Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial
Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial

Igal Sason
Shlomo Shamai
Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel
{sason, sshlomo}@ee.technion.ac.il
Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians
Foundations and Trends® in Communications and Information Theory, 2006, Volume 3, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.
Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial

Igal Sason and Shlomo Shamai

Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel {sason, sshlomo}@ee.technion.ac.il

Abstract

This article is focused on the performance evaluation of linear codes under optimal maximum-likelihood (ML) decoding. Though the ML decoding algorithm is prohibitively complex for most practical codes, their performance analysis under ML decoding allows to predict their performance without resorting to computer simulations. It also provides a benchmark for testing the sub-optimality of iterative (or other practical) decoding algorithms. This analysis also establishes the goodness of linear codes (or ensembles), determined by the gap between their achievable rates under optimal ML decoding and information theoretical limits. In this article, upper and lower bounds on the error probability of linear codes under ML decoding are surveyed and applied to codes and ensembles of codes on graphs. For upper bounds, we discuss various bounds where focus is put on Gallager bounding techniques and their relation to a variety of other reported bounds. Within the class of lower bounds, we address de Caen’s based bounds and their improvements, and also consider sphere-packing bounds with their recent improvements targeting codes of moderate block lengths.
# Contents

## 1 A Short Overview 1

1.1 Introduction 1

1.2 General approach for the derivation of improved upper bounds 3

1.3 On Gallager bounds: Variations and applications 6

1.4 Lower bounds on the decoding error probability 8

## 2 Union Bounds: How Tight Can They Be? 13

2.1 Union bounds 13

2.2 Union bounds for turbo-like codes 17

## 3 Improved Upper Bounds for Gaussian and Fading Channels 21

3.1 The methodology of the bounding technique 21

3.2 Improved upper bounds for the Gaussian channel 23

3.3 Improved upper bounds for fading channels 79

3.4 Concluding comments 88

## 4 Gallager-Type Upper Bounds: Variations, Connections and Applications 91
4.1 Introduction
4.2 Gallager bounds for symmetric memoryless channels
4.3 Interconnections between bounds
4.4 Special cases of the DS2 bound
4.5 Gallager-type bounds for the mismatched decoding regime
4.6 Gallager-type bounds for parallel channels
4.7 Some applications of the Gallager-type bounds
4.8 Summary and conclusions

5 Sphere-Packing Bounds on the Decoding Error Probability
5.1 Introduction
5.2 The 1959 Shannon lower bound for the AWGN channel
5.3 The 1967 sphere-packing bound
5.4 Sphere-packing bounds revisited for moderate block lengths
5.5 Concluding comments

6 Lower Bounds Based on de Caen's Inequality and Recent Improvements
6.1 Introduction
6.2 Lower bounds based on de Caen's inequality and variations
6.3 Summary and conclusions

7 Concluding Remarks
Acknowledgments
References
A Short Overview

Overview: Upper and lower bounds on the error probability of linear codes under maximum-likelihood (ML) decoding are shortly surveyed and applied to ensembles of codes on graphs. For upper bounds, we focus on the Gallager bounding techniques and their relation to a variety of other known bounds. Within the class of lower bounds, we address de Caen’s based bounds and their improvements, and sphere-packing bounds with their recent developments targeting codes of moderate block lengths. This serves as an introductory section, and a comprehensive overview is provided in the continuation of this tutorial.

1.1 Introduction

Consider the classical coded communication model of transmitting one of equally likely signals over a communication channel. Since the error performance of coded communication systems rarely admits exact expressions, tight analytical upper and lower bounds serve as a useful theoretical and engineering tool for assessing performance and for gaining insight into the effect of the main system parameters. As specific good codes are hard to identify, the performance of ensembles of codes
is usually considered. The Fano \cite{71} and Gallager \cite{82} bounds were introduced as efficient tools to determine the error exponents of the ensemble of random codes, providing informative results up to the ultimate capacity limit. Since the advent of information theory, the search for efficient coding systems has motivated the introduction of efficient bounding techniques tailored to specific codes or some carefully chosen ensembles of codes. A classical example is the adaptation of the Fano upper bounding technique \cite{71} to specific codes, as reported in the seminal dissertation by Gallager \cite{81} (to be referred to as the 1961 Gallager-Fano bound). The incentive for introducing and applying such bounds has strengthened with the introduction of various families of codes defined on graphs which closely approach the channel capacity limit with feasible complexity (e.g., turbo codes \cite{23}, repeat-accumulate codes \cite{1,49}, and low-density parity-check (LDPC) codes \cite{124, 156}). Clearly, the desired bounds must not be subject to the union bound limitation, since for codes of large enough block lengths, these ensembles of turbo-like codes perform reliably at rates which are considerably above the cutoff rate ($R_0$) of the channel (recalling that union bounds for long codes are not informative at the portion of the rate region above $R_0$, where the performance of these capacity-approaching codes is most appealing). Although maximum-likelihood (ML) decoding is in general prohibitively complex for long codes, the derivation of upper and lower bounds on the ML decoding error probability is of interest, providing an ultimate indication of the system performance. Further, the structure of efficient codes is usually not available, necessitating efficient bounds on performance to rely only on basic features, such as the distance spectrum and the input-output weight enumeration function (IOWEF) of the examined code (for the evaluation of the block and bit error probabilities, respectively, of a specific code or ensemble). These latter features can be found by analytical methods (see e.g., \cite{127}).

In classical treatments, due to the difficulty in the analytic characterization of optimal codes, random codes were introduced (\cite{71, 82, 83}). This is also the case with modern approaches and practical coding techniques, where ensembles of codes defined on graphs lend themselves to analytical treatment, while this is not necessarily the case for specifically chosen codes within these families. A desirable feature is to
identify efficient bounding techniques encompassing both specific codes and ensembles.

In Sections 2–4, we present various reported upper bounds on the ML decoding error probability, and exemplify the improvement in their tightness as compared to union bounds. These include the bounds of Berlekamp [22], Divsalar [52], Duman-Salehi [59], Engdahl-Zigangirov [67], Gallager-Fano 1961 bound [81], Hughes [97], Poltyrev [153], Sason-Shamai [170], Shulman-Feder [186], Viterbi ([208] and [209]), Yousefi-Khandani ([223] and [224]) and others. We demonstrate in Sections 3 and 4 the underlying connections that exist between them; the bounds are based on the distance spectrum or the IOWEFs of the codes. The focus of this presentation is directed towards the application of efficient bounding techniques on ML decoding performance, which are not subject to the deficiencies of the union bounds and therefore provide useful results at rates reasonably higher than the cut-off rate. In Sections 2–4 and references therein, improved upper bounds are applied to block codes and turbo-like codes. In addressing the Gallager bounds and their variations, we focus in [182] (and more extensively in Section 4) on the Duman and Salehi variation which originates from the standard Gallager bound. A large class of efficient recent bounds (or their Chernoff versions) is demonstrated to be a special case of the generalized second version of the Duman and Salehi bounds. Implications and applications of these observations are addressed in Section 4.

In Sections 5 and 6 we address lower bounds on the ML decoding error probability and exemplify these bounds on linear block codes. Here we overview a class of bounds which are based on de Caen’s bound and its improved version. We also review classical sphere-packing bounds and recent improvements for finite length codes.

We note that every section is self-contained and consequently, notations may (slightly) change from one section to another.

1.2 General approach for the derivation of improved upper bounds

In Sections 3–4 we present many improved upper bounds on the ML decoding error probability which are tighter than the union bound.
A Short Overview

The basic concept which is common to the derivation of the upper bounds within the class discussed in Sections 3 and 4 is the following:

\[ \Pr(\text{error}) = \Pr(\text{error}, y \in \mathcal{R}) + \Pr(\text{error}, y \notin \mathcal{R}) \]
\[ \leq \Pr(\text{error}, y \in \mathcal{R}) + \Pr(y \notin \mathcal{R}) \]  \hspace{1cm} (1.1)

where \( y \) is the received signal vector, and \( \mathcal{R} \) is an arbitrary region around the transmitted signal point which is interpreted as the “good region”. The idea is to use the union bound only for the joint event where the decoder fails to decode correctly, and in addition, the received signal vector falls inside the region \( \mathcal{R} \) (i.e., the union bound is used for upper bounding the first term in the right-hand side (RHS) of (1.1)). On the other hand, the second term in the RHS of (1.1) represents the probability of the event where the received signal vector falls outside the region \( \mathcal{R} \), and which is typically the dominant term for very low SNR, is calculated only one time (and it is not part of the event where the union bound is used). We note that in the case where the region \( \mathcal{R} \) is the whole observation space, the basic approach which is suggested above provides the union bound. However, since the upper bound in (1.1) is valid for an arbitrary region \( \mathcal{R} \) in the observation space, many improved upper bounds can be derived by an appropriate selection of this region. These bounds could be therefore interpreted as geometric bounds (see [52] and [182]). As we will see, the choice of the region \( \mathcal{R} \) is very significant in this bounding technique; different choices of this region have resulted in various different improved upper bounds which are considered extensively in Sections 3 and 4. For instance, the tangential bound of Berlekamp [22] used the basic inequality in (1.1) to provide a considerably tighter bound than the union bound at low SNR values. This was achieved by separating the radial and tangential components of the Gaussian noise with a half-space as the underlying region \( \mathcal{R} \). For the derivation of the sphere bound [90], Herzberg and Poltyrev have chosen the region \( \mathcal{R} \) in (1.1) to be a sphere around the transmitted signal vector, and optimized the radius of the sphere in order to get the tightest upper bound within this form. The Divsalar bound [52] is another simple and tight bound which relies on the basic inequality (1.1). The geometrical region \( \mathcal{R} \) in the Divsalar bound was chosen to be a sphere; in addition to the optimization of the radius of this sphere,
1.2. **General approach for the derivation of improved upper bounds**

The center of the sphere which does not necessarily coincide with the transmitted signal vector was optimized too. Finally, the tangential-sphere bound (TSB) which was proposed for binary linear block codes by Poltyrev [153] and for M-ary PSK block coded-modulation schemes by Herzberg and Poltyrev [91] selected $R$ as a circular cone of half-angle $\theta$, whose central line passes through the origin and the transmitted signal. It is one of the tightest upper bounds known to-date for linear codes whose transmission takes place over a binary-input AWGN channel (see Fig. 1.1 and [168, 170, 223]).

---

**Fig. 1.1** Various bounds for the ensemble of rate $-\frac{1}{3}$ turbo codes whose components are recursive systematic convolutional codes with generators $G_1(D) = G_2(D) = [1, 1+D, \frac{1+D^4}{1+D^2+D^4+D^8}]$. There is no puncturing of the parity bits, and the uniform interleaver between the two parallel concatenated (component) codes is of length 1000. It is assumed that the transmission of the codes takes place over a binary-input AWGN channel. The upper bounds on the bit error probability under optimal ML decoding are compared with computer simulations of the iterative Log-MAP decoding algorithm with up to 10 iterations.
We note that the bounds mentioned above are only a sample of various bounds reported in Section 3; all of these bounds rely on the inequality (1.1) where the geometric region $\mathcal{R}$ characterizes the resulting upper bounds on the decoding error probability. After providing the general approach, we outline some connections between these bounds and demonstrate a few possible applications.

### 1.3 On Gallager bounds: Variations and applications

In addressing the Gallager bounding techniques and their variations, we focus in Section 4 on variations of the Gallager bounds and their applications.

In the following, we present shortly the 1965 Gallager bound [82]. Suppose an arbitrary codeword $x^m$ (of length-$N$) is transmitted over a channel. Let $y$ designate the observation vector (of $N$ components), and $p_N(y|x^m)$ be the channel transition probability measure. Then, the conditional ML decoding error probability is given by

$$P_{e|m} = \sum_{y: \exists m' \neq m: p_N(y|x^{m'}) \geq p_N(y|x^m)} p_N(y|x^m).$$

If the observation vector $y$ is such that there exists $m' \neq m$ so that $p_N(y|x^{m'}) \geq p_N(y|x^m)$, then for arbitrary $\lambda, \rho \geq 0$, the value of the expression

$$\left( \sum_{m' \neq m} \left( \frac{p_N(y|x^{m'})}{p_N(y|x^m)} \right)^{\lambda} \right)^{\rho}$$

is clearly lower bounded by 1, and in general, it is always non-negative. The 1965 Gallager bound [82, 83] therefore states that

$$P_{e|m} \leq \sum_{y} p_N(y|x^m) \left( \sum_{m' \neq m} \left( \frac{p_N(y|x^{m'})}{p_N(y|x^m)} \right)^{\lambda} \right)^{\rho}, \quad \lambda, \rho \geq 0.$$ 

This upper bound is usually not easily evaluated in terms of basic features of particular codes, except for example, orthogonal codes and the special case of $\rho = 1$ and $\lambda = \frac{1}{2}$ (which yields the Bhattacharyya-union bound).
1.3. On Gallager bounds: Variations and applications

An alternative bounding technique which originates from the 1965 Gallager bound is the second version of the Duman and Salehi (DS2) bound (see [61 [182]). This bound is calculable in terms of the distance spectrum, not requiring the fine details of the code structure. A similar upper bound on the bit error probability is expressible in terms of the IOWEFs of the codes (or the average IOWEFs of code ensembles). By generalizing the framework of the DS2 bound, a large class of efficient bounds (or their Chernoff versions) is demonstrated to follow from this bound. Implications and applications of these observations are pointed out in [182], including the fully interleaved fading channel, resorting to either matched or mismatched decoding. The proposed approach can be generalized to geometrically uniform non-binary codes, finite state channels, bit-interleaved coded-modulation systems, parallel channels [119], and it can be also used for the derivation of upper bounds on the conditional decoding error probability. In Section 4, we present the suitability of variations on the Gallager bounds as bounding techniques for random and deterministic codes, which partially rely on insightful observations made by Divsalar [52]. Focus is put in [182] on geometric interpretations of the 1961 Gallager-Fano bound (see [71] and [81]). The interconnections between many reported upper bounds are illustrated in Section 4, where it is shown that the generalized DS2 bound particularizes to these upper bounds by proper selections of the tilting measure. Further details, extensions and examples are provided in Section 4.

The TSB [153] happens often to be the tightest reported upper bound for block codes which are transmitted over the binary-input additive white Gaussian noise (AWGN) channel and ML decoded (see e.g., [168] and [170]). However, in the random coding setting, it fails to reproduce the random coding error exponent (see [153]), while the DS2 bound does. In fact, also the Shulman-Feder bound [180] which is a special case of the latter bound achieves capacity for the ensemble of fully random block codes. This substantiates the claim that there is no uniformly best bound. However, we note that the loosened version of the TSB [52] (which involves the Chernoff inequality) maintains the asymptotic (i.e., for infinite block length) exponential tightness of the TSB of Poltyrev [153], and it is a special case of the DS2 bound.
In the following, we exemplify the use of the DS2 bounding technique for fully interleaved fading channels with faulty measurements of the fading samples.

Example 1.1. The Generalized DS2 bound for the Mismatched Regime. In [182], we apply the generalized DS2 bound to study the robustness of a mismatched decoding that is based on ML decoding with respect to the faulty channel measurements. We examine here the robustness of the decoder in case that a BPSK modulated signal is transmitted through a fully interleaved Rayleigh fading channel. For simplicity, the bounds are applied to the case of perfect phase estimation of the i.i.d. fading samples (in essence reducing the problem to a real channel). We also assume here that the estimated and real magnitudes of the Rayleigh fading samples have a joint distribution of two correlated bivariate Rayleigh variables with an average power of unity.

The bounds in Fig. 1.2 refer to the ensemble of uniformly interleaved rate $-\frac{1}{3}$ turbo codes whose components are recursive systematic convolutional codes: $G_1(D) = G_2(D) = \left[1, \frac{1+D^4}{1+D^2+D^3+D^4}\right]$ without puncturing of parity bits, and an interleaver length of $N = 1000$. Since for a fully interleaved Rayleigh fading channel with perfect side information on the fading samples, the matched channel cutoff rate corresponds to $\frac{E_b}{N_0} = 3.23$ dB then, according to the upper bounds depicted in Fig. 1.2 the ensemble performance of these turbo codes (associated with the ML decoding) is sufficiently robust in case of mismatched decoding, even in a portion of the rate region exceeding the channel matched cutoff rate. The proposed upper bounds depicted here were efficiently implemented in software, thus indicating their feasible computational complexity.

1.4 Lower bounds on the decoding error probability

1.4.1 De Caen inequality and variations

D. de Caen [45] suggested a lower bound on the probability of a finite union of events. While an elementary result (essentially, the Cauchy-Schwartz inequality), it was used to compute lower bounds on the
1.4. Lower bounds on the decoding error probability

Fig. 1.2 A comparison between upper bounds on the bit error probability for the ensemble of turbo codes considered in Example [11] where the transmission of these codes takes place over a fully interleaved Rayleigh fading channel with mismatched decoding. The bounds are based on the combination of the generalized DS2 bound and the tight form of the union bound applied to every constant Hamming-weight subcode. These bounds are plotted for $E_b/N_0 = 2.50, 2.75, 3.00$ and $3.25$ dB, as a function of the correlation coefficient between the actual i.i.d Rayleigh fading samples and their Rayleigh distributed estimations.

decoding error probability of linear block codes via their distance distribution (see [108] for the binary symmetric channel (BSC), and [181] for the Gaussian channel). In [39], Cohen and Merhav improved de Caen’s inequality by introducing an arbitrary non-negative weighting function which is subject to optimization. The concept of this improved bound is presented in the following statement and, like de Caen’s inequality, it follows from the Cauchy-Schwartz inequality.

**Theorem 1.2.** [39, Theorem 2.1] Let $\{A_i\}_{i \in I}$ be an arbitrary set of events in a probability space $(\Omega, \mathcal{F}, P)$, then the probability of the union...
of these events is lower bounded by
\[
P\left(\bigcup_{i \in I} A_i\right) \geq \sum_{i \in I} \left\{ \frac{\left(\sum_{x \in A_i} p(x)m_i(x)\right)^2}{\sum_{j \in I} \sum_{x \in A_i \cap A_j} p(x)m_i(x)^2} \right\}
\]
where \( m_i \) is an arbitrary non-negative function on \( \Omega \) such that the sums in the RHS converge. Further, equality is achieved when
\[
m_i(x) = m^*(x) \triangleq \frac{1}{\text{deg}(x)}, \quad \forall i \in I
\]
where for each \( x \in \Omega \)
\[
\text{deg}(x) \triangleq |\{i \in I \mid x \in A_i\}|.
\]

The lower bound on the union of events in Theorem 1.2 particularizes to de Caen’s inequality by the particular choice of the weighting functions \( m_i(x) = 1 \) for all \( i \in I \), which then gives
\[
P\left(\bigcup_{i \in I} A_i\right) \geq \sum_{i \in I} \frac{P(A_i)^2}{\sum_{j \in I} P(A_i \cap A_j)}.
\]
Cohen and Merhav relied on Theorem 1.2 for the derivation of improved lower bounds on the decoding error probability of linear codes under optimal ML decoding. They exemplified their bounds for BPSK modulated signals which are equally likely to be transmitted among \( M \) signals, and the examined communication channels were a BSC and an AWGN channel. In this context, the element \( x \) in Theorem 1.2 is replaced by the received vector \( y \) at the output of the communication channel, and \( A_i \) (where \( i = 1, 2, \ldots, M - 1 \)) consists of all the vectors which are closer in the Euclidean sense to the signal \( s^i \) rather than the transmitted signal \( s^0 \). Following [181], the bounds in [39] get (after some loosening in their tightness) final forms which solely depend on the distance spectrum of the code. Recently, two lower bounds on the
ML decoding error probability of linear binary block codes were derived by Behnamfar et al. [16] for BPSK-modulated AWGN channels. These bounds are easier for numerical calculation, but are looser than Cohen-Merhav bounds for low to moderate SNRs.

Note that de Caen’s based lower bounds on the decoding error probability (see [16], [39], [108] and [181]) are applicable for specific codes but not for ensembles; this restriction is due to the fact that Jensen’s inequality does not allow to replace the distance spectrum of a linear code in these bounds by the average distance spectrum of ensembles.

1.4.2 Sphere-packing bounds revisited for moderate block lengths

In the asymptotic case where the block length of a code tends to infinity, the best known lower bound on the decoding error probability for discrete memoryless channels (DMCs) with high levels of noise is the 1967 sphere-packing (SP67) bound [184]. Like the random coding bound of Gallager [82], the sphere-packing bound decreases exponentially with the block length. Further, the error exponent of the SP67 bound is a convex function of the rate which is known to be 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 exponent of the SP67 bound coincides with the error exponent of the random coding bound [184, Part 1]. For the AWGN channel, the 1959 sphere-packing (SP59) bound was derived by Shannon [183] by showing that the error probability of any code whose codewords lie on a sphere must be greater than the error probability of a code of the same length and rate whose codewords are uniformly distributed over that sphere.

The reason that the SP67 bound fails to provide useful results for codes of small to moderate block length is due to the original focus in [184] on asymptotic analysis. In their paper [203], Valembois and Fossorier have recently revisited the SP67 bound in order to make it applicable for codes of moderate block lengths, and also to extend its field of application to continuous output channels (e.g., the AWGN channel which is the communication channel model of the SP59 bound.
of Shannon [183]. The motivation for the study in [203] was strengthened due to the outstanding performance of codes defined on graphs with moderate block length. The remarkable improvement in the tightness of the SP67 was exemplified in [203] for the case of the AWGN channel with BPSK signaling, and it was shown that in some cases, the improved version of the SP67 presents an interesting alternative to the SP59 bound [183].
References


References


Full text available at: http://dx.doi.org/10.1561/0100000009
References


References


Full text available at: http://dx.doi.org/10.1561/0100000009
References


Full text available at: http://dx.doi.org/10.1561/0100000009
References


References


References


References


