}, x\Big) &\triangleq \frac{\ln 8}{N} - \frac{\ln\big(2 - \frac{1}{x^2}\big)}{N} \qquad (44) \\ O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) &\triangleq s(\rho)\ x\ \sqrt{\frac{8}{N}\mu_0''\big(s(\rho)\big)} + \frac{\ln 8}{N} - \frac{\ln\big(2 - \frac{1}{x^2}\big)}{N}. \qquad (45) \end{aligned}$$

Here, $s(\rho) \triangleq \frac{\rho}{1+\rho}$, the parameter ρ_x in the RHS of (43) is determined by solving the equation

$$R - O_1\Big(\frac{1}{N}, x\Big) = -\mu_0\big(s(\rho), f_{s(\rho)}\big) - \big(1 - s(\rho)\big)\mu_0'\big(s(\rho), f_{s(\rho)}\big) + \big(1 - s(\rho)\big)\ x\ \sqrt{\frac{2\mu_0''\big(s(\rho), f_{s(\rho)}\big)}{N}} \tag{46}$$

and the functions $\mu_0(s, f)$ and f_s are defined in (28) and (33), respectively.

Remark 3.1. The requirement that the support of the input distribution $\mathbf{q}_s$ which maximizes the sphere-packing error exponent consists of the entire input alphabet for all $s \in (0, 1)$ is crucial. It is basically what in essence makes the difference between the VF bound in [18] and the ISP bound here. This assumption allows us to show in (36) that $\mu_k(s, f_s)$ is independent of k (i.e., it is independent of the symbol in the channel input) and thus allows us to represent the bound in (38) and (39) independently of the composition $\mathbf{q}^m$ of the codeword $\mathbf{x}_m$. In case that the support of the input distribution $\mathbf{q}_s$ does not satisfy the above condition, the lower bound on the maximum block error probability becomes a function of the codeword index m_s which in turn is a function of the probability distribution f_s, and therefore requires the intermediate step of the derivation of the bound for fixed composition codes as in [14, 18]. This mutual dependency does not allow us in general to complete the proof for the general case.

Remark 3.2. In light of the previous remark, the ISP bound differs from the VF bound [18] (see Theorem 2.4) in the sense that the term $\frac{\log\binom{N+K-1}{K-1}}{N}$ is removed from $O_1(\frac{\ln N}{N}, x)$. Therefore, the shift in the rate of the error exponent of the ISP bound behaves asymptotically like $O_1\left(\frac{1}{N}\right)$ instead of $O_1\left(\frac{\ln N}{N}\right)$ (see (11), (12) and (44)). This difference indicates a tightening of the pre-exponent of the ISP bound (as compared to the SP67 and VF bounds) which is expected to be especially pronounced for small to moderate block lengths and when the size of the channel input alphabet is increased.

Remark 3.3. The rate loss as a result of the expurgation of the code by removing half of the codewords with the largest error probability was ignored in [18]. The term $\frac{\ln 4}{N}$, as it appears in the term $O_1(\frac{\ln N}{N}, x)$ of [18, Theorem 7], should be therefore replaced by $\frac{\ln 8}{N}$ (see (44)).

Remark 3.4. Under the mild condition discussed in Remark 3.1, the ISP bound is also applicable to continuous-input memoryless channels. In contrast to (12) which depends on the size of the channel input alphabet (K)) and requires it to be finite, the parallel expression in (44) which corresponds to the ISP bound is not subject to this requirement. This inherent difference stems from Remark 3.1. When the ISP bound is applied to a continuous-input memoryless channel, the distribution of the channel input, as used for the derivation of the bound for a DMC, is replaced by the probability density function of the continuous channel input. Similarly, the transition probability of a DMC is replaced by a transition density function for a memoryless channel with continuous input or output alphabets, and the sums are replaced by integrals. Note that these densities may include Dirac delta functions which appear at the points where the corresponding input distribution or the transition density function of the channel are discontinuous.

Discussion on Theorem 3.1 In the following, we refer to another possibility of tightening the ISP bound for codes of short to moderate block lengths. To this end, note that in the final step of the derivation of the ISP bound, we move from a lower bound on the maximal error probability to a lower bound on the average error probability. Similarly to the derivation of the SP67 bound [14], this is done by expurgating half of the codewords and applying the fact that the maximal error probability of the expurgated code, composed of half of the codewords whose error probabilities are lowest, is not greater than twice the average error probability of the entire code. The decision to consider a code composed of exactly *half* of the codewords is arbitrary, and one may consider an expurgated code which includes an arbitrary fraction α of the codewords with the lowest error probabilities (where $0 < \alpha < 1$). In this case, the average error probability of the entire code is at least a fraction $1-\alpha$ of the maximal error probability of the expurgated code, and the rate of the expurgated code is decreased by $\frac{\ln\left(\frac{1}{\alpha}\right)}{N}$. This modifies the final form in Theorem 3.1; more explicitly, the expressions $O_1\left(\frac{1}{N}, x\right)$ and $O_2\left(\frac{1}{\sqrt{N}}, x\right)$ in (44) and (45), respectively, are converted to

$$O_1\Big(\frac{1}{N}, x\Big) \triangleq \frac{\ln\left(\frac{4}{\alpha}\right)}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N} \tag{47}$$

$$O_2\Big(\frac{1}{\sqrt{N}}, x, \rho\Big) \triangleq s_x x\sqrt{\frac{8}{N}\mu_0''\big(s(\rho)\big)} + \frac{\ln\left(\frac{4}{1-\alpha}\right)}{N} - \frac{\ln\left(2-\frac{1}{x^2}\right)}{N}. \tag{48}$$

Note that (47) and (48) coincide with (44) and (45), respectively, for the case where $\alpha = \frac{1}{2}$. To exemplify the effect of the parameter α, we refer to a block code of length $N = 150$ bits and rate $0.9\frac{\text{bits}}{\text{channel use}}$ which is transmitted over the AWGN channel using BPSK modulation. In Figure 1, we plot the ISP bound for this scenario, where we set α to $\frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$. Firstly, it can be observed that the choice $\alpha = \frac{1}{2}$ is relatively good for the range of energy per bit to one-sided spectral noise density shown in Fig. 1 (and in fact, it is the best choice among these three values of α for the intermediate range of $\frac{E_\mathrm{b}}{N_0}$ which is not depicted in this figure). It can be also observed that for high block error probabilities, the smallest value of $\alpha = \frac{1}{4}$ gives the tightest lower bound among the bounds which correspond to the above three values of α; on the other hand, for low block error probabilities, the larger values of α give more appealing results (note that, however, the bound is useless for $\alpha \to 1^-$). This is due to the fact that the value ρ_x (see (46)) is monotonically increasing as the value of $\frac{E_\mathrm{b}}{N_0}$ is increased, and is zero for all values of $\frac{E_\mathrm{b}}{N_0}$ for which the shifted code rate (see the LHS of (46)) is above the corresponding channel capacity (thus, the communication is not reliable). For low values of $\frac{E_\mathrm{b}}{N_0}$ (which yield high error probabilities), the optimal ρ_x is very small; therefore, the fact that the rate of the

expurgated code is smaller than the code rate of the original code has little affect on the bound, but the factor $1-\alpha$ becomes larger as the value of α is decreased. This therefore implies that this bounding technique favors smaller values of α as the value of $\frac{E_b}{N_0}$ decreases. On the other hand, for large enough values of $\frac{E_b}{N_0}$ (which correspond to lower error probabilities), the optimal ρ_x becomes larger and eventually the penalty for the decreased code rate of the expurgated code caused by selecting a small value of α outweighs the advantage of the larger factor $1-\alpha$; hence, in the high SNR regime, this bounding technique favors large values of α (i.e., values of α closer to 1). For moderate values of $\frac{E_b}{N_0}$, the tradeoff between the rate of the expurgated code and the value of the factor $1-\alpha$ dictates the optimal choice of the expurgation parameter.

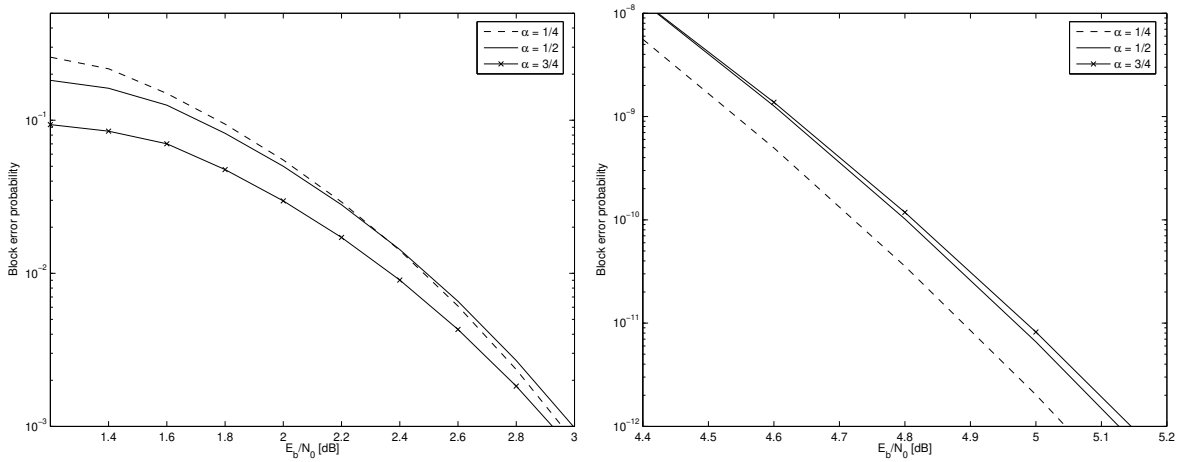

Figure 1: A comparison of the improved sphere-packing (ISP) lower bound from Section 3 for different values of the expurgation parameter α. The examined code is of block length $N = 150$ bits and rate $0.9\frac{\text{bits}}{\text{channel use}}$. The two plots refer to BPSK modulated signals whose transmission takes place over the AWGN channel, for lower (left plot) and higher (right plot) values of $\frac{E_b}{N_0}$.

3.2 Application of the New Bound to M-ary PSK Block Coded Modulation

In this section, we apply the ISP bound to the case where the codewords of a block code are modulated by an M-ary PSK scheme, transmitted over a complex AWGN channel and coherently detected. For simplicity of notation, we treat the channel inputs and outputs as two dimensional real vectors. Let $M = 2^k$ (where $k \in \mathbb{N}$) be the modulation parameter, denote the input to the channel by $\mathbf{X} = (x_1, x_2)$ where the possible input values are given by

$$\mathbf{X}_k = (\cos\theta_k, \sin\theta_k), \quad \theta_k \triangleq \frac{(2k+1)\pi}{M}, \quad k = 0, 1, \ldots, M-1. \tag{49}$$

We denote the channel output by $\mathbf{Y} = (y_1, y_2)$ where $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, and $\mathbf{N} = (n_1, n_2)$ is a Gaussian random vector with i.i.d. components each with zero-mean and variance σ^2. The conditional *pdf* of the channel output, given the transmitted symbol $\mathbf{X}_k$, is given by

$$\begin{aligned} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X}_k) &= \frac{1}{2\pi\sigma^2}\, e^{-\frac{(y_1-\cos\theta_k)^2+(y_2-\sin\theta_k)^2}{2\sigma^2}} \\ &= \frac{1}{2\pi\sigma^2}\, e^{-\frac{\|\mathbf{Y}-\mathbf{X}_k\|^2}{2\sigma^2}}, \quad \mathbf{Y} \in \mathbb{R}^2 \end{aligned} \tag{50}$$

where $\|\cdot\|$ designates the L_2 norm.

Due to the symmetry of the channel, then for every $0 < s < 1$, the optimal input distribution which maximizes the error exponent of the sphere-packing bound is given by $q_{k,s} = \frac{1}{M}$ for $k \in \{0, 1, \ldots, M-1\}$ and $s \in (0, 1)$. Hence, the ISP bound derived in Section 3.1 can be applied to lower bound the decoding error probability of M-ary PSK block coded modulated schemes whose transmission takes place over the AWGN channel. To this end, we derive in Appendix A the function $\mu_0(s, f_s)$ defined in (36) and calculate its first and second derivatives w.r.t. s. The final expressions for these functions are given below.

$$\begin{aligned}
\mu_0(s, f_s) &= (1-s) \ln\left(\theta(s)\right) \\
\mu_0'(s, f_s) &= -\ln(\theta(s)) + (1-s)\left(\frac{\beta(s)+\gamma(s)}{\theta(s)}\right) \\
\mu_0''(s, f_s) &= -2\,\frac{\beta(s)+\gamma(s)}{\theta(s)} + (1-s)\left[\frac{\beta'(s)+\gamma'(s)}{\theta(s)} - \left(\frac{\beta(s)+\gamma(s)}{\theta(s)}\right)^2\right]
\end{aligned}$$

where $\theta(s)$, $\beta(s)$, $\gamma(s)$, $\beta'(s)$ and $\gamma'(s)$ are introduced in Appendix A (see (A.1), (A.4), (A.5), (A.7) and (A.8), respectively).

The ISP bound is calculated by applying the equations above to Theorem 3.1. The above expressions can also be applied towards the calculation of the VF bound for this scenario. In Section 5, we present some numerical results which compare the tightness of the VF, ISP and SP59 bounds for M-ary PSK modulated signals transmitted over the AWGN channel.

4 The 1959 Sphere-Packing Bound of Shannon and Improved Algorithms for Its Calculation

The 1959 sphere-packing (SP59) bound derived by Shannon [13] provides a lower bound on the decoding error probability of an arbitrary block code whose transmission takes place over an AWGN channel. We begin this section by introducing the SP59 bound in its original form, along with asymptotic approximations derived in [13] which facilitate the estimation of the bound for large block lengths. We then review a theorem, introduced by Valembois and Fossorier [18], presenting a set of recursive equations which simplify the calculation of the bound. Both the original formula for the SP59 bound in [13] and the recursive method in [18] perform the calculation of the bound in the probability domain; this leads to various problems of over and under flows when calculating the exact value of the bound for codes with block lengths of $N = 1000$ or more. In this section, we present a theorem which facilitates the calculation of the SP59 bound in the logarithmic domain. This virtually eliminates the possibility of numerical errors in the calculation.

4.1 The 1959 Sphere-Packing Bound and Asymptotic Approximations [13]

Consider a block code $\mathcal{C}$ of length N, and rate R nats per channel use per dimension. It is assumed that all the codewords are mapped to signals with equal energy (e.g. PSK modulation); hence, all the signals representing codewords lie on an N-dimensional sphere centered at the origin, but finer details of the modulation used are not taken into account. This assumption implies that every Voronoi cell (i.e., the convex region containing all the points which are closer to the considered signal than to any other code signal) is a polyhedric cone which is limited by at most $\exp(NR) - 1$ hyper planes intersecting at the origin. As a measure of volume, Shannon introduces the solid angle of a cone which is defined to be the area of the sphere of unit radius which is cut out by the cone. Since the Voronoi cells partition the space $\mathbb{R}^N$, the sum of their

solid angles must be the area of an N-dimensional sphere of unit radius. The derivation of the SP59 bound relies on two main observations:

- Among the cones of a given solid angle, the lowest probability of error is given by the circular cone whose axis connect the code signal with the origin.
- It is best to share the total solid angle equally among the $\exp(NR)$ Voronoi regions.

As a corollary of these two observations follows the argument that the average Voronoi cell of any code cannot be better than a circular cone centered around the code signal with solid angle equal to $\exp(-NR)$ of the solid angle of $\mathbb{R}^N$. The solid angle of a circular cone is given by the following lemma.

Lemma 4.1 (Solid Angle of a Circular Cone [13]). The solid angle of a circular cone of half angle θ in $\mathbb{R}^N$ is given by

$$\Omega_N(\theta) = \frac{2\pi^{\frac{N-1}{2}}}{\Gamma(\frac{N-1}{2})} \int_0^\theta (\sin\phi)^{N-2}\, d\phi \ .$$

In particular, the solid angle of $\mathbb{R}^N$ is given by

$$\Omega_N(\pi) = \frac{2\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \ .$$

Theorem 4.1 (The 1959 Sphere-Packing (SP59) Bound [13]). Assume the transmission of an arbitrary block code of length N and rate R takes place over an AWGN channel with noise spectral density $\frac{N_0}{2}$. Then, under ML decoding, the error probability is lower bounded by

$$P_{\rm e}(\mathrm{ML}) > P_{\rm SPB}(N,\theta,A) \ , \quad A \triangleq \sqrt{\frac{2E_{\rm s}}{N_0}}$$

where $E_{\rm s}$ is the average energy per symbol, $\theta \in [0,\pi]$ satisfies the inequality $2^{-NR} \le \frac{\Omega_N(\theta)}{\Omega_N(\pi)}$,

$$\begin{aligned} P_{\rm SPB}(N,\theta,A) \ &\triangleq \ \frac{(N-1)e^{-\frac{NA^2}{2}}}{\sqrt{2\pi}} \int_\theta^{\frac{\pi}{2}} (\sin\phi)^{N-2}\, f_N(\sqrt{N}A\cos\phi)\, d\phi \\ &\quad + Q(\sqrt{N}A). \end{aligned} \tag{51}$$

and

$$f_N(x) \triangleq \frac{1}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \int_0^\infty z^{N-1} \exp\left(-\frac{z^2}{2} + zx\right)\, dz \ , \quad \forall\, x \in \mathbb{R},\ N \in \mathbb{N}. \tag{52}$$

By assumption, the transmitted signal is represented by a point which lies on the N-dimensional sphere of radius $\sqrt{NE_s}$ and which is centered at the origin, and the Gaussian noise is additive. The value $P_{\rm SPB}(N,\theta,A)$ in the RHS of (4.1) designates the probability that the received vector falls outside the N-dimensional circular cone of half angle θ whose main axis passes through the origin and the signal point which is represented by the transmitted signal. Hence, this function is monotonically decreasing in θ. The tightest lower bound on the decoding error probability is therefore achieved for $\theta_1(N,R)$ which satisfies

$$\frac{\Omega_N\big(\theta_1(N,R)\big)}{\Omega_N(\pi)} = \exp(-NR).$$

The calculation of $\theta_1(N,R)$ can become quite tedious. In order to simplify the calculation of the SP59 bound, [13] provides asymptotically tight upper and lower bounds on the ratio $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$.

Lemma 4.2 (Bounds on the Solid Angle [13]). The solid angle of a circular cone of half angle θ in the Euclidean space $\mathbb{R}^N$ satisfies the inequality

$$\frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\left(1-\frac{\tan^2\theta}{N}\right) \leq \frac{\Omega_N(\theta)}{\Omega_N(\pi)} \leq \frac{\Gamma(\frac{N}{2})(\sin\theta)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta}\ .$$

Corollary 4.1 (SP59 Bound (Cont.)). If θ^* satisfies the equation

$$\frac{\Gamma(\frac{N}{2})(\sin\theta^*)^{N-1}}{2\Gamma(\frac{N+1}{2})\sqrt{\pi}\cos\theta^*}\left(1-\frac{\tan^2\theta^*}{N}\right) = \exp(-NR) \tag{53}$$

then $\frac{\Omega_N(\theta^*)}{\Omega_N(\pi)} \geq \exp(-NR)$, and therefore

$$P_{\mathrm{e}}(\mathrm{ML}) > P_{\mathrm{SPB}}(N,\theta^*,A). \tag{54}$$

The use of θ^* instead of the optimal value $\theta_1(N,R)$ causes some loss in the tightness of the SP59 bound. However, due to the asymptotic tightness of the bounds on $\frac{\Omega_N(\theta)}{\Omega_N(\pi)}$, the loss in the tightness of the bound in Corollary 4.1 vanishes asymptotically as $N\to\infty$. In [18], it was numerically observed that the loss is marginal even for relatively small values of N and R. It was observed that the loss is smaller then 0.01 dB whenever the dimension of the code NR is greater than 20, and it becomes smaller then 0.001 dB when the dimension of the code exceeds 60.

For large block lengths, the calculation of the SP59 becomes extremely difficult due to over and under flows in the floating point operations. However, [13] presents some asymptotic formulas which give a very accurate estimation of the bound for large enough block lengths. These approximations allow the calculation to be made in the logarithmic domain which virtually eliminates the possibility of floating point errors.

Theorem 4.2. [13]: Defining

$$\begin{aligned} G(\theta) &\triangleq \frac{A\cos\theta+\sqrt{A^2\cos^2\theta+4}}{2} \\ E_L(\theta) &\triangleq \frac{A^2-AG(\theta)\cos\theta-2\ln\big(G(\theta)\sin\theta\big)}{2} \end{aligned}$$

then

$$P_{\mathrm{SPB}}(N,\theta,A) \geq \frac{\sqrt{N-1}}{6N(A+1)}e^{\frac{-(A+1)^2+3}{2}}e^{-N\,E_L(\theta)}. \tag{55}$$

This lower bound is valid for any block length N. However, the ratio of the left and right terms in (55) stays bounded away from one for all N. A more accurate approximation of $P_{\mathrm{SPB}}(N,\theta,A)$ is given by the next theorem, but without a determined inequality. As a consequence, the following approximation is not a proven theoretical lower bound on the error probability. For $N>1000$, however, its numerical values become almost identical to those of the exact bound, thus giving a useful estimation for the lower bound.

Theorem 4.3. [13]: Using the notation of Theorem 4.2, if $\theta > \cot^{-1}(A)$, then

$$P_{\mathrm{SPB}}(N,\theta,A) \approx \frac{\alpha(\theta)e^{-NE_L(\theta)}}{\sqrt{N}}$$

where

$$\alpha(\theta) \triangleq \left(\sqrt{\pi\,(1+G(\theta)^2)}\ \sin\theta\ \big(AG(\theta)\sin^2\theta-\cos\theta\big)\right)^{-1}.$$

4.2 An Algorithm for Calculating the 1959 Sphere-Packing Bound [18]

In [18, Section 2], Valembois and Fossorier review the SP59 bound and suggest a recursive algorithm to simplify its calculation. This algorithm is presented in the following theorem:

Theorem 4.4 (Recursive Equations for Simplifying the Calculation of the SP59 Bound). [18, Theorem 3]: The set of functions $\{f_N\}$ introduced in (52) can be expressed in the alternative form

$$f_N(x) = P_N(x) + Q_N(x)\exp(\frac{x^2}{2})\int_{-\infty}^{x}\exp(-\frac{t^2}{2})\,dt\,, \quad x\in\mathbb{R},\ N\in\mathbb{N} \tag{56}$$

where P_N and Q_N are two polynomials, determined by the same recursive equation for all $N \geq 5$

$$\begin{aligned} P_N(x) &= \frac{2N-5+x^2}{N-1}\,P_{N-2}(x) - \frac{N-4}{N-1}\,P_{N-4}(x)\,, \\ Q_N(x) &= \frac{2N-5+x^2}{N-1}\,Q_{N-2}(x) - \frac{N-4}{N-1}\,Q_{N-4}(x) \end{aligned} \tag{57}$$

with the initial conditions

$$\begin{aligned} &P_1(x) = 0, \quad Q_1(x) = 1 \\ &P_2(x) = \sqrt{\frac{2}{\pi}}, \quad Q_2(x) = \sqrt{\frac{2}{\pi}}\,x \\ &P_3(x) = \frac{x}{2}, \quad Q_3(x) = \frac{1+x^2}{2} \\ &P_4(x) = \sqrt{\frac{2}{\pi}}\frac{2+x^2}{3}, \quad Q_4(x) = \sqrt{\frac{2}{\pi}}\frac{3x+x^3}{3}. \end{aligned}$$

By observing the recursive equations for P_N and Q_N in (57), it can be noticed that the coefficients of the higher powers of x vanish exponentially as N increases. When performing the calculation using double-precision floating point numbers, these coefficients cause underflows when N is larger than several hundreds, and are replaced by zeros. Examining the expression for $P_{\mathrm{SPB}}(N,\theta,A)$ in (51), we observe that $f_N(x)$ (and therefore the polynomials $P_N(x)$ and $Q_N(x)$) is evaluated at $x \sim O(\sqrt{N})$. Hence, the replacement of the coefficients of the high powers of x by zeros causes a considerable inaccuracy in the calculation of P_{SPB} in (51). To exemplify the effect of these underflows, we study the coefficients of $P_{750}(x)$ as calculated using double precision floating point numbers. In this case, the coefficients of all the powers higher than 400 have caused underflows and have been replaced by zeros. The left plot of Figure 2 shows the coefficients of $P_{750}(x)$. Since $f_N(x)$ is evaluated at $x \sim O(\sqrt{N})$, one should examine the coefficients of $\tilde{P}_{750}(x) \triangleq P_{750}(\sqrt{750}\ x)$ which are plotted in the right plot of Figure 2. It can be seen that the dominant coefficients are those multiplying the powers of x between 400 and 520 which, as mentioned above, have been replaced by zeros due to underflows. This demonstrates the inaccuracy due to underflows in the coefficients of the high powers. To avoid this loss of dominant coefficients, it is possible to modify the recursive equations (57) in order to calculate the polynomials $\tilde{P}_N(x) \triangleq P_N(\sqrt{N}\ x)$ and $\tilde{Q}_N(x) \triangleq Q(\sqrt{N}\ x)$. However, as can be observed from the right plot of Figure 2, these coefficients become extremely large and cause overflows when N approaches 1000.

Considering the integrand in the RHS of (51), reveals another difficulty in calculating the SP59 bound for large values on N. For these values, the term $f_N(\sqrt{N}A\cos\phi)$ becomes very large and causes overflows, while the value of the term $(\sin\phi)^{N-2}$ becomes very small and causes underflows; this causes a "$0\cdot\infty$" phenomenon when evaluating the integrand at the RHS of (51).

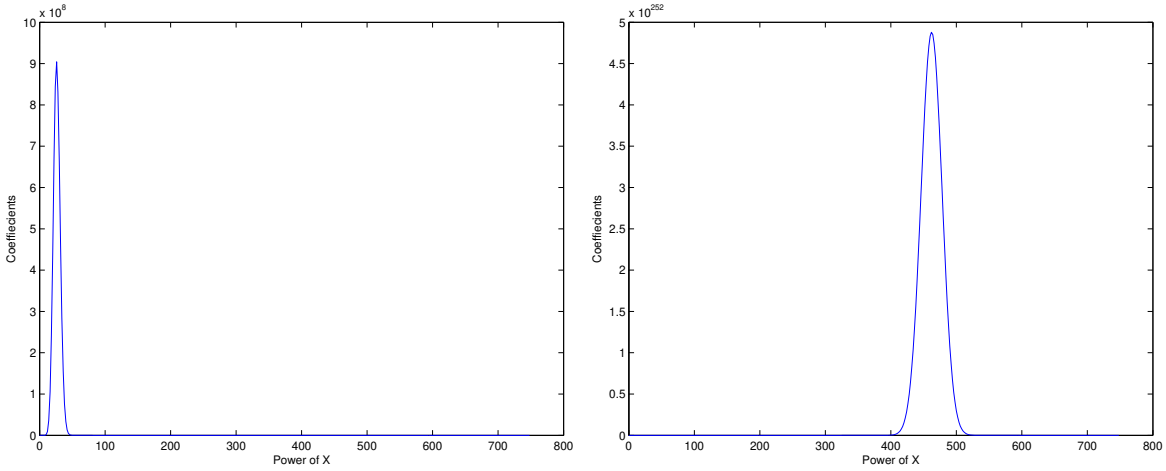

Figure 2: Coefficients of the polynomials $P_{750}(x)$ (right plot) and $\tilde{P}_{750}(x) = P_{750}(\sqrt{750}\,x)$ (left plot). Since the polynomials are even, only the coefficients multiplying the even powers of x have been plotted. It can be observed that in the right plot, the coefficients of powers of x between 400 and 520 are dominant. These coefficients have caused underflows in the calculation of $P_{750}(x)$ in the left plot.

4.3 A Log-Domain Approach for Computing the 1959 Sphere-Packing Bound

In this section, we present a method which enables the entire calculation of the integrand in the RHS of (51) in the log domain, thus circumventing the numerical over and under flows which become problematic in the calculation of the SP59 bound for large block lengths. We begin our derivation by representing the set of functions $\{f_N\}$ defined in (52) as a sum of exponents.

Proposition 4.1. The set of functions $\{f_N\}$ in (52) can be expressed in the form

$$f_N(x) = \sum_{j=0}^{N-1} \exp\big(d(N,j,x)\big)\ , \quad x \in \mathbb{R},\ N \in \mathbb{N}$$

where

$$\begin{aligned} d(N,j,x) \ \triangleq\ & \frac{x^2}{2} + \ln\Gamma\left(\frac{N}{2}\right) - \ln\Gamma\left(\frac{j}{2}+1\right) - \ln\Gamma(N-j) \\ & + (N-1-j)\ln\left(\sqrt{2}\,x\right) - \frac{\ln 2}{2} \\ & + \ln\left[1 + (-1)^j\ \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right]\ , \qquad \begin{array}{c} N \in \mathbb{N},\ x \in \mathbb{R} \\ j = 0, 1 \ldots, N-1 \end{array} \end{aligned} \tag{58}$$

and

$$\Gamma(a) \ \triangleq\ \int_0^\infty t^{a-1} e^{-t} dt\ , \quad \mathrm{Re}(a) > 0 \tag{59}$$

$$\tilde{\gamma}(x,a) \ \triangleq\ \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt\ , \quad x \in \mathbb{R},\ \mathrm{Re}(a) > 0 \tag{60}$$

designate the complete and incomplete Gamma functions, respectively.

Proof. The proof is given in Appendix B. □

Remark 4.1. It should be noted that the exponents $d(N, j, x)$ in (58) can be readily calculated by using standard mathematical functions. The function which calculates the natural logarithm of the Gamma function is implemented in the MATLAB software by `gammaln`, and in the Mathematica software by `LogGamma`. The function $\tilde{\gamma}(a, b)$ is implemented in MATLAB by `gammainc(x,N)` and in Mathematica by `GammaRegularized(N,0,x)`.

In order to perform the entire calculation of the function f_N in the log domain, we employ the function

$$\max{}^*(x_1, \ldots, x_m) \triangleq \ln\left(\sum_{i=1}^{m} e^{x_i}\right) , \quad m \in \mathbb{N}, \quad x_1, \ldots, x_m \in \mathbb{R} \tag{61}$$

which is commonly used in the implementation of the log-domain BCJR algorithm. The function $\max^*$ can be calculated in the log domain using the recursive equation

$$\max{}^*(x_1, \ldots, x_{m+1}) = \max{}^*\big(\max{}^*(x_1, \ldots, x_m), x_{m+1}\big) , \quad m \in \mathbb{N} \setminus \{1\},\ x_1, \ldots, x_{m+1} \in \mathbb{R}$$

with the initial condition

$$\max{}^*(x_1, x_2) = \max(x_1, x_2) + \ln\left(1 + e^{-|x_1 - x_2|}\right) .$$

By combining Proposition 4.1 and the definition of the function $\max^*$ in (61), we get a method of calculating the set of functions $\{f_N\}$ in the log domain.

Corollary 4.2. The set of functions $\{f_N\}$ defined in (52) can be rewritten in the form

$$f_N(x) = \exp\Big[\max{}^*\big(d(N, 0, x), d(N, 1, x), \ldots, d(N, N-1, x)\big)\Big] \tag{62}$$

where $d(N, j, x)$ is introduced in (58).

By combining (51) and (62), one gets the following theorem which provides an efficient algorithm for the calculation of the SP59 bound in the log domain.

Theorem 4.5 (Log domain calculation of the SP59 bound). The term $P_{\mathrm{SPB}}(N, \theta, A)$ in the RHS of (4.1) can be rewritten as

$$\begin{aligned} P_{\mathrm{SPB}}(N, \theta, A) \;=\; & \int_{\theta}^{\frac{\pi}{2}} \exp\Bigg[\ln(N-1) - \frac{NA^2}{2} - \frac{1}{2}\ln(2\pi) + (N-2)\ln\sin\phi \\ & \qquad + \max{}^*\Big(d(N, 0, \sqrt{N}A\cos\phi), \ldots, d(N, N-1, \sqrt{N}A\cos\phi)\Big)\Bigg] d\phi \\ & + Q(\sqrt{N}A) , \qquad N \in \mathbb{N},\ \theta \in [0, \tfrac{\pi}{2}],\ A \in \mathbb{R}^+ \end{aligned}$$

where $d(N, j, x)$ is defined in (58).

Using Theorem 4.5, it is possible to calculate the exact value of the SP59 lower bound for very large block lengths. Figure 3 shows a comparison of the exact value of the SP59 bound and its asymptotic value as given in Theorems 4.5 and 4.3, respectively. This comparison is shown for a code rate of 0.5 bits per channel use per dimension and block lengths of $N = 10^2, 10^3$ and 10^4. The calculations of the exact and asymptotic expressions were done using θ^* from (53); due to the large block lengths, the loss incurred by using this suboptimal value is negligible. It is observed that the asymptotic expression is indeed quite accurate for the two larger block lengths of $N = 1,000$ and $10,000$, and its accuracy is improved by increasing the block length and the SNR.

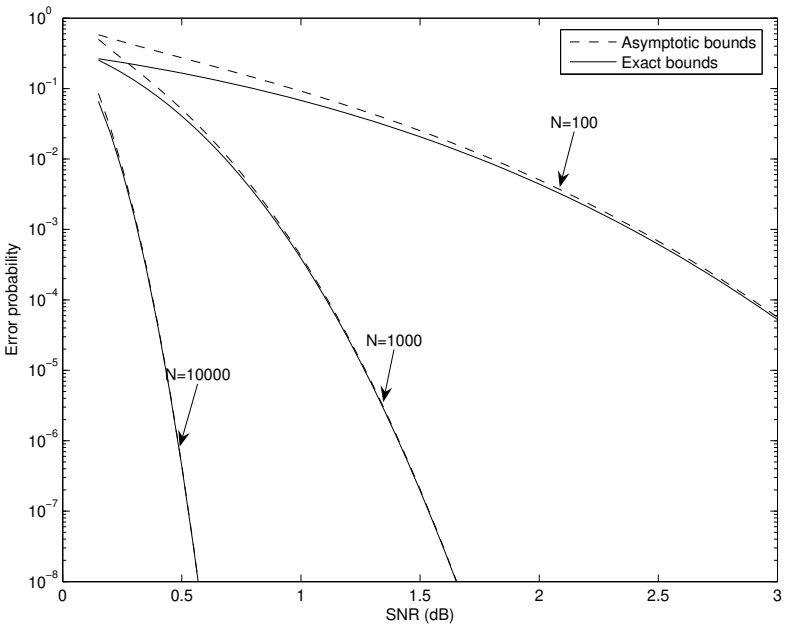

Figure 3: A comparison of the asymptotic and exact expressions for the SP59 bound (see Theorems 4.3 and 4.5, respectively). The examined block lengths are $N = 100, 1000$ and $10,000$ for a code rate of 0.5 bits per channel use per dimension.

5 Numerical Results

This section presents some numerical results which serve to demonstrate the improved tightness of the ISP bound derived in Section 3. We consider performance bounds for M-ary PSK block coded modulation with coherent detection over an AWGN channel, and for the binary erasure channel (BEC).

5.1 Performance Bounds for M-ary PSK Block Coded Modulation over the AWGN Channel

The ISP bound is particularized in Section 3.2 to M-ary PSK block coded modulation schemes whose transmission takes place over the AWGN channel, and where the received signals are coherently detected. The calculations of the function μ and its derivatives (see Appendix A) are useful for the calculation of the VF bound [18] as well. The SP59 bound reviewed in Section 4 provides a lower bound on the decoding error probability of M-ary PSK signaling over the AWGN channel, as a particular case of equi-energy signals. In the following, we exemplify the use of these lower bounds for the considered case. They are also compared to Gallager's random-coding upper bound [5] and the tangential-sphere upper bound [10] when applied to random block codes. This serves for the study of the tightness of the ISP bound (see Section 3) as compared to other upper and lower bounds. The numerical results shown in this section indicate that the recent variants of the SP67 bound provide an interesting alternative to the SP59 bound which is commonly used in the literature as an ultimate measure of performance for codes transmitted over the AWGN channel (see, e.g., [4, 7, 8, 9, 12, 16, 18, 19]). The advantage of the ISP bound over the VF bound in [18] is also exemplified in this section.

Fig. 4 presents a comparison of the SP59 bound [13], the VF bound [18], and the ISP bound derived in this paper (see Section 3). The comparison refers to block codes of length

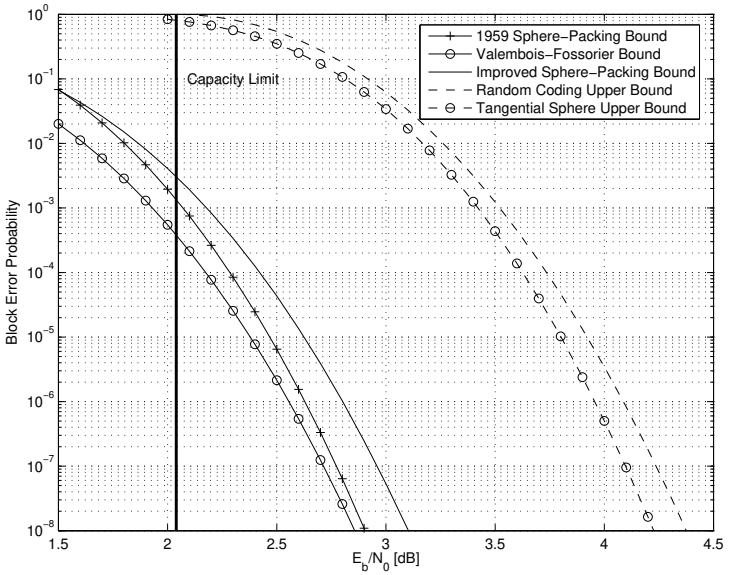

Figure 4: A comparison between upper and lower bounds on the ML decoding error probability for block codes of length $N = 500$ bits and code rate of $0.8\frac{\text{bits}}{\text{channel use}}$. This figure refers to BPSK modulated signals whose transmission takes place over an AWGN channel. The compared bounds are the 1959 sphere-packing (SP59) bound of Shannon [13], the Valembois-Fossorier (VF) bound [18], the improved sphere-packing (ISP) bound derived in Section 3, the random-coding upper bound of Gallager [5], and the tangential-sphere upper bound (TSB) of Poltyrev [6, 10] when applied to fully random block codes with the above block length and rate.

500 bits and rate $0.8\frac{\text{bits}}{\text{channel use}}$, which are BPSK modulated and transmitted over an AWGN channel. The plot also depicts the random-coding upper bound on the error probability, the tangential-sphere bound (TSB) [6, 10], and the capacity limit bound (CLB).[1] It is observed from this figure that even for relatively short block lengths, the ISP bound outperforms the SP59 bound for block error probabilities below 10^{-1}. For a block error probability of 10^{-5}, the ISP bound provides a gain of about 0.2 dB and 0.3 dB over the SP59 bound and the VF bound, respectively. For these code parameters, the TSB provides a tighter upper bound on the block error probability of random codes than Gallager's random-coding bound; e.g., the gain of the TSB over Gallager's bound is about 0.2 dB for a block error probability of 10^{-5}. Note that the random coding bound of Gallager is tighter than the TSB for large enough block lengths, as the latter bound does not reproduce the random coding error exponent for the AWGN channel [10]. However, this figure exemplifies the advantage of the TSB over the random coding bound of Gallager, when particularized to random block codes of relatively short block lengths; this advantage is especially pronounced for low code rates where the gap between the error exponents of these two bounds is reduced (see [12, p. 67]), but it is also reflected from Figure 4 for BPSK modulation with a code rate of $0.8\frac{\text{bits}}{\text{channel use}}$. The gap between the TSB and the ISP bound, as upper and lower bounds respectively, is less than 1.2 dB for all block error probabilities lower than 10^{-1}. Also, the ISP bound is more informative than the CLB for block error probabilities below 3×10^{-3}.

[1] Although the CLB refers to the asymptotic case where the block length tends to infinity, it is plotted in [18] and here as a reference, in order to examine whether the improvement in the tightness of the ISP is for rates above or below capacity.

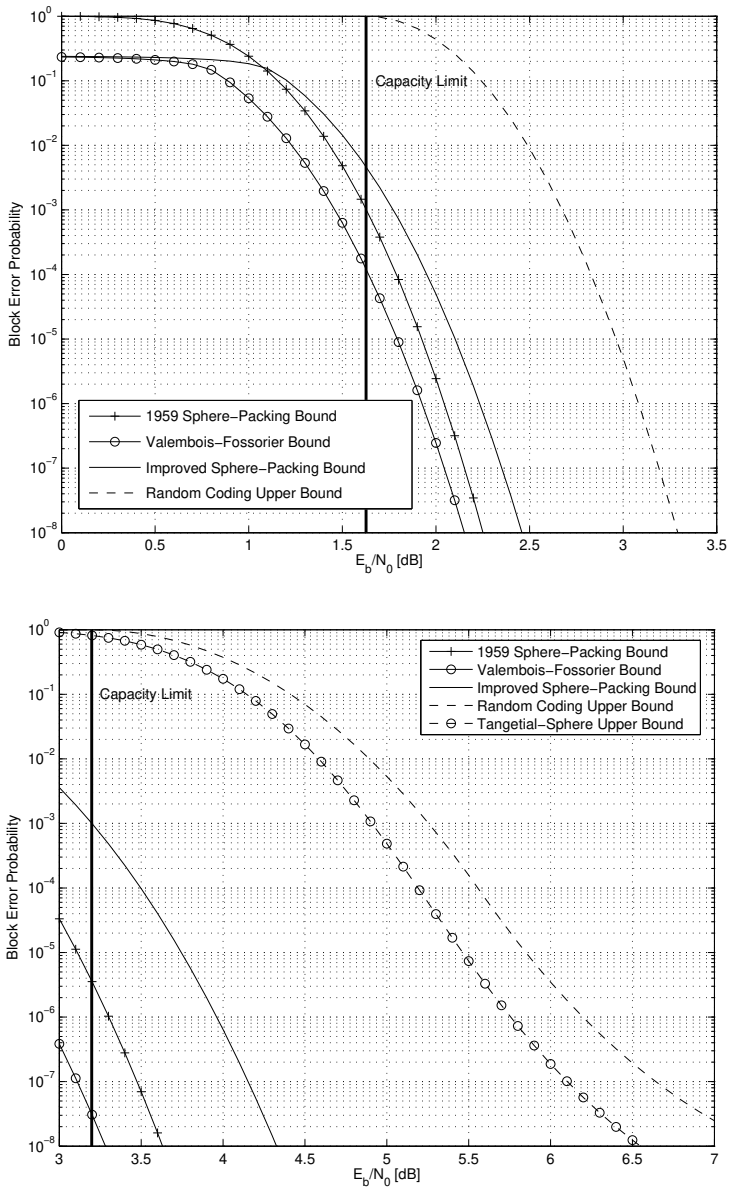

Figure 5: A comparison between upper and lower bounds on the ML decoding error probability, referring to short block codes which are QPSK modulated and transmitted over the AWGN channel. The compared lower bounds are the 1959 sphere-packing (SP59) bound of Shannon [13], the Valembois-Fossorier (VF) bound [18], and the improved sphere-packing (ISP) bound derived in Section 3; the compared upper bounds are the random-coding upper bound of Gallager [5] and the tangential-sphere bound (TSB) of Poltyrev [10]. The upper plot refers to block codes of length $N = 1024$ which are encoded by 768 information bits (so the rate is $1.5 \frac{\text{bits}}{\text{channel use}}$), and the lower plot refers to block codes of length $N = 300$ which are encoded by 270 bits whose rate is therefore $1.8 \frac{\text{bits}}{\text{channel use}}$.

Fig. 5 presents a comparison of the SP59, VF and ISP bounds where QPSK modulated signals are considered. The two plots in this figure refer to codes of short block lengths. The plots also depict the random-coding upper bound on the block error probability, the TSB of

Poltyrev [10], and the CLB. It can be observed from these two plots that even for relatively short block lengths, the ISP bound outperforms the SP59 bound for all error probabilities below 10^{-1} (this result is consistent with the upper plot of Fig. 9). In the upper plot of Fig. 5, referring to a block length of 1024 bits (i.e., 512 QPSK symbols) and a rate of $1.5\frac{\text{bits}}{\text{channel use}}$, it is observed that for a block error probability of 10^{-5}, the ISP bound provides a gain of about 0.2 dB and 0.3 dB over the SP59 and the VF bounds, respectively. The gap between the ISP lower bound and the random-coding upper bound is 0.85 dB for all block error probabilities lower than 10^{-1}. In the lower plot of Fig. 5, referring to a block length of 300 bits and a rate of $1.8\frac{\text{bits}}{\text{channel use}}$, the ISP bound improves significantly the SP59 bound and the VF bound (for a block error probability of 10^{-5}, the improvement in the tightness of the ISP over the SP59 and VF bounds is 0.7 dB and 1 dB, respectively). Additionally, the ISP bound is more informative than the CLB for block error probabilities below 10^{-3}, where the SP59 and VF bound outperform the capacity-limit only for block error probabilities of 3×10^{-6} and 2×10^{-8}, respectively. For random block codes of block length $N = 300$ and rate $1.8\frac{\text{bits}}{\text{channel use}}$ which are QPSK modulated with Gray's mapping and transmitted over the AWGN channel, the TSB [10] is tighter than the random coding bound (see the lower plot in Fig. 5 and the explanation referring to Fig. 4). The gap between the ISP bound and the TSB in this plot is about 1.7 dB for a block error probability of 10^{-5} (as compared to gaps of 2.4 dB (2.7 dB) between the TSB and the SP59 (VF) bound).

Figure 6 presents a comparison of the bounds for codes of block length 5580 bits and information block length of 4092, where both QPSK (upper plot) and 8-PSK (lower plot) constellations are considered. The modulated signals correspond to 2790 and 1680 symbols, respectively, and the code rates for these constellations are equal to 1.467 and 2.2 bits per channel use, respectively. For this larger block length and for both constellations, both of the SP67-based bounds (i.e., the VF and the ISP bounds) outperform the SP59 for all block error probabilities below 10^{-1}; the ISP bound gives a gain of 0.1 dB and 0.2 dB over the VF bound for the QPSK and 8-PSK constellations, respectively. For both modulations, the gap between the ISP lower bound and the random-coding upper bound of Gallager does not exceed 0.4 dB. In [3], Divsalar and Dolinar design codes with the considered parameters by using concatenated Hamming and Accumulator codes. They also present computer simulations of the performance of these codes under iterative decoding, when the transmission takes place over the AWGN and several common modulation schemes are applied. For an error probability of 10^{-4}, the gap between the simulated performance of these codes under iterative decoding, and the ISP lower bound, which gives an ultimate lower bound on the error probability of optimally designed codes under ML decoding, is approximately 1.4 dB for QPSK and 1.6 dB for 8-PSK signaling. This provides an indication on the performance of codes defined on graphs and their iterative decoding algorithms, especially in light of the feasible complexity of the decoding algorithm which is linear in the block length. To conclude, it is reflected from the results plotted in Fig. 6 that a gap of about 1.5 dB between the ISP lower bound and the performance of the iteratively decoded codes in [3] is mainly due to the imperfectness of these codes and their sub-optimal iterative decoding algorithm; this conclusion follows in light of the fact that for random codes of the same block length and rate, the gap between the ISP bound and the random coding bound is reduced to less than 0.4 dB.

While it was shown in Section 3 that the ISP bound is uniformly tighter than the VF bound (which in turn is uniformly tighter than the SP67 bound [14]), no such relations are shown between the SP59 bound and the recent improvements on the SP67 bound (i.e., the VF and ISP bounds). Fig. 7 presents regions of code rates and block lengths for which the ISP bound outperforms the SP59 bound and the CLB; it refers to BPSK modulated signals transmitted over the AWGN and considers block error probabilities of 10^{-4}, 10^{-5} and 10^{-6}.

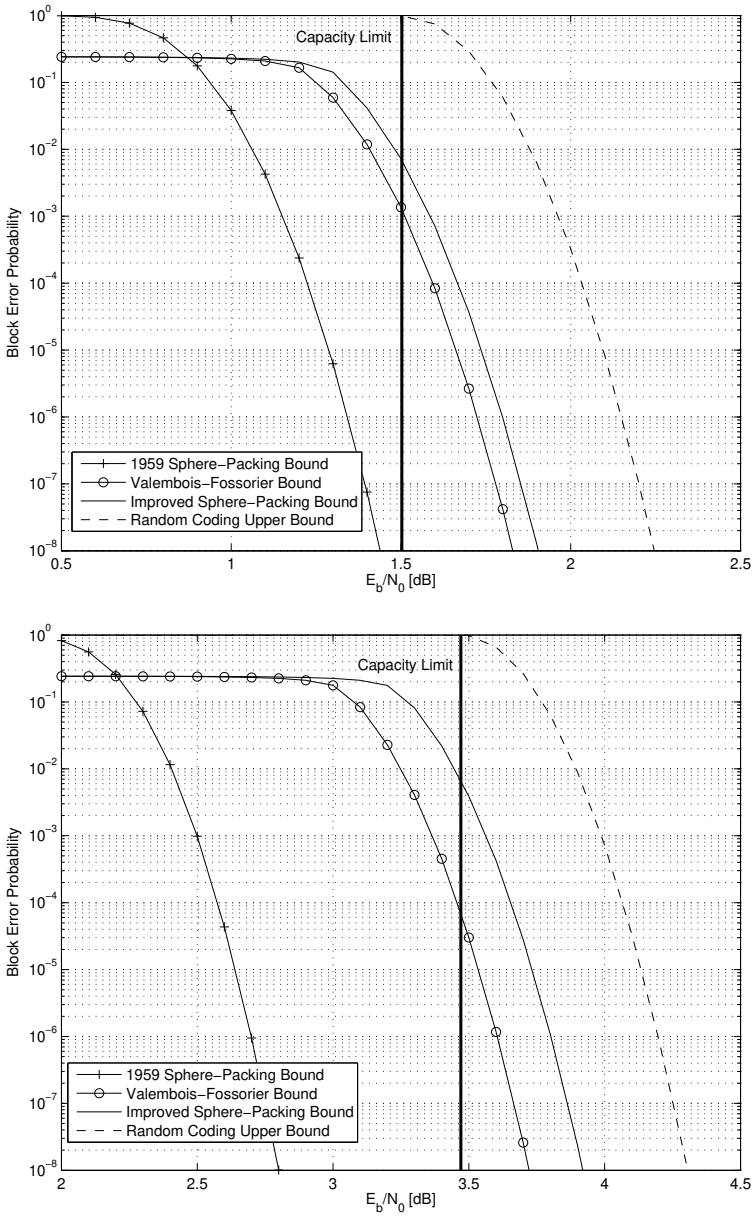

Figure 6: A comparison of upper and lower bounds on the ML decoding error probability for block codes of length $N = 5580$ bits and information block length of 4092 bits. This figure refers to QPSK (upper plot) and 8-PSK (lower plot) modulated signals whose transmission takes place over an AWGN channel; the rates in this case are 1.467 and 2.200 $\frac{\text{bits}}{\text{channel use}}$, respectively. The compared bounds are the 1959 sphere-packing (SP59) bound of Shannon [13], the Valembois-Fossorier (VF) bound [18], the improved sphere-packing (ISP) bound derived in Section 3, and the random-coding upper bound of Gallager [5].

It is reflected from this figure that for any rate $0 < R < 1$, there exists a block length $N(R)$ such that the ISP bound outperforms the SP59 bound for block lengths larger than $N(R)$ (the same property holds for the VF bound, but that the value of $N(R)$ depends on the considered SP67-based bound, and is significantly larger in the latter case). It is also observed that the

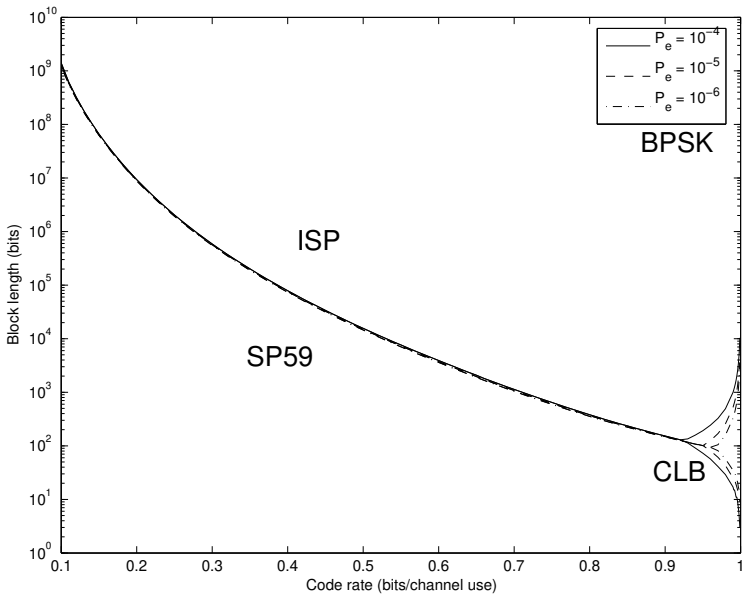

Figure 7: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for three different targets of block error probability (P_{e}). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [13], the improved sphere-packing (ISP) bound derived in Section 3, and the capacity-limit bound (CLB). The plot refers to BPSK modulated signals whose transmission takes place over the AWGN channel, and the considered code rates lie in the range between 0.1 and $1 \frac{\text{bits}}{\text{channel use}}$.

value $N(R)$ is monotonically decreasing with R, and it approaches infinity as we let R tend to zero. An intuitive explanation for this behavior can be given by considering the capacity limits of the binary-input and the energy constraint AWGN channels. For any value $0 \leq C < 1$, denote by $\frac{E_{\mathrm{b},1}(C)}{N_0}$ and $\frac{E_{\mathrm{b},2}(C)}{N_0}$ the values of $\frac{E_{\mathrm{b}}}{N_0}$ required to achieve a channel capacity of C bits per channel use for the binary-input and unconstrained-input AWGN channel, respectively (note that in the Gaussian regime, the un-constrained input distribution is also Gaussian). For any $0 \leq C < 1$, clearly $\frac{E_{\mathrm{b},1}(C)}{N_0} \geq \frac{E_{\mathrm{b},2}(C)}{N_0}$; however, the difference between these values is monotonically increasing with the capacity C, and, on the other hand, this difference approaches zero as we let C tend to zero. Since the SP59 bound only constrains the signals to be of equal energy, it gives a measure of performance for the energy constraint AWGN channel, where the SP67-based bounds consider the actual modulation and therefore refer to the binary-input AWGN channel. As the code rates become higher, the difference in the ultimate performance between the two channels is larger, and therefore the SP67 based techniques outperform the SP59 bound for smaller block lengths. For low code rates, the difference between the channels is smaller, and the SP59 outperforms the SP67 based bounding techniques even for larger block lengths due to the superior bounding technique which is specifically tailored for the AWGN channel. Fig 8 presents the regions of code rates and block lengths for which the VF bound (upper plot) and the ISP bound (lower plot) outperform the CLB and the SP59 bound when the the signals are BPSK modulated and transmitted over the AWGN channel; block error probabilities of 10^{-4}, 10^{-5} and 10^{-6} are examined. This figure is focused on high code rates, where the performance of the SP67 based bounds and their advantage over the SP59 bound is most appealing. From Figure 8, we have that for a code rate of 0.75 bits per channel use and an error probability of 10^{-6}, the VF bound becomes tighter than the SP59 for block lengths

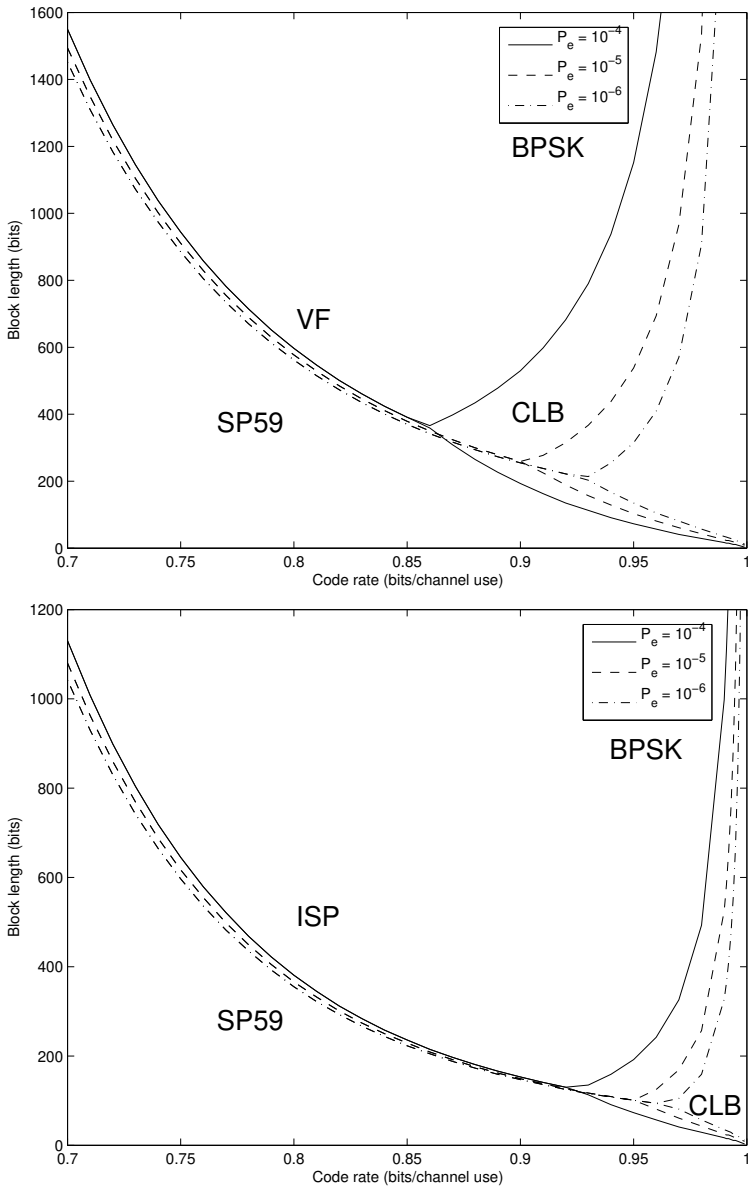

Figure 8: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for three different targets of block error probability (P_{e}). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [13], the capacity-limit bound (CLB), and the Valembois-Fossorier (VF) bound [18] (upper plot) or the improved sphere-packing (ISP) bound derived in Section 3 (lower plot). The plots refer to BPSK modulated signals whose transmission takes place over the AWGN channel, and the considered code rates lie in the range between 0.70 and $1 \frac{\text{bits}}{\text{channel use}}$.

exceeding 870 bits while the ISP bound reduces this value to 617 bits; moreover, when increasing the rate to 0.8 bits per channel use, the respective minimal block lengths reduce to 550 and 350 bits for the VF and ISP bounds, respectively. Fig 9 shows the regions of code rates and block lengths where the ISP outperform the CLB and SP59 bounds for QPSK (upper plot) and 8-PSK (lower plot) modulations. Comparing the lower plot of Fig. 8 which refers to BPSK modulation with the upper plot of Fig. 9 which refers to QPSK modulation, one can see that

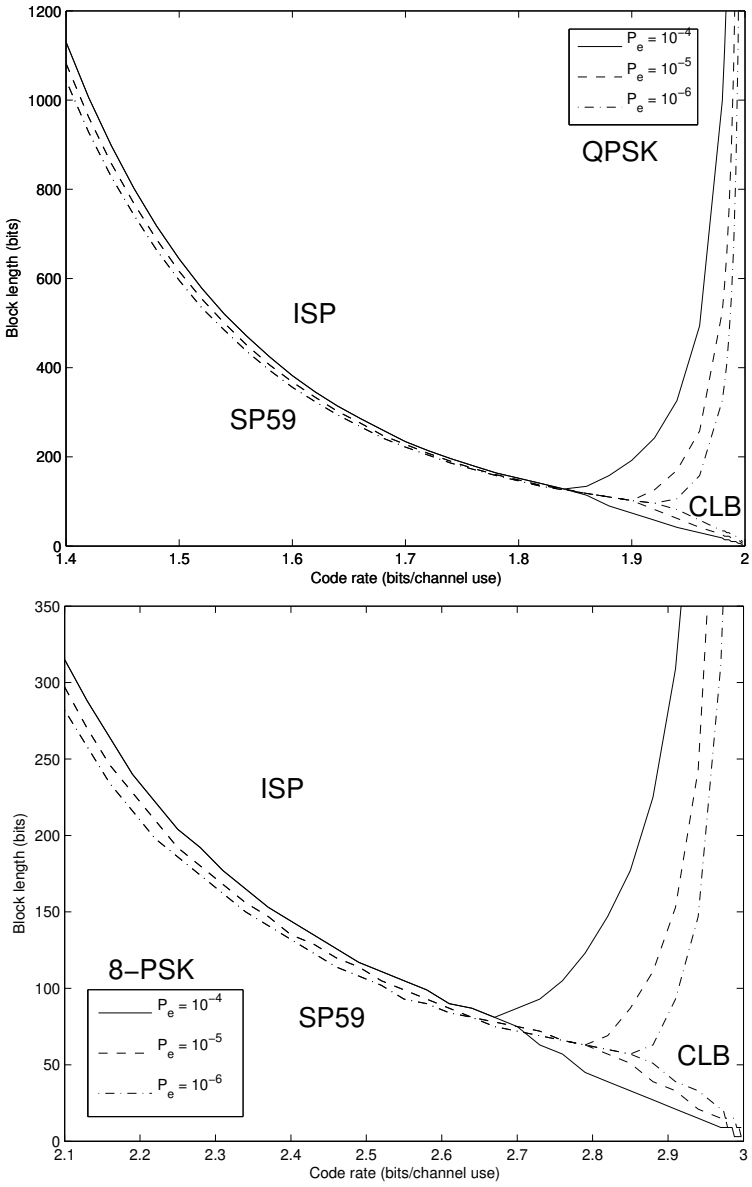

Figure 9: Regions in the two-dimensional space of code rate and block length, where a bound is better than the two others for three different targets of block error probability (P_{e}). The figure compares the tightness of the 1959 sphere-packing (SP59) bound of Shannon [13], the improved sphere-packing (ISP) bound derived in Section 3, and the capacity-limit bound (CLB). The plots refer to QPSK (upper plot) and 8-PSK (lower plot) modulated signals whose transmission takes place over the AWGN channel; the considered code rates lie in the range between 1.4 and $2\frac{\text{bits}}{\text{channel use}}$ for the QPSK modulated signals and between 2.1 and $3\frac{\text{bits}}{\text{channel use}}$ for the 8-PSK modulated signals.

the two graphs are virtually identical (when accounting for the doubling of the rate which is due to the use of both real and imaginary dimensions in the QPSK modulation). This is due to the fact that QPSK modulation posses no additional constraints on the channel and in fact, the real and imaginary planes can be serialized and decoded as in BPSK modulation. However, this property does not hold when replacing the ISP bound by the VF bound; this is due to the fact

the the VF bound considers a fixed composition subcode of the original code and the increased size of the alphabet causes a greater loss in the rate for QPSK modulation. When comparing the two plots of Fig. 9, it is evident that the block lengths for which the ISP bound becomes better than the SP59 bound decreases as the spectral efficiency of the modulation is increased (when normalizing the rate to units of information bits per code bit). An intuitive justification for this phenomenon is attributed to the fact that referring to the constellation points in (49), the mutual information between the code symbols in each dimension of the QPSK modulation is zero, while as the spectral efficiency of the PSK modulation is increased, the mutual information between the real and imaginary parts of each signal point is increased; thus, as the spectral efficiency is increased, this posses a stronger constraint on the possible positioning of the equal-energy signal points on the N-dimensional sphere. This intuition may suggest an explanation for the reason why as the spectral efficiency is increased, the advantage of the ISP bound which is exemplified for the M-ary PSK modulated signals over the SP59 bound which refers to the un-constrained input distribution holds even for smaller block lengths. This effect is expected to be more subtle for the VF bound since the increased alphabet size increases the reduction in the rate (by the quantity in (12)), which therefore causes the bound to be looser.

5.2 Performance Bounds for the Binary Erasure Channel

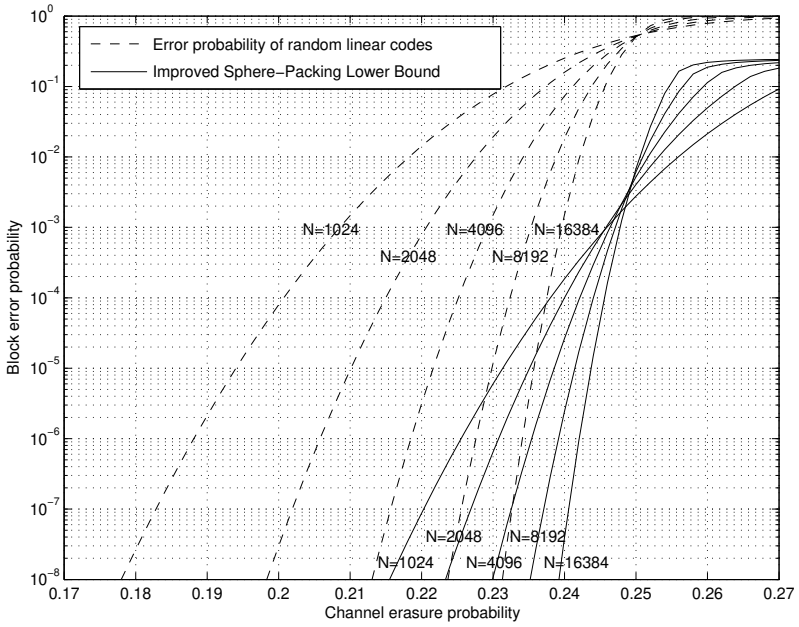

Figure 10: A comparison of the improved sphere-packing (ISP) lower bound from Section 3 and the exact decoding error probability of random binary linear block codes under ML decoding where the transmission takes place over the BEC (see [2, Eq. (3.2)]). The code rate examined is $0.75 \frac{\text{bits}}{\text{channel use}}$ and the block lengths are $N = 1024, 2048, 4096, 8192$ and 16384 bits.

In recent years, the BEC has been the focus of much attention in the field of iterative decoding techniques. The simplicity of this channel and the absolute reliability of the known values at the output lend themselves to a one-dimensional analysis of turbo-like codes and the performance of their iterative decoding algorithms in the case where the codes are transmitted over the BEC (see, e.g., [15]). For the asymptotic case where we let the block length tend

to infinity, several families which achieve the capacity of the BEC under iterative decoding have been constructed; these include low-density parity-check (LDPC), repeat-accumulate (RA) and accumulate-repeat-accumulate (ARA) codes. These discoveries motivate a study of the performance of iteratively decoded codes defined on graphs for moderate block lengths (see, e.g., [11]). In Figure 10, we compare the ISP lower bound and the exact block error probability of random linear block codes transmitted over the BEC as given in [2, Eq. (3.2)]). The figure refers to codes of rate 0.75 bits per channel use and various block lengths. It can be observed that for a block length of 1024 bits, the difference in the channel erasure probability for which the random coding bound and the ISP bound achieve an error probability of 10^{-5} is 0.037 while for a block length of 16384 bits, this gap is decreased to 0.008. This yields that the ISP bound is reasonably tight, and also suggests that this bound can be used in order to assess the imperfectness of turbo-like codes even for moderate block lengths.

6 Summary

This paper presents an improved sphere-packing (ISP) bound targeting codes of short to moderate block lengths, and it exemplifies some of its applications. The derivation of the ISP bound was stimulated by the remarkable performance and moderate complexity of turbo-like codes with short to moderate block lengths. We were motivated by recent improvements on the sphere-packing bound of [14] for finite block lengths, as suggested by Valembois and Fossorier in [18].

We first review the classical sphere-packing bounds, i.e., the 1959 sphere-packing bound (SP59) derived by Shannon for the Gaussian channel [13], and the 1967 sphere-packing (SP67) bound derived by Shannon, Gallager and Berlekamp for discrete memoryless channels [14]. The ISP bound, introduced in Section 3, is uniformly tighter than the classical SP67 bound [14] and the bound in [18]. Under a mild condition, the validity of the ISP bound is extended to general memoryless channels (even with continuous input and output alphabets); the basic observation which enables the derivation of the ISP bound is explained in Remark 3.1 (see p. 15).

We apply the ISP bound to M-ary PSK block coded modulation schemes whose transmission takes place over the AWGN channel and the received signals are coherently detected. The tightness of the ISP bound is exemplified by comparing it with upper and lower bounds on the ML decoding error probability and also with computer simulations of turbo-like codes under iterative decoding. The paper also presents a new algorithm which performs the entire calculation of the SP59 bound in the logarithmic domain, thus facilitating the exact calculation of the SP59 bound for all block lengths without the need for asymptotic approximations. It is shown that the ISP bound suggests an interesting alternative to the SP59 bound, where the latter is specialized for the AWGN channel.

High rate turbo-product codes with moderate block lengths (see [1]) exhibit a gap of 0.75–0.95 dB w.r.t. the information-theoretic limitation provided by the ISP bound. Based on numerical results in [17] for the ensemble of uniformly interleaved (1144, 1000) turbo-block codes whose components are random systematic linear block codes, the gap in $\frac{E_b}{N_0}$ between the ISP lower bound and an upper bound under ML decoding is 0.9 dB for a block error probability of 10^{-7}. These results exemplify the strength of the sphere-packing bounds for assessing the theoretical limitations of block codes and the power of iteratively decoded codes (see also [4, 7, 8, 12, 18]).

The ISP bound is especially attractive for block codes of high rate in terms of the range of the block lengths where this bound outperforms the SP59 bound and the capacity limit bound (see Figs. 5–9). Its effectiveness is especially pronounced for modulations of high spectral efficiency, due to the enhancement of its tightness as the size of the input alphabet is increased.

Appendices

Appendix A: Calculations Related to the VF and ISP Bounds for M-Ary PSK Modulated Schemes over the AWGN Channel

This appendix presents some technical calculations which yield the expressions for the function μ_0 defined in (36) and its derivatives. These expressions serve for the application of the VF bound in [18] and the ISP bound derived in Section 3 to block coded M-ary PSK modulation schemes transmitted over the AWGN channel and coherently detected.

From symmetry considerations, it is clear that the input distribution is uniform (i.e., $q_{k,s} = \frac{1}{M}$ for $k \in \{1, \ldots, M\}$ and $s \in (0,1)$.) Since the support of the vector $\mathbf{q}$ includes all the input alphabet, then from (36), the function $\mu_k(s, f_s)$ is independent of k. In the case of a continuous output alphabet, the sums in (36) are replaced by integrals, and the transition probabilities are replaced by transition probability density functions. Hence, we get by substituting (50) into (36) that

$$\mu_0(s, f_s) = (1-s) \ln \left(\theta(s) \right)$$

where

$$\theta(s) \triangleq \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} d\mathbf{Y}. \tag{A.1}$$

This can be rewritten in the form

$$\mu_0(s, f_s) \quad = \quad (1-s) \ln \left(\iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2}\ e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{-\frac{(1-s)\ (\|\mathbf{Y}-\mathbf{X}_k\|^2 - \|\mathbf{Y}-\mathbf{X}_0\|^2)}{2\sigma^2}} \right)^{\frac{1}{1-s}} d\mathbf{Y} \right).$$

By observing that

$$\|\mathbf{Y} - \mathbf{X}_k\|^2 - \|\mathbf{Y} - \mathbf{X}_0\|^2 = -2\langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle$$

which holds since $\|\mathbf{X}_l\|^2 = 1$ for every $l \in \{0, 1, \ldots, M-1\}$, we get

$$\mu_0(s, f_s) = (1-s) \ln \left(\iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2}\ e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} d\mathbf{Y} \right). \tag{A.2}$$

We now turn to calculate the derivative of μ_0 in (A.2) with respect to s while holding f_s constant, which gives

$$\begin{aligned} \frac{d}{ds}\mu_0(s, f_s) \quad = \quad & -\ln \left(\iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} d\mathbf{Y} \right) \\ & + (1-s) \frac{\displaystyle \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \frac{d}{ds} \left[\left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} \right] d\mathbf{Y}}{\displaystyle \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{1}{1-s}} d\mathbf{Y}}. \end{aligned}$$

To calculate $\frac{d}{ds}\left[\left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)^{\frac{1}{1-s}}\right]$, we apply the equality

$$\frac{d}{ds}\left(g(s)^{h(s)}\right) = h'(s)\ g(s)^{h(s)}\ \ln\big(g(s)\big) + h(s)\ g(s)^{h(s)-1}\ g'(s). \tag{A.3}$$

Straightforward calculus finally gives the equality

$$\frac{d}{ds}\mu_0(s,f_s) = -\ln(\theta(s)) + (1-s)\left(\frac{\beta(s)+\gamma(s)}{\theta(s)}\right)$$

where $\theta(s)$ is introduced in (A.1) and

$$\begin{aligned}\beta(s) \ &\triangleq \ \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \frac{1}{(1-s)^2} \ln\left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right) \\ &\qquad \cdot \left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)^{\frac{1}{1-s}} d\mathbf{Y} \end{aligned} \tag{A.4}$$

$$\begin{aligned}\gamma(s) \ &\triangleq \ \iint\limits_{\mathbb{R}^2} \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}} \frac{1}{1-s} \left(\frac{1}{M}\sum_{k=0}^{M-1} \frac{\langle \mathbf{Y},\mathbf{X}_0-\mathbf{X}_k\rangle}{\sigma^2}\ e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right) \\ &\qquad \left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)^{\frac{s}{1-s}} d\mathbf{Y}. \end{aligned} \tag{A.5}$$

Since the equality $\theta'(s) = \beta(s) + \gamma(s)$ holds for all s, then it gives

$$\begin{aligned}\frac{d^2}{ds^2}\mu_0(s,f_s) \ &= \ -\frac{d}{ds}\ln(\theta(s)) - \frac{\beta(s)+\gamma(s)}{\theta(s)} + (1-s)\frac{\partial}{\partial s}\left(\frac{\beta(s)+\gamma(s)}{\theta(s)}\right) \\ &= \ -2\ \frac{\beta(s)+\gamma(s)}{\theta(s)} + (1-s)\left[\frac{\beta'(s)+\gamma'(s)}{\theta(s)} - \left(\frac{\beta(s)+\gamma(s)}{\theta(s)}\right)^2\right]. \end{aligned} \tag{A.6}$$

We now calculate $\beta'(s)$, and get

$$\begin{aligned}\beta'(s) \ &= \ \frac{2\beta(s)}{1-s} + \frac{\gamma(s)}{1-s} + \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2} \left[\frac{1}{(1-s)^2}\ln\left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)\right]^2 \\ &\qquad \cdot \left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)^{\frac{1}{1-s}} d\mathbf{Y} \\ &\quad + \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2\ (1-s)^3} \ln\left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)\left(\frac{1}{M}\sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right)^{\frac{s}{1-s}} \\ &\qquad \cdot \left(\frac{1}{M}\sum_{k=0}^{M-1} \frac{\langle \mathbf{Y},\mathbf{X}_0-\mathbf{X}_k\rangle}{\sigma^2}\ e^{\frac{(1-s)\ \langle \mathbf{Y},\mathbf{X}_k-\mathbf{X}_0\rangle}{\sigma^2}}\right) d\mathbf{Y}\ . \end{aligned} \tag{A.7}$$

By calculating the derivative of the function γ in (A.5) with the aid of (A.3) gives

$$\begin{aligned}
\gamma'(s) &= \frac{\gamma(s)}{1-s} + \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2} \frac{1}{1-s} \left(\frac{1}{M} \sum_{k=0}^{M-1} \left(\frac{\langle \mathbf{Y}, \mathbf{X}_0 - \mathbf{X}_k \rangle}{\sigma^2} \right)^2 e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right) \\
&\qquad \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{s}{1-s}} d\mathbf{Y} \\
&\quad + \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2\ (1-s)^3} \left(\frac{1}{M} \sum_{k=0}^{M-1} \frac{\langle \mathbf{Y}, \mathbf{X}_0 - \mathbf{X}_k \rangle}{\sigma^2} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right) \\
&\qquad \ln \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right) \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{s}{1-s}} d\mathbf{Y} \\
&\quad + s \iint\limits_{\mathbb{R}^2} \frac{e^{-\frac{\|\mathbf{Y}-\mathbf{X}_0\|^2}{2\sigma^2}}}{2\pi\sigma^2} \left(\frac{1}{(1-s)\ M} \sum_{k=0}^{M-1} \frac{\langle \mathbf{Y}, \mathbf{X}_0 - \mathbf{X}_k \rangle}{\sigma^2} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^2 \\
&\qquad \left(\frac{1}{M} \sum_{k=0}^{M-1} e^{\frac{(1-s)\ \langle \mathbf{Y}, \mathbf{X}_k - \mathbf{X}_0 \rangle}{\sigma^2}} \right)^{\frac{2s-1}{1-s}} d\mathbf{Y}. \qquad \text{(A.8)}
\end{aligned}$$

Finally, the substitution of (A.1), (A.4), (A.5), (A.7) and (A.8) in (A.6) gives an explicit expression for the second derivative of μ_0.

Appendix B: Proof of Proposition 4.1

From the definition of f_N in (52), it follows that

$$\begin{aligned}
f_N(x) &= \frac{1}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \int_0^\infty z^{N-1} \exp(-\frac{z^2}{2} + zx)\ dz \\
&= \frac{e^{\frac{x^2}{2}}}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \int_0^\infty z^{N-1} \exp\left(-\frac{(z-x)^2}{2}\right)\ dz \\
&= \frac{e^{\frac{x^2}{2}}}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \int_{-x}^\infty (u+x)^{N-1} \exp\left(-\frac{u^2}{2}\right)\ du\ .
\end{aligned}$$

From the binomial formula, we get

$$f_N(x) = \frac{e^{\frac{x^2}{2}}}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \sum_{j=0}^{N-1} \left[\binom{N-1}{j} x^{N-1-j} \int_{-x}^\infty u^j \exp\left(-\frac{u^2}{2}\right)\ du \right]\ . \qquad \text{(B.1)}$$

We now examine the integral in the RHS of (B.1). For odd values of j, we get

$$\begin{aligned}
\int_{-x}^\infty u^j \exp\left(-\frac{u^2}{2}\right)\ du &= \int_{-x}^x u^j \exp\left(-\frac{u^2}{2}\right)\ du + \int_x^\infty u^j \exp\left(-\frac{u^2}{2}\right)\ du \\
&= \int_x^\infty u^j \exp\left(-\frac{u^2}{2}\right)\ du \\
&= \int_0^\infty u^j \exp\left(-\frac{u^2}{2}\right)\ du - \int_0^x u^j \exp\left(-\frac{u^2}{2}\right)\ du \qquad \text{(B.2)}
\end{aligned}$$

where the second equality follows since the integrand is an odd function and the domain of first integral is symmetric around zero. For even values of j, we get

$$\begin{aligned}\int_{-x}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right)\,du &= \int_{0}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right)\,du + \int_{-x}^{0} u^j \exp\left(-\frac{u^2}{2}\right)\,du \\ &= \int_{0}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right)\,du + \int_{0}^{x} u^j \exp\left(-\frac{u^2}{2}\right)\,du \qquad \text{(B.3)}\end{aligned}$$

where the second equality holds since the integrand is an even function. Combining (B.2) and (B.3) gives that for $j \in \{0, 1, \ldots, N-1\}$

$$\begin{aligned}\int_{-x}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right)\,du &= \int_{0}^{\infty} u^j \exp\left(-\frac{u^2}{2}\right)\,du + (-1)^j \int_{0}^{x} u^j \exp\left(-\frac{u^2}{2}\right)\,du \\ &\overset{(a)}{=} \int_{0}^{\infty} (2t)^{\frac{j-1}{2}} e^{-t}\,dt + (-1)^j \int_{0}^{\frac{x^2}{2}} (2t)^{\frac{j-1}{2}} e^{-t}\,dt \\ &= 2^{\frac{j-1}{2}} \int_{0}^{\infty} t^{\frac{j-1}{2}} e^{-t}\,dt \left[1 + (-1)^j \frac{\int_{0}^{\frac{x^2}{2}} t^{\frac{j-1}{2}} e^{-t}\,dt}{\int_{0}^{\infty} t^{\frac{j-1}{2}} e^{-t}\,dt}\right] \\ &= 2^{\frac{j-1}{2}} \Gamma\left(\frac{j+1}{2}\right) \left[1 + (-1)^j\, \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right]\end{aligned}$$

where (a) follows by substituting $t \triangleq \frac{u^2}{2}$ and the functions Γ and $\tilde{\gamma}$ are defined in (59) and (60), respectively. Substituting the last equality in (B.1) and also noting that

$$\binom{N-1}{j} = \frac{\Gamma(N)}{\Gamma(N-j)\,\Gamma(j+1)}\,, \quad N \in \mathbb{N},\ j \in \{0, 1, \ldots, N-1\}$$

we get

$$\begin{aligned}f_N(x) &= \frac{e^{\frac{x^2}{2}}}{2^{\frac{N-1}{2}}\Gamma(\frac{N+1}{2})} \sum_{j=0}^{N-1} \left\{ \frac{\Gamma(N)}{\Gamma(N-j)\,\Gamma(j+1)}\, x^{N-1-j}\, 2^{\frac{j-1}{2}} \right. \\ &\qquad \left. \cdot\Gamma\left(\frac{j+1}{2}\right) \left[1 + (-1)^j\, \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &= \sum_{j=0}^{N-1} \left\{ \frac{e^{\frac{x^2}{2}}}{\Gamma(N-j)}\, \frac{\Gamma(N)}{\Gamma\left(\frac{N+1}{2}\right)}\, \frac{\Gamma\left(\frac{j+1}{2}\right)}{\Gamma(j+1)}\, \frac{x^{N-1-j}}{2^{\frac{N-j}{2}}} \left[1 + (-1)^j\, \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &\overset{(a)}{=} \sum_{j=0}^{N-1} \left\{ \frac{e^{\frac{x^2}{2}}}{\Gamma(N-j)}\, \frac{2^{N-1}\,\Gamma\left(\frac{N}{2}\right)}{\sqrt{\pi}}\, \frac{2^{-j}\,\sqrt{\pi}}{\Gamma\left(\frac{j}{2}+1\right)}\, \frac{x^{N-1-j}}{2^{\frac{N-j}{2}}} \left[1 + (-1)^j\, \tilde{\gamma}\left(\frac{x^2}{2}, \frac{j+1}{2}\right)\right] \right\} \\ &\overset{(b)}{=} \sum_{j=0}^{N-1} \exp\big(d(N, j, x)\big)\end{aligned}$$

where (a) follows from the equality

$$\Gamma(2u) = \frac{2^{2u-1}}{\sqrt{\pi}}\,\Gamma(u)\,\Gamma\left(u + \frac{1}{2}\right), \quad u \neq 0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$$

and (b) follows from the definition of $d(N, j, x)$ in (58).

References

[1] J. Cuevas, P. Adde and S. Kerouedan, "Turbo decoding of product codes for Gigabit per second applications and beyond," *European Transactions on Telecommunications*, vol. 17, no. 1, pp. 45–55, Jan.–Feb. 2006.

[2] C. Di, D. Proietti, I. E. Telatar and R. Urbanke, "Finite-length analysis of low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1570–1579, June 2002.

[3] D. Divsalar and S. Doliner, "Concatenation of Hamming codes and accumulator codes with high-order modulations for high-speed decoding," Jet Propulsion Laboratory (JPL), IPN Progress Report 42-156, February 15, 2004. [Online] Available: `http://tmo.jpl.nasa.gov/progress_report/42-156/156G.pdf`.

[4] S. Doliner, D. Divsalar and F. Pollara, "Code performance as a function of block size," Jet Propulsion Laboratory (JPL), TMO Progress Report 42-133, May 15, 1998. [Online] Available: `http://tmo.jpl.nasa.gov/tmo/progress_report/42-133/133K.pdf`.

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

[6] H. Herzberg and G. Poltyrev, "The error probability of M-ary PSK block coded modulation schemes," *IEEE Trans. on Communications*, vol. 44, pp. 427–433, April 1996.

[7] D. E. Lazic, Th. Beth and M. Calic, "How close are turbo codes to optimal codes ?," *Proceedings of the International Symposium on Turbo Codes and Related Topics*, pp. 192–195, Brest, France, 3–5 September 1997.

[8] D. E. Lazic, Th. Beth and S. Egner, "Constrained capacity of the AWGN channel," *IEEE 1998 International Symposium on Information Theory (ISIT 1998)*, p. 237, Cambridge, MA, USA, 16–21 August, 1998.

[9] S. J. Macmullan nad O.M. Collins, "A comparison of known codes, randsom codes and the best codes," *IEE Trans. on Information Theory*, vol. 44, pp. 3009–3022, November 1998.

[10] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Information Theory*, vol. 40, pp. 1284–1292, July 1994.

[11] R. Urbanke, *Error floor calculator for the binary erasure channel*. [Online]. Available: `http://lthcwww.epfl.ch/research/efc/`.

[12] I. Sason and S. Shamai, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial*," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1-2, pp. 1–222, NOW Publishers, Delft, the Netherlands, July 2006.

[13] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.

[14] C. Shannon, R. Gallager and E. Berlekamp,"Lower bounds to error probability for decoding on discrete memoryless channels," *Information and Control*, vol. 10, Part 1: pp. 65–103, and Part 2: pp. 522–552, February/May 1967.

[15] A. Shokrollahi, "New sequences of time erasure codes approaching channel capacity," in *Proceedings of the 13th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, Lectures Notes in Computer Science 1719, Springer Verlag, pp. 65–76, 1999.

[16] O. Y. Takeshita, O. M. Collins, P. C. Massey and D. J. Costello, "On the frame-error rate of concatenated turbo codes," *IEEE Trans. on Communications*, vol. 49 pp. 602–608, April 2001.

[17] M. Twitto, I. Sason and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," *Proceedings 2006 IEEE International Symposium on Information Theory (ISIT 2006)*, pp. 714–718, Seattle, Washington, USA, 6–12 July 2006.

[18] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block length," *IEEE Trans. on Information Theory*, vol. 50, pp. 2998–3014, Decemeber 2004.

[19] L. Wei, "Near-optimum serial concatenation of single-parity codes with convolutional codes," *IEE Proceedings on Communications*, vol. 152, no. 4, pp. 397–403, August 2005.